Making Flexible Use of Sub-tasks: A Multiplex Interaction Network for Unified Aspect-based Sentiment Analysis

Guoxin Yu♠♡, Xiang Ao♠♡∗, Ling Luo♠♡, Min Yang♣, Xiaofei Sun◇, Jiwei Li◇, Qing He♠♡

♠Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing 100190, China.
♡University of Chinese Academy of Sciences, Beijing 100049, China.
♣Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
◇Shannon.AI, China.
{yuguoxin20g,aoxiang,luoling18s,heqing}@ict.ac.cn
min.yang@siat.ac.cn
{xiaofei_sun,jiwei_li}@shannonai.com

Abstract

Aspect Term Extraction (ATE), Opinion Term Extraction (OTE) and Aspect Sentiment Classification (ASC) are the essential building blocks of Aspect-based Sentiment Analysis (ABSA). They are typically treated as separate tasks and are individually studied by previous work. Recent studies intend to incorporate multiple sub-tasks into a unified framework, but suffer from the following major disadvantages: (1) ABSA models are extremely fragile when some sub-tasks are absent; (2) the interactive relations among sub-tasks are not adequate. To this end, we propose a multi-task learning approach named MIN (Multiplex Interaction Network) to make flexible use of sub-tasks for a unified ABSA. We divide the sub-tasks of ABSA into extractive sub-tasks and classification sub-tasks, and optimize these sub-tasks in a unified manner with multiplex interaction mechanisms. Specifically, we devise a pairwise attention to exploit bidirectional interactions between any arbitrary pair of extractive sub-tasks and a consistency-weighting to perform unidirectional interaction from an extractive sub-task to a classification sub-task. Since the proposed interaction mechanisms are task-agnostic, our model can also work well when some specific sub-tasks are absent. Extensive experiments on two widely used benchmarks with different numbers of sub-tasks demonstrate the superiority of the proposed model.

1 Introduction

Aspect-based sentiment analysis (ABSA), a fine-grained task of text sentiment analysis (Liu, 2012), aims at summarizing opinions in user comments towards different targets (also known as *aspects*). ABSA generally consists of three specific sub-tasks, namely, aspect term extraction (ATE), opinion term extraction (OTE) and aspect sentiment classification (ASC). For example, consider the sentence *"The price is reasonable although the service is inferior."*, ATE aims to extract a set of aspect terms from the sentence, i.e. {*price, service*}, OTE extracts the opinion words, i.e. {*reasonable, inferior*} , and ASC predicts sentiment polarity for each aspect that is *positive* over the first aspect *price* and *negative* for the second aspect *service*, respectively.

Prevailing solutions of ABSA treated ATE (Liu et al., 2015; Li and Lam, 2017; Angelidis and Lapata, 2018; Liao et al., 2019; Luo et al., 2019b; Ma et al., 2019), OTE (Wang et al., 2017; Wang and Pan, 2019) and ASC (Wang et al., 2016b; Chen et al., 2017; He et al., 2018; Li et al., 2018b; Du et al., 2019; Xu et al., 2021) as separate tasks and were individually studied for decades. These separate tasks need to be integrated into a pipeline for practical use (Hu et al., 2019; Phan and Ogunbona, 2020). The key problem with pipeline approaches is that errors can accumulate and that the pipeline model fails to fully exploit the interactive relations among different sub-tasks (He et al., 2019). Some recent efforts have been proposed to remedy these issues by using joint learning to enhance the interactions among sub-tasks (Wang et al., 2018; Li et al., 2019; He et al., 2019; Luo et al., 2019a; Chen and Qian, 2020; Peng et al., 2020) and achieved better performance than pipeline solutions. To name some, (Li et al., 2019) incorporated ATE and ASC and formulated the problem as a single sequence

---

Sentence	Food	is	pretty	good	but	the	orange	juice	is	horrific	.
ATE	B	O	O	O	O	O	B	I	O	O	O
OTE	O	O	O	B	O	O	O	O	O	B	O
ASC	POS	-	-	-	-	-	NEG	NEG	-	-	-

Table 1: A training instance sentence with gold ATE, OTE and ASC labels.

labeling task with a collapsed tagging scheme[1]. (Luo et al., 2019a) treated ATE and ASC as two sequence labeling problems to mitigate the confused representation brought by the collapsed tags. More recently, (He et al., 2019; Chen and Qian, 2020; Peng et al., 2020) proposed to utilize OTE as an auxiliary task and further facilitated performance.

Despite their success, existing solutions severely suffer from the following disadvantages: firstly, the approaches fusing ATE and ASC cannot benefit from OTE even though there exists opinion term annotations; secondly, the joint learning model might not work when any of the sub-tasks is absent; thirdly, the interactive relations among sub-tasks are not appropriately explored. For instance, (Luo et al., 2019a; Li et al., 2019) would fail to exploit OTE even if these exists labeled data for training. (Chen and Qian, 2020; Peng et al., 2020) might become fragile when the sub-task of OTE or ATE is absent, due to their special designed complex interactions among the sub-tasks of ATE, OTE and ASC. In a nutshell, the key reason for these disadvantages is that existing studies only consider specific interactive relations among a fixed number of sub-tasks, but are inadequate to use the available sub-tasks flexibly.

In this paper, we propose a simple yet effective neural method named MIN (Multiplex Interaction Network) that can make flexible use of sub-tasks for unified ABSA task. The flexibility comes in two folds: First, the number of sub-tasks is not restricted and any related sub-task can be integrated into the unified framework. Second, the interactive relations among distinct sub-tasks are flexibly modeled in an explicit manner.

To tackle the major challenge of how to manage ABSA with arbitrary number of sub-tasks and allow sub-tasks to share interactive information in a unified learning manner, we divide the sub-tasks of ABSA into two categories, namely extractive sub-tasks (e.g. ATE and OTE) and classification sub-tasks (e.g. ASC), and formulate each sub-task as

a sequence labeling problem. MIN is built upon a multi-task learning framework of the shared-private scheme (Collobert and Weston, 2008). Sub-tasks first jointly extract the low-level shared features using multi-layer CNN encoder/pre-trained models. Then the private features of each extractive sub-task are independently learned by a multi-layer CNN network with a pairwise attention mechanism. Such *pair-attention* can capture bidirectional interactions between any two extractive sub-tasks accordingly. Moreover, we devise a *consistency-weighting* mechanism to exploit unidirectional interactions from extractive sub-tasks to classification sub-tasks and assist in private feature extractions in classification sub-tasks. The proposed interaction mechanisms (including pair-attention and consistency-weighting) give the model the flexibility to handle different ABSA situations with different number of sub-tasks of different types. In addition, MIN adopts an information feedback mechanism that first aggregates information from all available tasks and then propagates useful information back to individual tasks, allowing different sub-tasks to positively influence each other. Experiments on two widely used ABSA benchmarks using different numbers of sub-tasks demonstrate the proposed MIN outperforms both pipeline and joint learning SOTA baselines on various settings.

2 Related Work

Separate learning mainly focuses on one of the sub-tasks in ABSA, e.g. aspect term extraction (ATE), opinion term extraction (OTE) or aspect sentiment classification (ASC). For the ATE task, previous methods can be divided into unsupervised and supervised models, respectively. In unsupervised methods, (Liu et al., 2015) utilized rules about grammar dependency relations between opinion words and aspects for aspect terms extraction. (Luo et al., 2019b) presented a neural framework that leverages sememes to enhance lexical semantics for long-tailed aspect extraction. (Liao et al., 2019) utilized the capability of coupling global and local representation to discover aspect terms. For supervised methods (Wang et al., 2016b; Li

---

[1]{B, I, E}-{POS, NEU, NEG} indicates the beginning, inside and end of an aspect-term with positive, neutral or negative sentiment, respectively. O denotes NULL sentiment.

and Lam, 2017; Angelidis and Lapata, 2018; Ma et al., 2019), they were usually formulated as a sequence tagging problem, and various neutral networks with attention mechanisms were proposed to solve the task. For ASC, (Liu and Zhang, 2017; Cheng et al., 2017; He et al., 2018; Tang et al., 2019; Liang et al., 2019; Lei et al., 2019) attempted to exploit contextual and positional proximity of aspect terms for prediction by attentional neural networks. And (Tian et al., 2020) proposed to learn a unified sentiment representation for different sentiment analysis tasks. Recently, capsule network (Du et al., 2019; Chen and Qian, 2019), and graph convolution networks (Zhang et al., 2019) were also utilized in ASC and achieved SOTA performance. These separate learning approaches may have disadvantages in practical applications as they need to be pipelined and the interactions between different sub-tasks are totally neglected.

Joint learning strives to combine sub-tasks of ABSA into a unified learning process. For example, some studies proposed to handle ATE and ASC in a pipeline or an integrated model. The pipeline models (Hu et al., 2019; Phan and Ogunbona, 2020) are extract-then-classify processes and were proposed to solve the two tasks successively. However, they can still derive error accumulations. For integrated models, (Wang et al., 2018; Li et al., 2019) solved ATE and ASC by collapsed tagging that is a unified tagging scheme to link the two tasks. (Luo et al., 2019a) considered the relationship between the two tasks and attempted to investigate useful information from one task to another. Some works (Wang et al., 2017; Dai and Song, 2019; Chen et al., 2020; Zhao et al., 2020) integrated ATE and OTE in the same framework to illustrate these two tasks can benefit from each other. Then emerging methods (He et al., 2019; Chen and Qian, 2020; Peng et al., 2020) proposed to inject OTE as an auxiliary task to further improve the performance of ABSA. However, the number of sub-tasks and interactions among them in existing integrated methods are fixed, which can be restricted when sub-tasks vary in practice.

3 The Multiplex Interaction Network

3.1 Task Definition

All the sub-tasks related to ABSA are categorized into extractive and classification sub-tasks, respectively, in our framework. The extractive sub-tasks extract meaningful subsequences of sen-

tences, such as aspect terms and opinion terms, etc. The classification sub-tasks classify parts of sentences into different classes, e.g. sentiment polarities, genres and etc. We formulate all sub-tasks as sequence labeling problems in our joint learning framework. Given a sequence of tokens $X = \{x_1, x_2, ..., x_n\}$,

- An extractive sub-task $\mathcal{T}$ is to predict a sequence tag $Y^{\mathcal{T}} = \{y_1^{\mathcal{T}}, y_2^{\mathcal{T}}, ..., y_n^{\mathcal{T}}\}$ for the sentence $X = \{x_1, x_2, ..., x_n\}$, where $y_i^{\mathcal{T}} \in \{B, I, O\}$. Specifically, $\{B, I, O\}$ denotes beginning of, inside of and out of the extracted targets. For example, the first and the second rows of Table 1 demonstrate the sequence tags of aspect term extraction (ATE) and opinion term extraction (OTE) respectively. In the example, "food" and "orange juice" are annotated as aspect terms, and "good" and "horrific" are labeled as opinion terms.

- A classification sub-task $\mathcal{C}$ is to predict a sequence tag $Y^{\mathcal{C}} = \{y_1^{\mathcal{C}}, y_2^{\mathcal{C}}, ..., y_n^{\mathcal{C}}\}$ for the sentence $X = \{x_1, x_2, ..., x_n\}$, where $y_i^{\mathcal{C}}$ is one of the class labels. $y_i^{\mathcal{C}}$ will be marked as "-" if the token x_i does not belong to any category. For example, the last row of Table 1 demonstrates the labels of the aspect sentiment classification (ASC) task for the example sentence. In this example, the "food" is labeled as "POS", and "orange" as well as "juice" is labeled as "NEG", according to the ground truth sentiment labels of the two aspect terms. Other tokens are labeled as "-" as they do not have golden labels in this task.

3.2 Model Overview

For the convenience of explanation, we take two extractive sub-tasks, namely ATE and OTE, and a classification sub-task ASC to describe our MIN approach hereafter. The overall architecture of MIN is shown in Figure 1. MIN consists of a shared encode layer, several features extraction modules for different sub-tasks, and an information feedback mechanism.

Private features extraction modules accept the same representation from a shared encode layer then extract their private task-oriented features by specific multi-layer CNN algorithms. In order to enhance flexibility, MIN conducts task interactions for two facets. Firstly, we exploit the bidirectional interaction between two different extractive sub-tasks by a proposed *pair-attention*. Using pair-
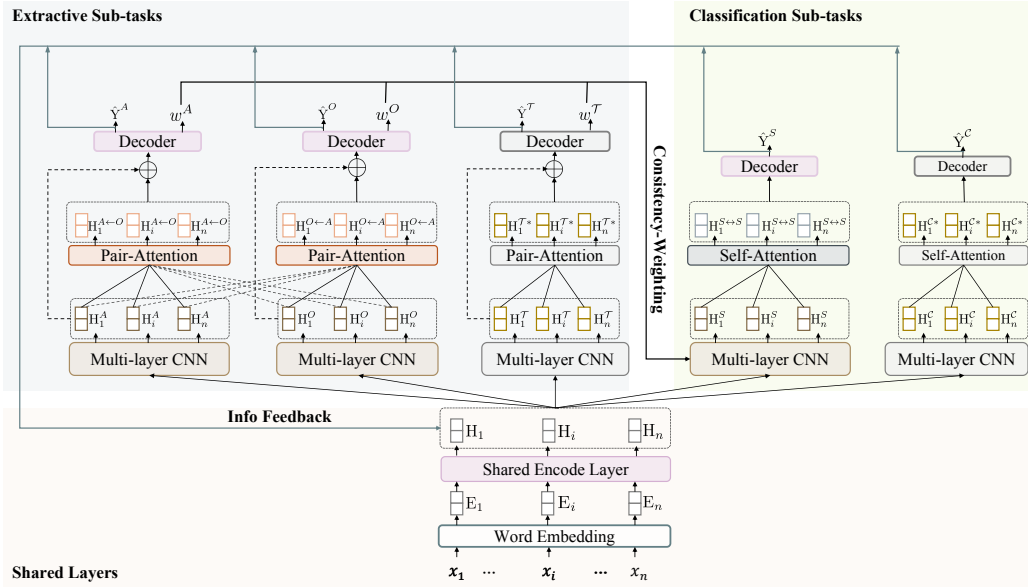
Figure 1: Model Overview of MIN. ATE, OTE and ASC are shown with corresponding interactions. A possible extractive sub-task and a possible classification sub-task are marked in grey colors and some details of them are omitted for easy viewing.

attention, extractive sub-tasks can exchange helpful clues with each other during features extraction. Secondly, we use a *consistency-weighting* mechanism to perform the unidirectional interaction from an extractive sub-task to a classification sub-task. In this way, features with highlight information of extractive sub-tasks will be passed to other classification sub-tasks and assist in their features extractions. Then every sub-task predicts the corresponding sequence labels, i.e., $\hat{Y}^A$, $\hat{Y}^O$ and $\hat{Y}^S$, by its decode layer. The model also adopts an information feedback mechanism that concatenates representations of all sub-tasks to fine-tune the shared representations. In the following, we first describe the MIN model in more detail and then illustrate the learning process.

3.3 Shared Representation Generation

For a sequence of tokens $\{x_1, x_2, ..., x_n\}$, we map the word sequence with either pre-trained double word embeddings (Xu et al., 2018) or pre-trained Bert model to generate a sequence of word vectors $E = \{E_1, ..., E_i, ..., E_n\} \in \mathbb{R}^{d_e \times n}$, where n denotes the sentence length and d_e denotes the dimension of word vectors. Then we simply feed E into a multi-layer CNN encoder to generate shared features $H = \{H_1, ..., H_i, ..., H_n\} \in \mathbb{R}^{d_h \times n}$, where d_h denotes the dimension of hidden vectors.

3.4 Features Extraction for Extractive Sub-tasks

3.4.1 Multi-layer CNN

For the extractive sub-tasks, we use a multi-layer CNN structure proposed by (Xu et al., 2018) to learn private features of each task separately. Specifically, there are many 1D-convolution filters in each CNN layer, and each filter has a fixed kernel size of $k = 2c + 1$. As a result, each filter performs convolution operation on a window of k word representations, and compute the representation for the i-th word along with $2c$ nearby words in its context. We can extract private features H^A of ATE and H^O of OTE by the above multi-layer CNN algorithm,

$$
\begin{aligned}
H^A = MC(H), H^A \in \mathbb{R}^{d_p \times n}, \\
H^O = MC(H), H^O \in \mathbb{R}^{d_p \times n},
\end{aligned}
\tag{1}
$$

where MC indicates the multi-layer CNN algorithm, d_p denotes the dimension of the private features for extractive sub-tasks.

3.4.2 Pair-attention

In order to exploit potential interactions, we develop a pair-attention to exchange helpful clues between ATE and OTE based on their semantic correlations. For ATE, the pair-attention matrix

$A^{A \leftarrow O} \in \mathbb{R}^{n \times n}$ is computed as follows:

$$score_{ij}^{(i \neq j)} = (H_i^A)^\top H_j^O,$$
$$A_{ij}^{A \leftarrow O} = \frac{\exp(score_{ij})}{\sum_{k=1}^n \exp(score_{ik})}. \quad (2)$$

By applying a weighted sum operation of pair-attention to the private features H^O of OTE, we get an interactive feature matrix $H^{A \leftarrow O}$ for ATE:

$$H_i^{A \leftarrow O} = \sum_{j=1}^n (A_{ij}^{A \leftarrow O} \cdot H_j^O). \quad (3)$$

Similarly, according to Eq. 2, we exchange H^A and H^O to compute pair-attention matrix $A^{O \leftarrow A} \in \mathbb{R}^{n \times n}$ then use $A^{O \leftarrow A}$ and H^A to get interactive feature matrix $H^{O \leftarrow A}$ for OTE by Eq. 3. In this way, the model exploits bidirectional interactions between the two sub-tasks. After pair-attention, we concatenate the private and the interactive features of the same sub-task, and then feed them to a fully-connected layer to predict the sequence tags as follows:

$$\hat{Y}^A = \text{softmax}(W^A(H^A \oplus H^{A \leftarrow O})),$$
$$\hat{Y}^O = \text{softmax}(W^O(H^O \oplus H^{O \leftarrow A})). \quad (4)$$

Note that for other extractive sub-tasks, we can also compute the pair-attention with related extractive sub-tasks in a similar manner.

3.5 Features Extraction for Classification Sub-tasks

Recall that a classification sub-task may benefit from extractive sub-tasks. To name some, knowing the location of aspect terms will enhance the matching degree of aspect terms with corresponding sentiment polarities. For example, the second row and the last row of Table 1 have consistent labels[2] since we expect to perform sentiment prediction for aspect terms than other words. Similarly, the sentiment polarities of aspect terms are mainly influenced by surrounding opinion words. For example, we can find "NEG" labels near to the opinion word "horrific" in the example sentence shown in Table 1.

3.5.1 Consistency-weighting

Inspired by the above observations, we devise a consistency-weighting mechanism to learn private features of ASC task.

---

[2]"Food" is the beginning of an aspect term and it has a positive sentiment; while "pretty" is out of an aspect term and thus has no sentiment in ATE and ASC sub-tasks.

First, we compute a set of weights to pass information with salient aspect and opinion terms to ASC. Specifically, the weights of a token i w.r.t. aspect and opinion terms are computed as follows:

$$w_i^A = \sum_{j=1}^k (\alpha - \beta \cdot |a_j - i|),$$
$$w_i^O = \sum_{j=1}^k (\alpha' - \beta' \cdot |o_j - i|) \quad (5)$$

where k denotes the first k tokens most likely to be aspect or opinion terms, and a_j and o_j are the indexes of j-th aspect and opinion term in the sentence. α, α', β and β' are pre-specified constants. We can simply multiply the weights w^A and w^O with interactive features of ATE and OTE to get a distilled representation of every token:

$$H_i^{S \leftarrow A} = w_i^A \cdot H_i^A, H^{S \leftarrow A} \in \mathbb{R}^{d_p \times n},$$
$$H_i^{S \leftarrow O} = w_i^O \cdot H_i^O, H^{S \leftarrow A} \in \mathbb{R}^{d_p \times n}. \quad (6)$$

Then the shared features are concatenated with such distilled representation, and the private features of ASC are extracted by the same multi-layer CNN algorithm as in the ATE and OTE sub-tasks.

$$H^S = MC(H^{S \leftarrow A} \oplus H^{S \leftarrow O} \oplus H),$$
$$H^S \in \mathbb{R}^{d_p \times n}. \quad (7)$$

By consistency-weighting, the information of aspect and opinion terms are highlighted during the features extraction for the classification sub-task.

3.5.2 Self-Attention

We adopt a self-attention mechanism to learn long-distant dependent information in a sentence for the classification sub-task. Specifically, the self-attention matrix $A^{S \leftrightarrow S}$ is computed as follows:

$$score_{ij}^{(i \neq j)} = ((H_i^S)^\top W^{S \leftrightarrow S} H_j^S) \cdot \frac{1}{|i - j|},$$
$$A_{ij}^{S \leftrightarrow S} = \frac{\exp(score_{ij})}{\sum_{k=1}^n \exp(score_{ik})}, A^{S \leftrightarrow S} \in \mathbb{R}^{n \times n}, \quad (8)$$

where $H_i^S W^{S \leftrightarrow S}(H_j^S)^T$ and $\frac{1}{|i-j|}$ denote the semantic relevance and distance between H_i and H_j, respectively. After getting the outputs of the self-attention layer, we also exploit a fully-connected layer with softmax activation as a decoder:

$$H_i^{S \leftrightarrow S} = \sum_{j=1}^n (A_{ij}^{S \leftrightarrow S} \cdot H_j^S),$$
$$\hat{Y}^S = \text{softmax}(W^S H^{S \leftrightarrow S}), \quad (9)$$

where $\hat{Y}^S$ is the predicted label sequence for ASC.

Dataset		Sentence	Opinion	Aspect		
				Pos	Neu	Neg
Res14	Train	3,044	3,484	2,164	807	637
	Test	800	1,008	728	196	196
Lap14	Train	3,048	2,504	994	870	464
	Test	800	674	341	128	169

Table 2: The statistics of datasets.

Model	Res14			Lap14		
	AE-F1	AS-F1	O-F1	AE-F1	AS-F1	O-F1
DECNN-TNet*	82.79	70.45	65.80	79.38	68.69	57.39
DECNN-TCaps*	82.79	71.77	66.84	79.38	69.61	57.71
PIPELINE-MIN	84.00	71.75	68.36	78.43	71.45	59.19
MNN*	85.84	67.93	–	79.91	58.30	–
E2E-ABSA*	83.92	68.38	66.60	77.34	68.24	55.88
DOER*	84.63	64.50	68.55	80.21	60.18	56.71
MIN	84.80	73.91	69.57	79.94	71.57	60.39

Table 3: Comparison results for double-task methods. We divided all methods into pipeline and unified models. The best scores are in bold face and the second best ones are underlined. "*" denotes the results are taken from their own papers or (Chen and Qian, 2020).

3.6 Information Feedback Mechanism

MIN exploits an information feedback mechanism to interactively fine-tune the shared feature representations. We denote t as the times of the information feedback. In each round, we utilize the hidden representation before decoder in every sub-task to update the shared features. Specifically, we concatenate the hidden representations with the shared feature matrix, and then use a fully-connected layer with softmax activation to squeeze the vectors to a fixed dimension.

$$\mathrm{H}_t = \mathrm{softmax}(\mathrm{W}(\mathrm{H}_{t-1}^{A \leftarrow O} \oplus \mathrm{H}_{t-1}^{O \leftarrow A} \oplus \mathrm{H}_{t-1}^{S \leftrightarrow S} \oplus \mathrm{H})), \quad (10)$$

where t denotes the t-th feedback update.

3.7 Training Procedure

The overall loss $\mathcal{L}$ consists of cross-entropy losses for all sub-tasks according to final predicted sequence labels.

$$\mathcal{L}^A = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij}^A \cdot \log(\hat{y}_{ij}^A)),$$

$$\mathcal{L}^O = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij}^O \cdot \log(\hat{y}_{ij}^O)), \quad (11)$$

$$\mathcal{L}^S = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij}^S \cdot \log(\hat{y}_{ij}^S)),$$

where N denotes the number of training instances, n_i denotes the number of tokens in the i-th instance. The overall loss is the weighted sum of the sub-tasks' losses:

$$\mathcal{L} = a \cdot \mathcal{L}^A + b \cdot \mathcal{L}^O + c \cdot \mathcal{L}^S, \quad (12)$$

where a, b, c are task coefficients, and we set all of them as 1. Following (He et al., 2019), during model training, we only compute ASC loss on tokens which are related to aspect terms, i.e., if a token is not aspect term, we will ignore predicted the sentiment on it. We minimize the $\mathcal{L}$ and determine a suitable t by grid search for information feedback mechanism during the experiment.

4 Experiments

4.1 Datasets

We adopt two widely used datasets from SemEval2014 Task 4 (Pontiki et al., 2014) in our experiment which contain reviews about restaurant and laptop, and the statistics are shown in Table 2. Note that original datasets have ground truth labels for aspect terms and corresponding sentiment polarities, while labels for opinion terms are annotated by (Wang et al., 2016a).

4.2 Compared Methods

We compare our method with recent ABSA methods. To demonstrate the flexibility of our method, we divide the baselines into two categories, namely double-task (ATE and ASC) and triple-task (ATE, OTE and ASC) methods, considering that existing methods are usually designed for fixed sub-tasks.

Double-task methods contain (1) pipeline methods: we take top-performing method **DECNN** (Xu et al., 2018) for ATE, **TNet** (Li et al., 2018a) and **TCaps** (Chen and Qian, 2019) for ASC to construct two pipeline models following (Chen and Qian, 2020). We conduct a pipeline setting of our proposed MIN denoted as **PIPELINE-MIN**, which trains ATE and ASC separately, for fair comparison. (2) unified methods: **MNN** (Wang et al., 2018) and **E2E-ABSA** (Li et al., 2019) jointly solve ATE and ASC by using collapsed tagging schemes, and **DOER** (Luo et al., 2019a) treats these two tasks as two sequence labeling problems and utilized a dual cross-shared RNN.

Triple-task methods aim to solve ATE, OTE and ASC simultaneously, including (1) pipeline methods: following (Chen and Qian, 2020), we combine **CMLA** (Xu et al., 2018) for ATE, **TNet** and **TCaps** for ASC to construct two pipeline baselines.

OTE is integrated into ATE. (2) unified methods: **IMN** (He et al., 2019) is an interactive multi-task model jointly trained on ATE and ASC where OTE is also integrated into ATE. While **RACL** (Chen and Qian, 2020) is trained on ATE, OTE and ASC in parallel that considers four relations among the three sub-tasks.

4.2.1 Settings

Following (He et al., 2019), we adopt double embedding in word embedding layer of **MIN**, where each word embedding is a concatenation of *general embedding* (Pennington et al., 2014) with 300 dimensions and *domain embedding* (Xu et al., 2018) with 100 dimensions. We set the hidden size $d_e = 400$, $d_h = 300$, $d_p = 300$, and the kernel size, number of shared CNN layers to 5 and 2, individually. And the numbers of information feedback in Res14 and Lap14 are set to 2 and 3 respectively. The layers of multi-layer CNN for ATE, OTE and ASC are set to $\{2, 2, 1\}$ and the kernel size is 5 in two dataset. Adam optimizer (Kingma and Ba, 2015) with a learning rate of 1e-4 and a batch size of 8 are utilized for all datasets.

We also combine **MIN** with BERT$_{Large}$ to get **MIN-BERT** and $\{d_e, d_h, d_p\}$ is set to $\{1024, 300, 300\}$ with a learning rate of $1e-5$ by grid search. The layers of multi-layer CNN for ATE, OTE, ASC and the number of information feedback are set to $\{1, 1, 1, 2\}$ in the two datasets. The other parameters are the same as the **MIN** model with the double embedding. We meanwhile alternate IMN and RACL with BERT$_{Large}$ for further detailed comparison. We use the same metrics as (He et al., 2019), i.e., AE-F1, OE-F1, AS-F1 and O-F1, representing macro F1 scores for ATE, OTE, ASC and overall performance for complete ABSA. We select the model with the best O-F1 on the development set for test.

4.3 Experimental Results

4.3.1 Overall Performance

Table 3 and 4 demonstrate the performance of our method and the compared double-task (i.e., ATE and ASC) and triple-task methods (i.e., ATE, OTE and ASC), respectively. We have several consistent observations from the two tables. Firstly, most unified models perform better than pipeline models, which proves the advantages of exploiting interactions between sub-tasks. Secondly, our MIN can be flexibly applied to either double-task or triple-task scenario and can achieve better performance than

Model	Res14				Lap14			
	AE-F1	OE-F1	AS-F1	O-F1	AE-F1	OE-F1	AS-F1	O-F1
CMLA-TNet*	81.91	83.84	69.69	64.49	77.49	76.06	68.30	55.94
CMLA-TCaps*	81.91	83.84	71.32	65.68	77.49	76.06	69.49	56.30
IMN*	84.01	85.64	71.90	68.32	78.46	78.14	69.92	57.66
RACL*	**85.37**	85.32	74.46	70.67	**81.99**	79.76	71.09	60.63
MIN	85.27	**86.85**	**76.39**	**70.92**	82.24	**80.56**	**72.60**	**61.35**
IMN-BERT	84.06	85.10	75.67	70.72	77.55	81.00	75.56	61.73
RACL-BERT	86.38	**87.18**	**81.61**	75.42	81.79	79.72	73.91	63.40
MIN-BERT	**87.91**	85.66	80.48	**76.02**	**83.22**	**81.80**	74.95	**64.83**

Table 4: Comparison results for triple-task methods. We divided all methods into pipeline, unified models and models based on BERT-large. "*", bolded and underlined scores have the same meanings as Table 3.

ATE+OTE	Res14		Lap14	
	AE-F1	OE-F1	AE-F1	OE-F1
IMN	84.83	86.32	78.31	77.58
RACL	85.47	86.48	**81.83**	78.19
MIN	**85.78**	**86.71**	81.74	**78.29**
OTE+ASC	OE-F1	AS-F1	OE-F1	AS-F1
IMN	NA	NA	NA	NA
RACL	81.27	63.16	79.32	65.45
MIN	**86.14**	**71.15**	**79.58**	**68.13**
ATE+ASC	AE-F1	AS-F1	AE-F1	AS-F1
IMN	84.78	70.46	79.22	69.65
RACL	**85.66**	70.78	79.76	68.87
MIN	84.80	**73.91**	**79.94**	**71.57**

Table 5: Comparison results for different combinations of ATE, OTE and ASC. The best scores are in bold. 'NA' denotes not available.

all the compared methods on O-F1. We conjecture the possible reason is that the proposed generic interaction mechanism among different sub-task types is independent on any specific sub-task. Thus, our MIN can work well regardless of the number or the type of sub-tasks. Thirdly, we enhance the performance of some compared models with word embeddings provided by BERT$_{Large}$ (Devlin et al., 2018) and report the corresponding results in Table 4. We can observe that our MIN-BERT also shows superiority especially on O-F1.

4.3.2 Flexibility

To further investigate the flexibility of MIN, we compare our MIN with variants of strong competitors in triple-task methods by varying the combination of sub-tasks. We remove one of the three sub-tasks and get three different combinations, namely "ATE+OTE", "ATE+ASC" and "OTE+ASC". Note that the double-task methods in table 3 cannot be easily adapted for this comparison due to their special design for specific sub-task combinations.

Table 5 exhibits the corresponding results on two datasets. We can see from the table that MIN

Model	Res14	Lap14
Full Model	70.92	61.35
w/o Pair-attention	69.82	59.37
w/o Consistency-weighting	69.46	60.84
w/o Info Feedback	69.07	59.17
w/o Self-Attention	67.63	57.03

Table 6: Results of ablation study on the model for "ATE+OTE+ASC". 'w/o' denotes without.

(a) Results on Res14. (b) Results on Lap14.
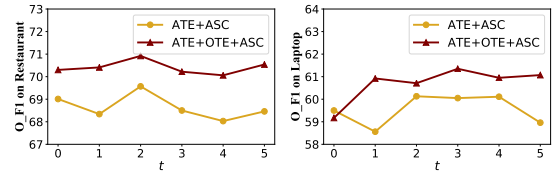
Figure 2: Effects of parameter t.

outperforms both IMN and RACL on most of the cases, and the two baselines cannot work well when the sub-task combination varies since both of them designed specialized interaction mechanisms for a fixed number of sub-tasks. For example, IMN adopts OTE as an auxiliary task for ATE and thus cannot provide results on "OTE+ASC". RACL, which designs four kinds of collaborative relations among ATE, OTE and ASC, degrades when arbitrary one sub-task is unavailable. However, our MIN separates the sub-tasks by different categories and leverages two category-level interaction mechanisms. It derives a trade-off that the interactions on remaining sub-tasks are still functional although one sub-task is removed. That is where the flexibility of our model comes from.

4.3.3 Ablation Test

In order to verify the effectiveness of different interactive mechanisms in our model, we conduct ablation tests on "ATE+OTE+ASC" unified ABSA. We remove different interaction mechanisms between sub-tasks in turn to observe the effectiveness of them. As the O-F1 results shown in Table 6, all interaction mechanisms are effective in MIN as expected. For example, on Res14, the scores of model without consistency-weighting and the model without self-attention drop largely. It is believed that sentiment polarities are mainly influenced by corresponding aspect and opinion words.

4.3.4 Case Study

Next, we illustrate two cases generated by IMN, RACL and our MIN for further investigation in Table 7. In Case (1), IMN extracts "choice" as an extra aspect without considering the relations between ATE and OTE. In addition, RACL incorrectly identifies "choice" and "craving" as aspect and opinion terms. We conjecture the possible reasons might be the "choice" is paired with "exquisite" as an aspect and the "craving" is recognized as a term expressing opinions by the specifically-designed interaction relations in RACL.

However, our proposed pair-attention of MIN is more general, and correctly extracts all the aspect terms and opinion terms. Case (2) explores the effectiveness of our proposed consistency-weighting. Specifically, although IMN extracts the correct opinion word "too sweet", the final sentiment of "mole sauce" is still predicted wrong with a weak help of opinion terms. Meanwhile, although RACL exploits the relation between OTE and ASC, it recognizes the "sweet" incompletely, which might because RACL cannot update representation without the information feedback mechanism. As a result, "sweet" interferes sentiment prediction and leads to a wrong polarity for "mole sauce". In contrast, MIN correctly predicts sentiment polarities by using consistency-weighting based on complete aspect and opinion terms.

4.3.5 Times of Information Feedback

Finally, we investigate the sensitivity of the parameter t in the information feedback mechanism. Figure 2 (a) and (b) show the results of O-F1 on Res14 and Lap14 as t increases. We observe that MIN achieves better performance when $t = 2$ or $t = 3$. And there is no considerable improvement with a further increase of t, because redundant iterations might derive overfitting.

5 Conclusion

In this paper, we propose a novel neural model MIN to make flexible use of sub-tasks for the unified ABSA task. We introduce two sub-task agnostic interaction mechanisms, i.e., pair-attention and consistency-weighting, to exploit interactions among different sub-tasks and an information feedback mechanism to fine-tune the shared feature representation. Experiments with a different number of sub-tasks show the flexibility of MIN. The proposed framework can be potentially applied to similar tasks such as named entity recognition, semantic relation extraction, and etc.

Case	IMN		RACL		MIN	
	ATE,ASC	OTE	ATE,ASC	OTE	ATE,ASC	OTE
(1) If you 're craving some [**serious**] [**indian food**]pos and desire a [**cozy**] [**ambiance**]pos, this is quite and [**exquisite**] choice.	[indian food]pos [ambiance]pos [choice✗]pos	[serious] [cozy] [exquisite]	[indian food]pos [ambiance]pos [choice✗]pos	[craving✗] [serious] [cozy] [exquisite]	[indian food]pos [ambiance]pos	[serious] [cozy] [exquisite]
(2) The [**fajita**]neg we tried was [**tasteless**] and [**burned**] and the [**mole sauce**]neg was way [**too sweet**] .	[fajita]neg [mole sauce]pos✗	[tasteless] [burned] [too sweet]	[fajita]neg [mole sauce]neu✗	[tasteless] [burned] [sweet✗]	[fajita]neg [mole sauce]neg	[tasteless] [burned] [too sweet]

Table 7: Case study. The abbreviations pos, neu and neg in the table represent positive, neutral and negative. ✗ indicates incorrect prediction. Notice that the ground-truth aspect and opinion terms are marked in red and blue in the brackets, while the sentiment polarity labels for the aspect terms are demonstrated as corresponding subscripts.

Acknowledgements

References

Stefanos Angelidis and Mirella Lapata. 2018. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3675–3686, Brussels, Belgium. Association for Computational Linguistics.

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 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, Copenhagen, Denmark. Association for Computational Linguistics.

Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020. Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6515–6524, Online. Association for Computational Linguistics.

Zhuang Chen and Tieyun Qian. 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, Florence, Italy. Association for Computational Linguistics.

Zhuang Chen and Tieyun Qian. 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, Online. Association for Computational Linguistics.

Jiajun Cheng, Shenglin Zhao, Jiani Zhang, Irwin King, Xin Zhang, and Hui Wang. 2017. Aspect-level sentiment classification with heat (hierarchical attention) network. In *CIKM*, pages 97–106.

Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *ICML*, pages 160–167.

Hongliang Dai and Yangqiu Song. 2019. Neural aspect and opinion term extraction with mined rules as weak supervision. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5268–5277, Florence, Italy. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, Jianxin Liao, Tong Xu, and Ming Liu. 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, Hong Kong, China. Association for Computational Linguistics.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *ACL*, pages 504–515.

Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR*.

Zeyang Lei, Yujiu Yang, Min Yang, Wei Zhao, Jun Guo, and Yi Liu. 2019. A human-like semantic cognition network for aspect-level sentiment classification. In *AAAI*, volume 33, pages 6650–6657.

Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018a. Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 946–956, Melbourne, Australia. Association for Computational Linguistics.

Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. A unified model for opinion target extraction and target sentiment prediction. In *AAAI*, volume 33, pages 6714–6721.

Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018b. Aspect term extraction with history attention and selective transformation. In *IJCAI*, pages 4194–4200.

Xin Li and Wai Lam. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2886–2892, Copenhagen, Denmark. Association for Computational Linguistics.

Bin Liang, Jiachen Du, Ruifeng Xu, Binyang Li, and Hejiao Huang. 2019. Context-aware embedding for targeted aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4678–4683, Florence, Italy. Association for Computational Linguistics.

Ming Liao, Jing Li, Haisong Zhang, Lingzhi Wang, Xixin Wu, and Kam-Fai Wong. 2019. Coupling global and local context for unsupervised aspect extraction. 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 4579–4589, Hong Kong, China. Association for Computational Linguistics.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Jiangming Liu and Yue Zhang. 2017. Attention modeling for targeted sentiment. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 572–577, Valencia, Spain. Association for Computational Linguistics.

Qian Liu, Zhiqiang Gao, Bing Liu, and Yuanlin Zhang. 2015. Automated rule selection for aspect extraction in opinion mining. In *IJCAI*.

Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019a. DOER: Dual cross-shared RNN for aspect term-polarity co-extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 591–601, Florence, Italy. Association for Computational Linguistics.

Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019b. Unsupervised neural aspect extraction with sememes. In *IJCAI*, pages 5123–5129.

Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3538–3547, Florence, Italy. Association for Computational Linguistics.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *AAAI*, pages 8600–8607.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 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. Association for Computational Linguistics.

Minh Hieu Phan and Philip O. Ogunbona. 2020. Modelling context and syntactical features for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3211–3220, Online. Association for Computational Linguistics.

Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. *SemEval 2014*, page 27.

Jialong Tang, Ziyao Lu, Jinsong Su, Yubin Ge, Linfeng Song, Le Sun, and Jiebo Luo. 2019. Progressive self-supervised attention learning for aspect-level sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 557–566, Florence, Italy. Association for Computational Linguistics.

Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual*

Meeting of the Association for Computational Linguistics, pages 4067–4076, Online. Association for Computational Linguistics.

Feixiang Wang, Man Lan, and Wenting Wang. 2018. Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multitask learning. In *IJCNN*, pages 1–8.

Wenya Wang and Sinno Jialin Pan. 2019. Transferable interactive memory network for domain adaptation in fine-grained opinion extraction. In *AAAI*, volume 33, pages 7192–7199.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 616–626, Austin, Texas. Association for Computational Linguistics.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *AAAI*.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016b. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.

Chi Xu, Hao Feng, Guo-Xin Yu, Min Yang, X. Wang, Yan Song, and Xiang Ao. 2021. Discovering protagonist of sentiment with aspect reconstructed capsule network. In *DASFAA*.

Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. Double embeddings and CNN-based sequence labeling for aspect extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 592–598, Melbourne, Australia. Association for Computational Linguistics.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *EMNLP*, pages 4560–4570.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. SpanMlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3239–3248, Online. Association for Computational Linguistics.