Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis

Hai Wan,[1] Yufei Yang,[1] Jianfeng Du,[2]* Yanan Liu,[1] Kunxun Qi,[2] Jeff Z. Pan[3]

[1]School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, P.R.China
[2]Guangzhou Key Laboratory of Multilingual Intelligent Processing, Guangdong University of Foreign Studies,
Guangzhou 510006, P.R.China
[3]Department of Computing Science, The University of Aberdeen, Aberdeen AB24 3UE, UK
wanhai@mail.sysu.edu.cn, jfdu@gdufs.edu.cn, {yangyf35, liuyn56}@mail3.sysu.edu.cn,
kunxunqi@foxmail.com, jeff.z.pan@abdn.ac.uk

Abstract

Aspect-based sentiment analysis (ABSA) aims to detect the targets (which are composed by continuous words), aspects and sentiment polarities in text. Published datasets from SemEval-2015 and SemEval-2016 reveal that a sentiment polarity depends on both the target and the aspect. However, most of the existing methods consider predicting sentiment polarities from either targets or aspects but not from both, thus they easily make wrong predictions on sentiment polarities. In particular, where the target is implicit, *i.e.*, it does not appear in the given text, the methods predicting sentiment polarities from targets do not work. To tackle these limitations in ABSA, this paper proposes a novel method for target-aspect-sentiment joint detection. It relies on a pre-trained language model and can capture the dependence on both targets and aspects for sentiment prediction. Experimental results on the SemEval-2015 and SemEval-2016 restaurant datasets show that the proposed method achieves a high performance in detecting target-aspect-sentiment triples even for the implicit target cases; moreover, it even outperforms the state-of-the-art methods for those subtasks of target-aspect-sentiment detection that they are competent to.

Introduction

Sentiment analysis, aiming to detect the sentiment expressed in text, is a fundamental task in natural language processing. Since the sentiments in a sentence can be complex and varied by different aspects, *aspect-based sentiment analysis* (ABSA) is proposed to refine sentiment analysis. It often aims to detect fine-grained opinions towards different aspects. Recently ABSA has gained more and more attention especially with the rise of social media and public opinion.

SemEval-2015 Task 12 (Pontiki et al. 2015) and SemEval-2016 Task 5 (Pontiki et al. 2016) formalize ABSA as a task for target-aspect-sentiment detection from a sentence, where the target is composed of continuous words in the sentence, the aspect is drawn from a predefined vocabulary, and the sentiment is a polarity (positive, negative, or neutral) towards the target and the aspect. Figure 1 gives three examples in the restaurant dataset from SemEval-2016

```
id = 1189674:3
   Admittedly some nights inside the restaurant were rather warm, but
the open kitchen is part of the charm.

target: open kitchen  aspect: AMBIENCE#GENERAL  sentiment: positive
target: restaurant    aspect: AMBIENCE#GENERAL  sentiment: negative
```

```
id = 1632445:5
   Even though its good seafood, the prices are too high.

target: seafood       aspect: FOOD#QUALITY      sentiment: positive
target: seafood       aspect: FOOD#PRICES       sentiment: negative
```

```
id = 1726473:4
   The food arrived 20 minutes after I called, cold and soggy.

target: NULL          aspect: SERVICE#GENERAL   sentiment: negative
target: food          aspect: FOOD#QUALITY      sentiment: negative
```

Figure 1: Three examples in the restaurant dataset from SemEval-2016 Task 5.

Task 5. The first example shows that the sentiment cannot be determined by the sentence and the aspect. The second example shows that the sentiment cannot be determined by the sentence and the target. That is, the sentiment depends on both the target and the aspect. Moreover, the target of the last example can be given implicitly; *i.e.*, it is assigned NULL and does not contain any word in the sentence.

Two restaurant datasets are published by the above two competitions, respectively. They give new challenges to ABSA. The primary challenge is that the sentiment actually depends on both the target and the aspect. Most existing studies in ABSA do not handle this dual dependence. For example, some studies such as (Wang et al. 2016) and (Xue and Li 2018) predict sentiments from aspects alone, while other studies such as (Schmitt et al. 2018) and (Sun, Huang, and Qiu 2019) predict sentiments from targets alone. Moreover, most existing studies such as (Ma, Li, and Wang 2018; Wang, Lan, and Wang 2018; Luo et al. 2019b) ignore implicate target cases. However, from the restaurant datasets, there are about one fourth of opinions that have implicit targets (see Table 1). Thus handling implicit target cases is also a non-neglectable challenge in ABSA.

The above challenges call for a solution to the task of *target-aspect-sentiment detection* (TASD), namely detecting target-aspect-sentiment triples from a sentence. By considering that the target and the aspect are highly related and that

detecting the target benefits detecting the aspect and vice versa, we propose to detect the three elements (target, aspect and sentiment) simultaneously. To this end, we develop a novel neural based method for target-aspect-sentiment joint detection. The method separates the joint detection problem into two subproblems on the basis of aspect-sentiment pairs, where for every given aspect-sentiment pair, one subproblem determines whether targets exist and is reducible to a binary text classification problem, and the other extracts all targets and is reducible to a sequence labeling problem. Both subproblems are solved by a single neural model built upon the pre-trained language model BERT (Devlin et al. 2019). The neural model is trained by minimizing a combined loss function about two subproblems.

We conduct experiments on the aforementioned two restaurant datasets. The results show that our proposed method achieves a high performance in the TASD task and works significantly better than solving the two subproblems separately. We also evaluate five subtasks of TASD, namely aspect-sentiment joint detection (ASD), target-sentiment joint detection (TSD), target-aspect joint detection (TAD), aspect detection (AD) and target detection (TD). The results show that our method outperforms the state-of-the-art methods for all those subtasks that they are competent to.

The main contributions of this work include:

- We propose a solution to capture the dual dependence of sentiments on both targets and aspects and to handle implicit target cases.

- We propose a neural model built upon the pre-trained language model BERT, which can be used to predict target-aspect-sentiment together.

- We empirically verify that the proposed method achieves the state-of-the-art performance in the task of target-aspect-sentiment joint detection as well as all its subtasks.

Related Work

Most existing studies for ABSA focus on the five subtasks of TASD, namely ASD, TSD, TAD, AD and TD.

The ASD task is designed to detect aspects and sentiments simultaneously. (Schmitt et al. 2018) tackles this task by an end-to-end CNN model. A more recent study (Sun, Huang, and Qiu 2019) reduces the problem of ASD to a set of binary classification problem and solves it by fine-tuning a pre-trained language model.

The TSD task aims to jointly detect targets and sentiments. In the earliest, (Mitchell et al. 2013) and (Zhang, Zhang, and Vo 2015) reduce the TSD task to a sequence labeling problem and solve it by a CRF decoder with hand-crafted linguistic features. Recently neural models are widely used. (Li et al. 2019) proposes a unified model composed of two stacked LSTM networks to tackle the TSD task. (Luo et al. 2019b) presents a dual cross-shared RNN model for TSD which uses sentiment lexicon and part-of-speech of words as auxiliary information. (Hu et al. 2019) introduces a span-based pipeline framework and solves the TSD task by fine-tuning a pre-trained language model.

There is little work specifically designed to the TAD task, which aims to detect targets and aspects together. In contrast,

there have been much work addressing the AD task, which aims to detect aspects only. Pioneer studies on AD such as (Kiritchenko et al. 2014; Xenos et al. 2016) train SVM classifiers to detect aspects. (Liu, Cohn, and Baldwin 2018) introduces neural models to improve the performance of the AD task. In the studies (He et al. 2017; Xue et al. 2017; Ma, Peng, and Cambria 2018; Movahedi et al. 2019), different attention mechanisms are introduced to a neural model to detect aspects in a more accurate way.

The TD task aims to extract targets only. Traditionally, the targets are extracted by CRF (Jakob and Gurevych 2010; Yin et al. 2016) or by syntactic patterns (Liu, Xu, and Zhao 2013). Lately, neural models such as CNN (Xue et al. 2017; Xu et al. 2018) and RNN (Li and Lam 2017; Xue et al. 2017; Luo et al. 2019a) are widely used in target extraction. In the studies (Wang et al. 2017; He et al. 2017; Li et al. 2018), different attention mechanisms are also introduced to a neural model to extract targets more accurately.

All the above tasks cannot capture the dual dependence of sentiments on both targets and aspects. Only the TASD task can capture this dual dependence, which aims to jointly detect target-aspect-sentiment triples. As far as we know, there is only one study (Brun and Nikoulina 2018) addressing the TASD task. It proposes a method relying on available parsers and domain-specific semantic lexicons, but this method performs poorly as shown in our experiments.

Besides the above tasks, there are two simpler tasks for ABSA, where one aims to classify the sentiment according to a given aspect, which has been investigated *e.g.* in (Wang et al. 2016; Xue and Li 2018), and the other aims to classify the sentiment according to a given target, studied *e.g.* in (Zeng, Ma, and Zhou 2019). Since these tasks rely on prerequisite tasks such as AD or TD to fulfill ABSA, they are incomparable to the TASD task.

Target-Aspect-Sentiment Detection

Problem Definition

Given a sentence S consisting of n words $s_1, \ldots, s_n$, a predefined set A of aspects and a predefined set P of sentiment polarities, the TASD task aims to detect all triples (t, a, p) that S entails in the natural language meaning, where t (called a *target*) is a subsequence of S, a is an aspect in A and p is a sentiment polarity (simply called a *sentiment*) in P. The target t can be empty, denoted by NULL. This case is referred to as an implicit target case. Since the general goal of ABSA is detecting fine-grained opinions, we also call the triple (t, a, p) an *opinion*. Consider the last example given in Figure 1. There are two opinions (NULL, SERVICE#GENERAL, negative) and (food, FOOD#QUALITY, negative) detected from the sentence with id 1726473:4.

Problem Reduction

Considering that the text classification problem and the sequence labeling problem are well-studied in natural language processing, we attempt to reduce the problem of TASD to a set of text classification problems and sequence labeling problems such that the state-of-the-art methods for the reduced problems can be adapted to the TASD task.
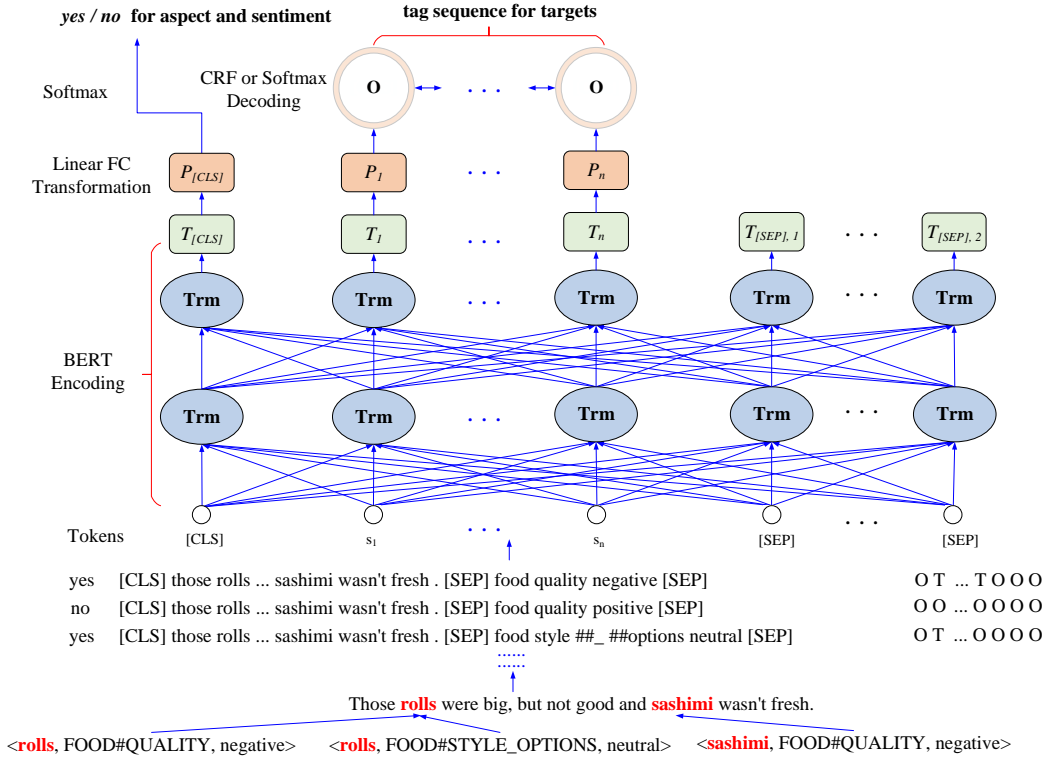
Figure 2: The architecture and a running example for the TAS-BERT model. TAS-BERT takes a sentence-aspect-sentiment token sequence "[CLS]···[SEP]···[SEP]" as input. It outputs "*yes/no*" for predicting whether targets exist for the aspect-sentiment pair and a tag sequence for extracting the targets.

To this end, we divide a given problem which detects all opinions from a sentence S, a set A of aspects and a set P of sentiments into $|A||P|$ problems on the basis of all aspect-sentiment pairs, where $|X|$ is the cardinality of the set X.

Every resulting problem is further separated into two subproblems, where one determines whether targets exist for the given aspect-sentiment pair, and the other extracts the targets corresponding to the given aspect-sentiment pair. The first subproblem can be reduced to a binary text classification problem, with "*yes*" indicating that at least one target (including the implicit target) exists and "*no*" indicating that no target exists. The second subproblem can be reduced to a sequence labeling problem using either the BIO tagging scheme or the TO tagging scheme, where "B" (resp. "I") denotes the starting (resp. an internal) word of a target, "T" a word inside a target and "O" a word outside any target.

The results of the two subproblems can be merged to get opinions. Given a sentence S and an aspect-sentiment pair (a, p), if the first subproblem outputs "*no*", there will be no opinion of the form (t, a, p) that can be detected from S. Otherwise, suppose there are n subsequences of S output by the second subproblem where $n \geq 0$. If $n = 0$, there is only one opinion (NULL, a, p) with implicit target detected from S, otherwise there are n opinions (t, a, p) detected from S for t a subsequence output by the second subproblem.

The TAS-BERT Model

We propose a neural based model for solving the aforementioned two subproblems together. In order to guarantee a high prediction performance, the model is built upon a pre-trained language model BERT (Devlin et al. 2019) and is named TAS-BERT.[1] The proposed model consists of five components, including a BERT encoder, two linear fully-connected (FC) layers, a softmax decoder for binary classification of "*yes/no*", and either a conditional random field (CRF) decoder (Ma and Hovy 2016) or a softmax decoder for sequence labeling, as shown in Figure 2.

The training set of the TASD task is given by a set of sentence-opinion pairs where one sentence corresponds to multiple opinions. In order to learn the proposed model over the training set, we need to preprocess the training set to obtain a tuple $(S, a, p, f, \mathbf{T})$ for every combination of S, a and p, where S is a sentence in the training set, a an aspect appearing in the training set, p a sentiment appearing in the training set, f a "*yes/no*" label, and $\mathbf{T}$ is a sequence of labels in either the BIO or the TO tagging scheme. The label f and the label sequence $\mathbf{T}$ are constructed as follows. Suppose $(t_1, a, p), \ldots, (t_k, a, p)$ are all opinions corresponding to S in the training set where k can be zero. If $k = 0$, we set f as

[1]Code and experimental datasets for TAS-BERT are available at https://github.com/sysulic/TAS-BERT

"*no*" and construct $\mathbf{T}$ as an all-O label sequence with length n where n is the number of words in S no matter which tagging scheme is used. Otherwise, we set f as "*yes*" and encode the k targets $t_1, \ldots, t_k$ to a label sequence $\mathbf{T}$ using the specified tagging scheme. We always treat the implicit target NULL as the last target t_k. In case two targets t_i and t_j (where $i < j$) have overlapped words, we ignore the latter target t_j. In case the implicit target exists, we construct $\mathbf{T}$ as an all-O label sequence if the implicit target is t_1, or ignore the implicit target otherwise. According to the restaurant datasets published by SemEval-2015 Task 12 (Pontiki et al. 2015) and SemEval-2016 Task 5 (Pontiki et al. 2016), this preprocessing step ignores only a very few opinions and can retain as many implicit target cases as possible.

In the training phase, when given a tuple $(S, a, p, f, \mathbf{T})$ for S a sentence consisting of n words $s_1, \ldots, s_n$, a an aspect, p a sentiment, f a "*yes/no*" label and $\mathbf{T}$ a label sequence, we first construct a token sequence "[CLS], $s_1, \cdots, s_n$, [SEP], $a_1, \cdots, a_m$, p, [SEP]" composed of $n + m + 4$ tokens, where $a_1, \cdots, a_m$ are words that constitute a, and [CLS] and [SEP] are tokens specifically introduced in BERT. This token sequence is fed into the BERT encoder, outputting a sequence of d-dimensional vectors $T_{[CLS]}, T_{s_1}, \ldots, T_{s_n}, T_{[SEP],1}, T_{a_1}, \ldots, T_{a_m}, T_p, T_{[SEP],2}$ at the final layer of BERT, where the two [SEP] tokens correspond to two different vectors $T_{[SEP],1}$ and $T_{[SEP],2}$.

The first vector $T_{[CLS]}$ is used to predict a "*yes/no*" label through a FC layer followed by a softmax decoder. More precisely, the probability distribution vector $g \in \mathbb{R}^2$ on the "*yes/no*" label is defined below.

$$P_{[CLS]} = \tanh(W_1 T_{[CLS]} + b_1) \qquad (1)$$
$$g = \text{softmax}(P_{[CLS]}) \qquad (2)$$

where $W_1 \in \mathbb{R}^{d \times 2}$ and $b_1 \in \mathbb{R}^2$ are trainable parameters.

The next n vectors $T_{s_1}, \ldots, T_{s_n}$ are used to predict a label sequence in the specified tagging scheme. These vectors are fed into another FC layer followed by a CRF decoder or a softmax decoder. In more details, the vector P_{s_i} computed from T_{S_i} (where $1 \le i \le n$) by the FC layer is defined as

$$P_{s_i} = \tanh(W_2 T_{s_i} + b_2) \qquad (3)$$

where $W_2 \in \mathbb{R}^{d \times o}$ and $b_2 \in \mathbb{R}^o$ are trainable parameters, and where $o = 3$ if the BIO tagging scheme is used or $o = 2$ if the TO tagging scheme is used.

In case a CRF decoder is used, the probability for predicting a label sequence $\mathbf{T}$, denoted by $p(\mathbf{T} \mid \mathbf{P})$, can be computed in the way presented by (Ma and Hovy 2016), where $\mathbf{P}$ is a $n \times o$ matrix composed of $P_{s_1}, \ldots, P_{s_n}$. In case a softmax decoder is used, the probability distribution vector $h_i \in \mathbb{R}^o$ (where $1 \le i \le n$) on the ith label in $\mathbf{T}$ is defined as $\text{softmax}(P_{s_i})$.

The loss value for predicting the "*yes/no*" label f is defined as

$$\text{loss}_g = -\sum_{i=1}^{2} I(\text{yn}(i) = f) \log(g_i) \qquad (4)$$

where g_i is the ith element of g, $\text{yn}(1) = yes$, $\text{yn}(2) = no$, and $I(X) = 1$ if X is true or $I(X) = 0$ otherwise.

The loss value for predicting the label sequence $\mathbf{T} = \langle t_1, \ldots, t_n \rangle$ is defined as

$$\text{loss}_h = -\log(p(\mathbf{T} \mid \mathbf{P})) \qquad (5)$$

if a CRF decoder is used; otherwise, it is defined as

$$\text{loss}_h = -\sum_{i=1}^{n} \sum_{j=1}^{o} I(\text{map}(j) = t_i) \log(h_{ij}) \qquad (6)$$

where h_{ij} is the jth element of h_i, and $\text{map}(1) = \text{B}$, $\text{map}(2) = \text{I}$, $\text{map}(3) = \text{O}$ when the BIO tagging scheme is used, or $\text{map}(1) = \text{T}$, $\text{map}(2) = \text{O}$ when the TO tagging scheme is used.

The proposed TAS-BERT model is trained by minimizing the following combined loss function over all training tuples.

$$\text{loss} = \sum_{i=1}^{N} \text{loss}_g{}^i + \text{loss}_h{}^i \qquad (7)$$

where N is the number of training tuples, and $\text{loss}_g{}^i$ and $\text{loss}_h{}^i$ denote the two loss values for the ith training tuple.

In the prediction phase, when a sentence S is given, we construct a triple (S, a, p) for every aspect a appearing in the training set and every sentiment p appearing in the training set. Afterwards, we feed (S, a, p) into the TAS-BERT model, yielding a "*yes/no*" label for the first subproblem as well as a possibly empty set of subsequences of S for the second subproblem. Finally, we apply the method described in the previous subsection to merge the results of the two subproblems and to obtain opinions for (S, a, p).

Experiments

Datasets

We conducted experiments on two datasets in the restaurant domain, where one (denoted Res15) is from SemEval-2015 Task 12 and the other (denoted Res16) is from SemEval-2016 Task 5. Although most existing studies experimented on the dataset from SemEval-2014 Task 4, we did not use this dataset since it does not provide target-aspect-sentiment triples and is unsuitable for the TASD task.

Table 1 reports the statistics on the two experimental datasets. These statistics reveal the limitations of most subtasks of TASD. The TSD task and the TD task ignore all opinions with implicit targets. This is a crucial limitation since there are about one fourth of opinions that have implicit targets. All opinions can be grouped by the same sentence and the same target. We call every resulting group a *target-sharing opinion group*. It can be seen from the table that there exist some target-sharing opinion groups that have multiple sentiments. Since the TSD task does not consider aspects, any method for TSD is impossible to predict completely correct sentiments for multi-sentiment target-sharing opinion groups. On the other hand, all opinions can also be grouped by the same sentence and the same target. We call every resulting group a *aspect-sharing opinion group*. It can also be seen that there exist some aspect-sharing opinion groups that have multiple sentiments. Since the ASD task does not consider targets, any method for ASD is impossible to predict completely correct sentiments for multi-sentiment aspect-sharing opinion groups.

Table 1: The statistics on Res15 and Res16.

Datasets		Sentences	Opinions		target-sharing opinion groups		aspect-sharing opinion groups	
			all opinions	opinions with implicit targets	all groups	multi-sentiment groups	all groups	multi-sentiment groups
Res15	Train	1315	1654	375 (22.67%)	1545	22 (1.42%)	1478	27 (1.83%)
	Test	685	845	248 (29.35%)	761	21 (2.76%)	775	14 (1.81%)
Res16	Train	2000	2507	627 (25.01%)	2312	44 (1.90%)	2258	42 (1.86%)
	Test	676	859	208 (24.21%)	810	16 (1.98%)	743	8 (1.08%)

Experimental Setup

In our experiments, we considered both the BIO and the TO tagging schemes. Since the CRF decoder in our TAS-BERT model only works well for tagging schemes that have transition constraints, we only considered the combination of the CRF decoder and the BIO tagging scheme as well as the combination of the softmax decoder and the TO tagging scheme. In addition, the restaurant datasets contain a number of words that are out of the vocabulary of the pre-trained language model BERT[2], thus we considered two methods for dealing with the unknown words. One method is replacing an unknown word with the longest prefix match in the vocabulary, called the Longest Prefix Match (LPM) method. The other method is splitting an unknown word into several recognizable words based on the vocabulary, called the Split Word (SW) method. Hence, we obtained four variants of the proposed model, denoted by TAS-BERT-LPM-BIO-CRF, TAS-BERT-LPM-TO, TAS-BERT-SW-BIO-CRF and TAS-BERT-SW-TO, respectively. To train these models, we set the dropout probability as 0.1 for all layers, the max sequence length as 128, the learning rate as 2e-5, and the maximum number of epochs as 30.

Comparison Methods

We compared our method with the following methods on the TASD task and all its subtasks.[3]

- **E2E-TBSA**: E2E-TBSA (Li et al. 2019) is a unified model for TSD. We used the published code to evaluate on Res15 and Res16.

- **DOER**: DOER (Luo et al. 2019b) is a dual cross-shared RNN model for TSD. We used the published code to evaluate on Res15 and Res16.

- **BERT-pair-NLI-B**: BERT-pair-NLI-B (Sun, Huang, and Qiu 2019) is a BERT based model for ASD. We used the published code to evaluate on Res15 and Res16 for both ASD and its subtask AD.

- **SemEval-Top**: SemEval-Top represents the the best scores in the SemEval competitions. They involve three subtasks AD, TD and TAD.

- **MTNA**: MTNA (Xue et al. 2017) is a multi-task model based on RNN and CNN. The paper reported results on Res15 and Res16 for both AD and TD.

---

[2]We used the uncased BERT base model available at https://github.com/google-research/bert.

[3]We cannot reproduce the reported results of TAN and DE-CNN due to problems in the published code, thus we only give the reported results for Res16 alone in Table 2.

- **TAN**: TAN (Movahedi et al. 2019) is a neural model with the multi-attention mechanism for AD. The paper only reported results on Res16.

- **Sentic LSTM + TA + SA**: It (Ma, Peng, and Cambria 2018) augments the LSTM network with target-level attention and sentence-level attention. The source code is not provided and the paper only reported results on Res15.

- **DE-CNN**: DE-CNN (Xu et al. 2018) is a CNN model for TD. The paper only reported results on Res16.

- **THA + STN**: THA + STN (Li et al. 2018) is neural model for TD with a bi-linear attention layer and a FC layer. The paper reported results on both Res15 and Res16.

- **BERT-PT**: BERT-PT (Xu et al. 2019) is a BERT based model for TD. We used the published code to evaluate on both Res15 and Res16.

- **baseline-1-f_lex**: baseline-1-f_lex (Brun and Nikoulina 2018) is a pipeline method for TASD. The source code is not provided and the paper only reported results on Res15 for both TASD and its subtask ASD.

For all BERT based methods including ours, we used the same BERT model. For ASD, TAD, AD and TASD, we evaluated on the full datasets. For TD and TSD, we evaluated on the partial datasets without implicit targets. In particular for TSD, since existing methods are unable to handle opinions with implicit targets while our method can, we evaluated our method on both the full datasets and the partial datasets.

Result Analysis

We used micro-F1 score (in percent) as the evaluation metric for all tasks. For the four variants of our method, we used the output of the TASD task to estimate the corresponding micro-F1 scores for all subtasks of TASD. The comparison results are reported in Table 2.

- **Results on ASD**: Currently, there are fewer studies on ASD. The baseline-1-f_lex method performs poorly, probably because the linguistic features are extracted from an NLP pipeline, which leads to error accumulation for the joint task. BERT-pair-NLI-B, which subtly transforms the problem of ASD into a binary classification problem, achieves significantly better results. However, it adopts the assumption that the sentiment depends on the aspect alone, thus it cannot work correctly for multi-sentiment aspect-sharing opinion groups and cannot make use of the dependence relationship between targets and aspects to further improve the performance. All four variants of our method except TAS-BERT-LPM-BIO-CRF on Res16 achieve better results than comparison methods.

Table 2: Comparison results for six tasks on ABSA, where "-" denotes unreported results. For the TSD task, scores outside brackets are for test sets without implicit targets, whereas scores in brackets are for the full test sets.

ASD			TSD			TAD		
Method	Res15	Res16	Method	Res15	Res16	Method	Res15	Res16
baseline-1-f_lex	-	63.50	E2E-TBSA	53.00	63.10	SemEval-Top	42.90	52.61
BERT-pair-NLI-B	63.67	72.70	DOER	56.33	65.91			
TAS-BERT-LPM-BIO-CRF	65.07	72.08	TAS-BERT-LPM-BIO-CRF	64.49 (63.25)	73.42 (70.70)	TAS-BERT-LPM-BIO-CRF	61.09	70.93
TAS-BERT-LPM-TO	67.75	73.87	TAS-BERT-LPM-TO	63.82 (62.85)	69.29 (67.94)	TAS-BERT-LPM-TO	60.86	67.55
TAS-BERT-SW-BIO-CRF	68.50	74.12	TAS-BERT-SW-BIO-CRF	**66.11** (64.29)	**75.68** (**72.92**)	TAS-BERT-SW-BIO-CRF	**63.37**	**71.64**
TAS-BERT-SW-TO	**70.42**	**76.33**	TAS-BERT-SW-TO	64.84 (**65.02**)	73.34 (71.02)	TAS-BERT-SW-TO	62.60	69.98
AD			TD			TASD		
Method	Res15	Res16	Method	Res15	Res16	Method	Res15	Res16
SemEval-Top	62.68	73.03	SemEval-Top	70.05	72.34	baseline-1-f_lex	-	38.10
MTNA	65.97	76.42	MTNA	67.73	72.95			
Sentic LSTM + TA + SA	73.82	-	DE-CNN	-	74.37			
TAN	-	78.38	THA + STN	71.46	73.61			
BERT-pair-NLI-B	70.78	80.25	BERT-PT	73.15	77.97			
TAS-BERT-LPM-BIO-CRF	73.74	81.12	TAS-BERT-LPM-BIO-CRF	74.10	79.77	TAS-BERT-LPM-BIO-CRF	54.76	64.66
TAS-BERT-LPM-TO	74.22	81.67	TAS-BERT-LPM-TO	72.17	75.15	TAS-BERT-LPM-TO	55.47	62.29
TAS-BERT-SW-BIO-CRF	76.34	81.57	TAS-BERT-SW-BIO-CRF	**75.00**	**81.37**	TAS-BERT-SW-BIO-CRF	57.51	**65.89**
TAS-BERT-SW-TO	**76.40**	**82.77**	TAS-BERT-SW-TO	71.54	78.10	TAS-BERT-SW-TO	**58.09**	65.44

- **Results on TSD**: Existing studies on TSD often perform joint extraction of targets and sentiments by transforming TSD into a sequence labeling problem. Based on this idea, both E2E-TBSA and DOER are unable to distinguish absence of opinions from opinions with implicit targets. On the contrary, our TAS-BERT model makes use of the result from a binary classification problem to distinguish absence of opinions from opinions with implicit targets, thus it can work correctly for opinions with implicit targets. All four variants of our method achieve significantly better results than comparison methods.

- **Results on TAD**: Due to little work on TAD, we can only compare the results achieved by our method with the best results in the SemEval competitions. All four variants of our method achieve significantly better results than the champions in the SemEval competitions.

- **Results on AD and TD**: Almost all four variants of our method achieve better results than comparison methods. This confirms that the dependence relationship between aspects and targets helps both aspect detection and target detection. In particular, a method for pure AD cannot distinguish different aspects for different targets in the same sentence, while a method for pure TD cannot distinguish different targets for different aspects in the same sentence.

- **Results on TASD**: There is little work on this task, and the only method baseline-1-f_lex that we can find only evaluated on Res16. baseline-1-f_lex uses the linguistic features obtained from a NLP pipeline and extracts three elements of an opinion separately. All four variants of our method achieve much better results than this baseline.

In addition, by comparing the results achieved by the four variants of our method, we can see that SW methods always outperforms LPM methods in dealing with unknown words. When target extraction is not involved, namely in the AD and ASD tasks, TAS-BERT-SW-TO achieves the best results. When target extraction is involved, namely in the

Table 3: Comparison results between our proposed joint model and the separate models. C1 refers to the micro-F1 score (in percent) on the full dataset; C2 refers to the micro-F1 score (in percent) on the partial dataset consisting of opinions with implicitly targets only; C3 refers to the ratio (in percent) of occurrences of all-O sequences in case "*no*" is predicted.

Method		Res15			Res16		
		C1	C2	C3	C1	C2	C3
TAS-BERT-LPM-BIO-CRF	separate	52.03	40.00	98.84	60.33	41.95	**99.03**
	joint	**54.76**	**43.21**	**98.86**	**64.66**	**49.63**	99.00
TAS-BERT-LPM-TO	separate	51.45	40.64	98.86	61.71	45.96	98.98
	joint	**55.47**	**44.50**	**98.96**	**62.29**	**49.52**	**99.08**
TAS-BERT-SW-BIO-CRF	separate	55.24	43.45	98.86	62.42	44.83	**99.11**
	joint	**57.51**	**45.75**	**98.96**	**65.89**	**50.59**	99.08
TAS-BERT-SW-TO	separate	53.92	46.63	98.97	62.90	45.78	99.07
	joint	**58.09**	**51.33**	**99.03**	**65.44**	**51.61**	**99.16**

TD, TSD, TAD and TASD tasks, TAS-BERT-SW-BIO-CRF almost achieves the best results except in two cases on Res15. We conjecture that it is due to the fact that the proportion of opinions with implicit targets on Res15 is larger than that on Res16. It is likely that the combination of the TO tagging scheme and the softmax decoder outperforms the combination of the BIO tagging scheme and the CRF decoder in handling opinions with implicit targets.

Ablation Study

A straightforward weakened variant of our method can be obtained by separately training two models for the first subproblem (namely "*yes/no*" classification) and for the second subproblem (namely sequence labeling), respectively, and by combining the results output by the two models to form opinions, where the first model minimizes $loss_g$ defined in Equation (4) and the second model minimizes $loss_h$ defined in Equation (5) or (6) for all training tuples.

We compared our method with this weakened variant on

Table 4: The results of our method *vs.* the results of two best existing methods (DOER for TSD task, BERT-pair-NLI-B for ASD task). Underlined sentiments in are incorrect while words in italic outline the reason why no relevant prediction is made.

Text	Gold	Method	Prediction	Type
lobster was good, nothing spectacular.	{lobster, FOOD#QUALITY, neutral}	DOER	{lobster, <u>positive</u>}	Error-3
		BERT-pair-NLI-B	{FOOD#QUALITY, <u>negative</u>}	Error-2
		Our method	{lobster, FOOD#QUALITY, neutral}	Correct
In fact, many want to return a second time during their visit.	{NULL, RESTAURANT#GENERAL, positive}	DOER	*unable to find opinions with implicit targets*	Error-1
		BERT-pair-NLI-B	{RESTAURANT#GENERAL, <u>negative</u>}	Error-2
		Our method	{NULL, RESTAURANT#GENERAL, positive}	Correct
Always busy, but they are good at seating you promptly and have quick service.	{service, SERVICE#GENERAL, positive}	DOER	{service, positive} *unable to find opinions with implicit targets*	Correct Error-1
	{NULL, SERVICE#GENERAL, positive}	BERT-pair-NLI-B	{SERVICE#GENERAL, positive} *removed due to duplication*	Correct Error-2
		Our method	{service, SERVICE#GENERAL, positive} *removed due to conflicts between targets*	Correct Error-4

four different settings that correspond to the aforementioned four variants of our method. The comparison results are reported in Table 3. It can be seen that, our method significantly outperforms the weakened variant on both the full datasets and the partial datasets that consist of opinions with implicit targets only. In addition, for both methods the ratio of occurrences of all-O sequences in case "*no*" is predicted is always very high, and the two methods are comparable in this metric. These results show that the joint training framework in our method is superior to the traditional separate training mechanism. They further demonstrate the effectiveness of exploiting the dependence relationship between targets and aspects in target-aspect-sentiment detection.

Case Study

We observed four main error types. The type Error-1 is the ignorance of opinions with implicit targets. The type Error-2 is the ignorance of targets. The type Error-3 is the ignorance of aspects. The type Error-4 is the conflict between different targets which happens when the aspect and the sentiment are the same and either the targets are overlapped or one of multiple targets is NULL. We select three representative examples to compare our method with two state-of-the-art methods DOER and BERT-pair-NLI-B, as shown in Table 4.

As for the first example, the sentiment should be predicted based on both the target and the aspect. However, DOER only gives a prediction of sentiments based on the target "lobster" whereas BERT-pair-NLI-B just takes aspect "FOOD#QUALITY" into consideration. This ignorance of targets or aspects leads to false sentiment prediction in DOER and BERT-pair-NLI-B. In contrast, our method can capture the dual dependence of sentiments on both targets and aspects and predict the correct sentiments.

When it comes to the second example, DOER is unable to find opinions with implicit targets, while BERT-pair-NLI-B is unable to give correct prediction due to ignorance of targets. Our method, in contrast, can deal with the opinion with implicit target in a correct way.

As for the third example, "service" and implicit target appear simultaneously. DOER just provides one prediction since it is unable to handle implicit targets. When "service" and implicit target refer to the same aspect

"SERVICE#GENERAL", BERT-pair-NLI-B gives only one prediction, which is not completely correct either. Since a conflict between targets happens when the aspect and the sentiment are the same while one of multiple targets is NULL, our method fails to detect the opinion with implicit target. This failure is caused by data construction in our method. There should exist optimizations to tackle this kind of failure cases but they are beyond the scope of this paper.

Conclusions and Future Work

In order to capture the dual dependence of sentiments on both targets and aspects and to handle implicit target cases in aspect-based sentiment analysis, we have proposed a novel method for target-aspect-sentiment joint detection. It reduces the joint detection problem to binary text classification problems and sequence labeling problems, solved by the proposed TAS-BERT model which is built upon the pre-trained language model BERT. Experimental results on the SemEval-2015 and SemEval-2016 restaurant datasets demonstrate that our method achieves a high performance in detecting target-aspect-sentiment triples even for the implicit target cases and outperforms the state-of-the-art methods for five subtasks of target-aspect-sentiment detection.

For future work, we will enhance the proposed method by optimizing data construction and further improve the performance in joint detection.

Acknowledgments

References

Brun, C., and Nikoulina, V. 2018. Aspect based sentiment analysis into the wild. In *WASSA@EMNLP*, 116–122.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, 4171–4186.

He, R.; Lee, W. S.; Ng, H. T.; and Dahlmeier, D. 2017. An unsupervised neural attention model for aspect extraction. In *ACL*, 388–397.

Hu, M.; Peng, Y.; Huang, Z.; Li, D.; and Lv, Y. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *ACL*, 537–546.

Jakob, N., and Gurevych, I. 2010. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *EMNLP*, 1035–1045.

Kiritchenko, S.; Zhu, X.; Cherry, C.; and Mohammad, S. 2014. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *SemEval@COLING*, 437–442.

Li, X., and Lam, W. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *EMNLP*, 2886–2892.

Li, X.; Bing, L.; Li, P.; Lam, W.; and Yang, Z. 2018. Aspect term extraction with history attention and selective transformation. In *IJCAI*, 4194–4200.

Li, X.; Bing, L.; Li, P.; and Lam, W. 2019. A unified model for opinion target extraction and target sentiment prediction. In *AAAI*, 6714–6721.

Liu, F.; Cohn, T.; and Baldwin, T. 2018. Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis. In *NAACL-HLT*, 278–283.

Liu, K.; Xu, L.; and Zhao, J. 2013. Syntactic patterns versus word alignment: Extracting opinion targets from online reviews. In *ACL*, 1754–1763.

Luo, H.; Li, T.; Liu, B.; Wang, B.; and Unger, H. 2019a. Improving aspect term extraction with bidirectional dependency tree representation. *IEEE/ACM Trans. Audio, Speech & Language Processing* 27(7):1201–1212.

Luo, H.; Li, T.; Liu, B.; and Zhang, J. 2019b. DOER: dual cross-shared RNN for aspect term-polarity co-extraction. In *ACL*, 591–601.

Ma, X., and Hovy, E. H. 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *ACL*, 1064–1074.

Ma, D.; Li, S.; and Wang, H. 2018. Joint learning for targeted sentiment analysis. In *EMNLP*, 4737–4742.

Ma, Y.; Peng, H.; and Cambria, E. 2018. Targeted aspect-based sentiment analysis via embedding common-sense knowledge into an attentive LSTM. In *AAAI*, 5876–5883.

Mitchell, M.; Aguilar, J.; Wilson, T.; and Durme, B. V. 2013. Open domain targeted sentiment. In *EMNLP*, 1643–1654.

Movahedi, S.; Ghadery, E.; Faili, H.; and Shakery, A. 2019. Aspect category detection via topic-attention network. *CoRR* abs/1901.01183.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; and Androutsopoulos, I. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *SemEval@NAACL-HLT*, 486–495.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; Al-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; Clercq, O. D.; Hoste, V.; Apidianaki, M.; Tannier, X.; Loukachevitch, N. V.; Kotelnikov, E. V.; Bel, N.; Zafra, S. M. J.; and Eryigit, G. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *SemEval@NAACL-HLT*, 19–30.

Schmitt, M.; Steinheber, S.; Schreiber, K.; and Roth, B. 2018. Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In *EMNLP*, 1109–1114.

Sun, C.; Huang, L.; and Qiu, X. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *NAACL-HLT*, 380–385.

Wang, Y.; Huang, M.; Zhu, X.; and Zhao, L. 2016. Attention-based LSTM for aspect-level sentiment classification. In *EMNLP*, 606–615.

Wang, W.; Pan, S. J.; Dahlmeier, D.; and Xiao, X. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *AAAI*, 3316–3322.

Wang, F.; Lan, M.; and Wang, W. 2018. Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multi-task learning. In *IJCNN*, 1–8.

Xenos, D.; Theodorakakos, P.; Pavlopoulos, J.; Malakasiotis, P.; and Androutsopoulos, I. 2016. AUEB-ABSA at semeval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In *SemEval@NAACL-HLT*, 312–317.

Xu, H.; Liu, B.; Shu, L.; and Yu, P. S. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction. In *ACL*, 592–598. Association for Computational Linguistics.

Xu, H.; Liu, B.; Shu, L.; and Yu, P. S. 2019. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *NAACL-HLT*, 2324–2335.

Xue, W., and Li, T. 2018. Aspect based sentiment analysis with gated convolutional networks. In *ACL*, 2514–2523.

Xue, W.; Zhou, W.; Li, T.; and Wang, Q. 2017. MTNA: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In *IJCNLP*, 151–156.

Yin, Y.; Wei, F.; Dong, L.; Xu, K.; Zhang, M.; and Zhou, M. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. In *IJCAI*, 2979–2985.

Zeng, J.; Ma, X.; and Zhou, K. 2019. Enhancing attention-based LSTM with position context for aspect-level sentiment classification. *IEEE Access* 7:20462–20471.

Zhang, M.; Zhang, Y.; and Vo, D. 2015. Neural networks for open domain targeted sentiment. In *EMNLP*, 612–621.