

# Synchronous Double-channel Recurrent Network for Aspect-Opinion Pair Extraction

Shaowei Chen, Jie Liu\*, Yu Wang, Wenzheng Zhang, Ziming Chi

College of Artificial Intelligence, Nankai University, Tianjin, China

{chenshaowei, yuwang17, wzzhang}@mail.nankai.edu.cn

jliu@nankai.edu.cn

zimingchi@163.com

## Abstract

Opinion entity extraction is a fundamental task in fine-grained opinion mining. Related studies generally extract aspects and/or opinion expressions without recognizing the relations between them. However, the relations are crucial for downstream tasks, including sentiment classification, opinion summarization, etc. In this paper, we explore **Aspect-Opinion Pair Extraction (AOPE)** task, which aims at extracting aspects and opinion expressions in pairs. To deal with this task, we propose **Synchronous Double-channel Recurrent Network (SDRN)** mainly consisting of an opinion entity extraction unit, a relation detection unit, and a synchronization unit. The opinion entity extraction unit and the relation detection unit are developed as two channels to extract opinion entities and relations simultaneously. Furthermore, within the synchronization unit, we design **Entity Synchronization Mechanism (ESM)** and **Relation Synchronization Mechanism (RSM)** to enhance the mutual benefit on the above two channels. To verify the performance of SDRN, we manually build three datasets based on SemEval 2014 and 2015 benchmarks. Extensive experiments demonstrate that SDRN achieves state-of-the-art performances.

## 1 Introduction

Opinion entity extraction, which aims at identifying aspects and/or opinion expressions in review sentences, is an important task in fine-grained opinion mining. Recently, there have been considerable studies focused on this task. Specifically, Liu et al. (2012), Li and Lam (2017) and Li et al. (2018) explored aspect term extraction, and Fan et al. (2019) extracted opinion phrases with given aspects. Meanwhile, many studies dealt with aspect and opinion

\*Corresponding author.

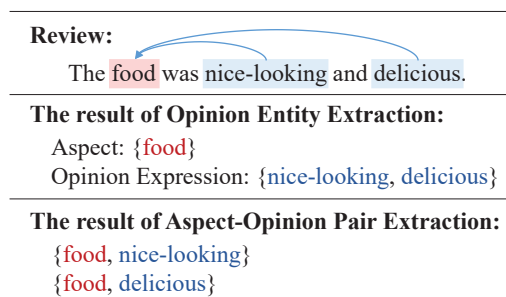


Figure 1: An example of task comparisons. The aspects and the opinion expressions are marked with red and blue, respectively.

term co-extraction (Xu et al., 2013; Liu et al., 2015; Wang et al., 2017; Yu et al., 2019; Wang and Pan, 2019; Dai and Song, 2019). These studies have shown the importance of opinion entity extraction and achieved great progress. However, they neglect to recognize the relations between aspects and opinion expressions.

While aspect-opinion relation detection is one of the key parts of an opinion mining system (Hu and Liu, 2004; Popescu and Etzioni, 2005; Zhuang et al., 2006), it is neglected or assumed given beforehand, which leaves a significant gap to subsequent opinion mining tasks. For instance, as shown in Figure 1, we can obtain the aspect  $\{food\}$  and the opinion expressions  $\{nice-looking, delicious\}$  from opinion entity extraction. Although both *nice-looking* and *delicious* express positive sentiment, they further describe *food* from the appearance and taste perspectives, respectively. Therefore, only with the relations between aspects and opinion expressions, e.g., the pair  $\langle food, delicious \rangle$ , can the more fine-grained subsequent tasks be executed, such as pair-level sentiment classification, pair-level opinion clustering, etc.

To bridge the gap between opinion entity extraction and subsequent tasks, we explore Aspect-Opinion Pair Extraction (AOPE) task, which aims at extracting aspects and opinion expressions along

with their relations. Specially, AOPE is not only necessary for subsequent tasks, but also beneficial to both opinion entity extraction and relation detection. However, the studies on AOPE are very limited. Early works (Hu and Liu, 2004; Zhuang et al., 2006) approach aspect-opinion pair extraction in a pipeline manner by dividing it into two isolated tasks. Yang and Cardie (2013), Klinger and Cimiano (2013b) and Katiyar and Cardie (2016) attempted to extract opinion entities and relations jointly without considering the interaction between opinion entity extraction and relation detection, which limits the performance.

Therefore, AOPE remains a rather challenging task. First, the relational structure of aspects and opinion expressions within a sentence can be complicated, requiring the model to be effective and flexible in detecting relations. For example, the relations can be one-to-many, many-to-one, and even embedded or overlapped. Second, opinion entity extraction and relation detection are not two independent tasks as in other multitask learning problems but rely on each other, hence posing a key challenge on how to fuse and learn the two subtasks properly. Third, how to synchronize opinion entity extraction with relation detection and make them mutually promotion is another primary challenge.

To address the aforementioned challenges, we propose **Synchronous Double-channel Recurrent Network (SDRN)**. Specifically, we first utilize Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) to learn context representations. Then, the double-channel recurrent network, which consists of an opinion entity extraction unit and a relation detection unit, is constructed to extract aspects, opinion expressions, and relations simultaneously. To enable the information interaction between the above two channels, we design a synchronization unit which contains **Entity Synchronization Mechanism (ESM)** and **Relation Synchronization Mechanism (RSM)**. Extensive experiments verify that our model achieves state-of-the-art performances. In summary, our contributions are three-fold:

- We explore AOPE task, which is valuable and critical for downstream tasks but remains under-investigated.
- We propose an end-to-end neural model, SDRN<sup>1</sup>. By adopting BERT as the encoding

layer, SDRN can learn richer context semantics. By designing the double-channel network and two synchronization mechanisms, SDRN could process opinion entity extraction and relation detection jointly and make them mutually beneficial.

- We manually build three datasets based on SemEval 2014 and 2015 benchmarks for AOPE task. Extensive experiments are conducted to verify that our model achieves state-of-the-art performances.

## 2 Related Work

Aspect-opinion pair extraction is a critical task in fine-grained opinion mining. Early studies approach this task in a pipeline manner. Hu and Liu (2004) used association mining to identify aspects and extract the adjacent adjectives as opinions. Zhuang et al. (2006) extracted aspects and opinion expressions first, and then mined the relations with dependency relation templates. Popescu and Etzioni (2005) proposed an unsupervised model to extract aspects and corresponding opinions from reviews with pre-defined rules. Although the above methods achieved great progress, they generally suffered from error propagation.

To avoid error propagation, recent studies propose joint learning methods. Klinger and Cimiano (2013a) adopted Imperatively Defined Factor graph (IDF) to analyze the inter-dependencies between aspects and opinion expressions. Klinger and Cimiano (2013b) presented a joint inference model based on IDF to extract aspect terms, opinion terms, and their relations. Yang and Cardie (2013) employed Integer Linear Programming (ILP) to identify opinion-related entities and their associated relations jointly. However, these works generally based on shallow machine learning methods and depended on hand-crafted features.

To automatically capture features, neural network methods have been applied to various fine-grained opinion mining tasks. Xu et al. (2018) used Convolutional Neural Network (CNN) to extract aspects. Wang et al. (2016), Wang et al. (2017), Yu et al. (2019) and Wang and Pan (2019) used deep learning methods to deal with aspect and opinion term co-extraction. Li et al. (2018) focused on aspect term extraction and adopted attention mechanism to exploit the latent relations between aspect and opinion terms. Hu et al. (2019) took BERT to extract aspects and corresponding sentiments.

<sup>1</sup><https://github.com/NKU-IPLab/SDRN>

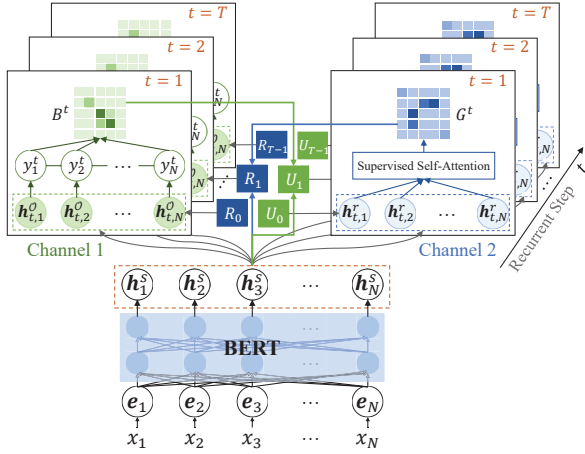


Figure 2: The framework of **Synchronous Double-channel Recurrent Network (SDRN)**.

For AOPE, [Katiyar and Cardie \(2016\)](#) explored LSTM-based models to jointly extract opinion entities and their relations with three optimization methods. But this method neglects to learn the interaction between opinion entity extraction and relation detection.

Therefore, AOPE is still under-investigated and needs more researches. In this paper, we further explore this task and propose a neural model SDRN.

### 3 Model

Given a review sentence  $S$ , **Aspect-Opinion Pair Extraction (AOPE)** task aims to obtain a collection of aspect-opinion pairs  $C = \langle a_m, o_m \rangle_{m=1}^M$  from  $S$ , where  $a_m$  and  $o_m$  represent the aspect and the opinion expression, respectively<sup>2</sup>.

To deal with AOPE task, we propose **Synchronous Double-channel Recurrent Network (SDRN)**. The overall framework of SDRN is illustrated in [Figure 2](#). Specifically, we first adopt BERT as the encoding layer to learn the context representations. Then, an opinion entity extraction unit and a relation detection unit are constructed as double channels to extract aspects, opinion expressions, and relations simultaneously. Furthermore, a synchronization unit is designed to enable information interaction between the double channels. To capture high-level representations, we recurrently execute the above units. After multiple recurrent steps, we adopt an inference layer to obtain aspect-opinion pairs.

<sup>2</sup>Note that  $a_m$  or  $o_m$  could be a single word or a phrase.

### 3.1 Encoding Layer

Given a review sentence  $S$ , we first tokenize it using the WordPiece vocabulary ([Wu et al., 2016](#)) and add tokens [CLS] and [SEP] to the beginning and the end of the tokenized sentence, respectively. As a result, we obtain the input sequence  $X = \{x_1, x_2, \dots, x_N\}$  with  $N$  tokens for each sentence.

Inspired by the success of BERT ([Devlin et al., 2019](#)), we adopt it as the encoder to learn the contextual semantics. For each token  $x_i$ , the initial embedding  $e_i$  is constructed by summing the corresponding token embedding  $e_i^w$ , segment embedding  $e_i^s$ , and position embedding  $e_i^p$ . Then, the embedding sequence  $E = \{e_1, e_2, \dots, e_N\}$  is fed into BERT, which consists of stacked Transformer blocks with multiple self-attention heads ([Vaswani et al., 2017](#)). We take the output of the last Transformer block as the context representation sequence  $H^s = \{h_1^s, h_2^s, \dots, h_N^s\}$ .

### 3.2 Double-channel Recurrent Network

#### 3.2.1 Opinion Entity Extraction Unit

The opinion entity extraction unit, which aims at extracting the aspects and the opinion expressions, is developed as a channel of SDRN. To deal with this sequence labeling task, we couple Conditional Random Field (CRF) ([Lafferty et al., 2001](#)) upon the encoding layer, which serves as the opinion entity extraction unit. Formally, CRF adopts a state score matrix  $P \in \mathbb{R}^{N \times K}$  to model the mappings between tokens and labels, and a transition score matrix  $Q \in \mathbb{R}^{K \times K}$  to model the relations between adjacent labels, where  $K$  denotes the dimension of the label space<sup>3</sup>. For a sequence of predicted labels  $Y^t = \{y_1^t, y_2^t, \dots, y_N^t\}$  at the  $t$ -th recurrent step, we define its score as follows:

$$S(X, Y^t) = \sum_{i=1}^N Q_{y_{i-1}^t, y_i^t} + \sum_{i=1}^N P_{i, y_i^t}^t, \quad (1)$$

$$P^t = H_t^o W_p + b_p, \quad (2)$$

where  $H_t^o = \{h_{t,1}^o, h_{t,2}^o, \dots, h_{t,N}^o\}$  denotes the input hidden representation sequence at the  $t$ -th recurrent step for the opinion entity extraction unit, which is calculated with the context representation sequence  $H^s$  and the relation synchronization semantics  $R_{t-1}$ . The details will be described in

<sup>3</sup>Following the BIO tagging scheme, we define five labels, including BA (beginning of aspect), IA (inside of aspect), BP (beginning of opinion expression), IP (inside of opinion expression), and O (others).

Section 3.3.2. The matrices  $W_p \in \mathbb{R}^{d_o \times K}$  and  $b_p \in \mathbb{R}^{N \times K}$  are model parameters, where  $d_o$  denotes the dimension of hidden representation  $\mathbf{h}_{t,i}^o$ .

Then, the probability of the predicted sequence  $Y^t$  can be calculated as follows:

$$p(Y^t | X) = \frac{\exp(S(X, Y^t))}{\sum_{\tilde{Y}^t \in Y_X^t} \exp(S(X, \tilde{Y}^t))}, \quad (3)$$

where  $Y_X^t$  denotes all possible label sequences. During training, we maximize the likelihood probability  $p(Y | X)$  of gold label sequence at the last step. During decoding, we use the Viterbi algorithm to find the label sequence with the maximum score.

### 3.2.2 Relation Detection Unit

To extract opinion entities and relations simultaneously, we design a relation detection unit as another channel of SDRN. Considering the complicated relations between aspects and opinion expressions, we devise a *supervised self-attention* mechanism as the relation detection unit to flexibly model token-level relations without the sequential limitation.

At the  $t$ -th recurrent step, we first compute the attention matrix  $G^t \in \mathbb{R}^{N \times N}$  whose element  $g_{i,j}^t$  represents the degree of correlation between the  $i$ -th token and the  $j$ -th token as follows:

$$g_{i,j}^t = \frac{\exp\left(\gamma\left(\mathbf{h}_{t,i}^r, \mathbf{h}_{t,j}^r\right)\right)}{\sum_{k=1}^N \exp\left(\gamma\left(\mathbf{h}_{t,i}^r, \mathbf{h}_{t,k}^r\right)\right)}, \quad (4)$$

$$\gamma\left(\mathbf{h}_{t,i}^r, \mathbf{h}_{t,j}^r\right) = \tanh\left(\mathbf{h}_{t,i}^r W_r^1 + \mathbf{h}_{t,j}^r W_r^2\right) W_r^3, \quad (5)$$

where  $\gamma$  is a score function, and  $\mathbf{h}_{t,i}^r$  denotes the input hidden representation of the  $i$ -th token for the relation detection unit. Note that the hidden representation sequence  $H_t^r = \{\mathbf{h}_{t,1}^r, \mathbf{h}_{t,2}^r, \dots, \mathbf{h}_{t,N}^r\}$  is calculated with the context representation sequence  $H^s$  and the entity synchronization semantics  $U_{t-1}$ . The details will be described in Section 3.3.1. The matrices  $W_r^1 \in \mathbb{R}^{d_r \times d_r}$ ,  $W_r^2 \in \mathbb{R}^{d_r \times d_r}$ , and  $W_r^3 \in \mathbb{R}^{d_r \times 1}$  are model parameters, where  $d_r$  is the dimension of hidden representation  $\mathbf{h}_{t,i}^r$ .

At the last step  $T$ , we further introduce supervision information into the calculation of the attention matrix  $G^T$  by maximizing the likelihood probability as follows:

$$p(Z|X) = \prod_{i=1}^N \prod_{j=1}^N p(z_{i,j}|x_i, x_j), \quad (6)$$

where the standard relation matrix  $Z \in \mathbb{R}^{N \times N}$  consists of element  $z_{i,j}$ , and the relation probability  $p(z_{i,j}|x_i, x_j)$  can be calculated as follows:

$$p(z_{i,j}|x_i, x_j) = \begin{cases} g_{i,j}^T, & \text{if } z_{i,j} = 1 \\ 1 - g_{i,j}^T, & \text{if } z_{i,j} = 0 \end{cases}, \quad (7)$$

where  $z_{i,j} = 1$  denotes the fact that there is a relation between the  $i$ -th token and the  $j$ -th token, and vice versa. With this supervision information, the attention can be guided to capture the correlations between the tokens more effectively.

### 3.3 Synchronization Unit

Since the above two channels are interdependent, it is important to synchronize their information and make them mutually beneficial. To this end, we design **Entity Synchronization Mechanism (ESM)** and **Relation Synchronization Mechanism (RSM)** to update the hidden representation sequences  $H_t^o$  and  $H_t^r$  by exchanging the high-level information.

#### 3.3.1 Entity Synchronization Mechanism

Considering that opinion entities are generally phrases, both opinion entity semantics and token-level interactions are crucial in detecting relations. For instance, given an aspect ‘hot dog’ and an opinion expression ‘tasty’, there is no relation between ‘hot’ and ‘tasty’ when only token-level interaction is considered, but it is easy to detect the relation if we utilize the semantics of aspect ‘hot dog’.

Accordingly, we design ESM to capture the corresponding entity semantics for each token and integrate these semantics into the hidden representation sequence  $H_{t+1}^r$ . Specifically, based on the predicted label sequence  $Y^t$  and its probability obtained from the opinion entity extraction unit, each entity semantics  $\mathbf{u}_{t,i}$  of the  $i$ -th token at the  $t$ -th recurrent step can be calculated as follows:

$$\mathbf{u}_{t,i} = \sum_{j=1}^N \varphi(B_{i,j}^t) \mathbf{h}_j^s, \quad (8)$$

$$\varphi(B_{i,j}^t) = \frac{B_{i,j}^t}{\sum_{k=1}^N B_{i,k}^t}, \quad (9)$$

where  $B_{i,j}^t$  is the label probability of the  $j$ -th token if the  $i$ -th token and the  $j$ -th token belong to the same entity; otherwise,  $B_{i,j}^t$  is zero. And  $\varphi(\cdot)$  is a normalization function.

To integrate both the context representation  $\mathbf{h}_i^s$  and the entity semantics  $\mathbf{u}_{t,i}$ , we calculate the hidden representation  $\mathbf{h}_{t+1,i}^r$  as follows:

$$\mathbf{h}_{t+1,i}^r = \sigma(\mathbf{u}_{t,i} W_r^4 + \mathbf{h}_i^s W_r^5), \quad (10)$$

where  $W_r^4 \in \mathbb{R}^{d_s \times d_r}$  and  $W_r^5 \in \mathbb{R}^{d_s \times d_r}$  are model parameters,  $d_s$  is the dimension of context representation, and  $\sigma$  is the activation function which can be tanh or sigmoid function. Note that we use zero matrix to initialize the entity semantics sequence  $U_0 = \{\mathbf{u}_{0,1}, \mathbf{u}_{0,2}, \dots, \mathbf{u}_{0,N}\}$ .

### 3.3.2 Relation Synchronization Mechanism

Since the relations between opinion entities can provide clues for opinion entity extraction, it's important to encode the relation semantics. For example, if 'overrated' is used to modify 'pizza', this relation could provide guidance to extract the aspect 'pizza' and the opinion expression 'overrated'.

Thus, we design RSM to capture the semantics which reflect the relations and update the hidden representation sequence  $H_{t+1}^o$ . Concretely, at the  $t$ -th recurrent step, we can calculate the relation semantics  $\mathbf{r}_{t,i}$  of the  $i$ -th token with the correlated degree  $g_{i,j}^t$  from the relation detection unit:

$$\mathbf{r}_{t,i} = \sum_{j=1}^N \varphi(\phi(g_{i,j}^t)) \mathbf{h}_j^s, \quad (11)$$

$$\phi(g_{i,j}^t) = \begin{cases} g_{i,j}^t, & \text{if } g_{i,j}^t \geq \beta \\ 0, & \text{if } g_{i,j}^t < \beta \end{cases}, \quad (12)$$

where  $\varphi(\cdot)$  is the same normalization function as Eq.(9). To avoid noise, we utilize  $\phi(\cdot)$  to filter correlated scores below the given threshold  $\beta$ .

Then, we combine the relation semantics  $\mathbf{r}_{t,i}$  and context representation  $\mathbf{h}_i^s$  to obtain the hidden representation  $\mathbf{h}_{t+1,i}^o$ :

$$\mathbf{h}_{t+1,i}^o = \sigma(\mathbf{r}_{t,i} W_o^1 + \mathbf{h}_i^s W_o^2), \quad (13)$$

where  $W_o^1 \in \mathbb{R}^{d_s \times d_o}$  and  $W_o^2 \in \mathbb{R}^{d_s \times d_o}$  are model parameters. Similar to ESM, the initial relation semantics sequence  $R_0 = \{\mathbf{r}_{0,1}, \mathbf{r}_{0,2}, \dots, \mathbf{r}_{0,N}\}$  is set to zero.

Particularly, the integration methods used in ESM and RSM can also make the proposed SDRN easy to optimize, which is similar to the shortcut connections (He et al., 2016).

### 3.4 Joint Learning

To synchronously learn the proposed two channels, we fuse the loss functions from the two channels. For opinion entity extraction unit, given the gold label sequence  $Y$ , we minimize the negative log-likelihood loss function at the last step as follows:

$$\mathcal{L}_E = \log \sum_{\tilde{Y} \in Y_X^T} \exp\left(S\left(X, \tilde{Y}\right)\right) - S(X, Y). \quad (14)$$

For the relation detection unit, we convert the gold annotation to a one-hot matrix, where 0 denotes no relations, and 1 represents the existence of relations between two tokens. Then, we minimize the cross-entropy loss between the predicted distribution  $\hat{p}(z_{i,j}|x_i, x_j)$  at the last step and the gold distribution  $p(z_{i,j}|x_i, x_j)$  as follows:

$$\mathcal{L}_R = - \sum_{i=1}^N \sum_{j=1}^N p(z_{i,j}|x_i, x_j) \log[\hat{p}(z_{i,j}|x_i, x_j)]. \quad (15)$$

Then, the two parts are combined to construct the loss objective of the entire model:

$$\mathcal{L}(\theta) = \mathcal{L}_E + \mathcal{L}_R. \quad (16)$$

The optimization problems in Eq. (16) can be solved by using any gradient descent method. In this paper, we adopt the BERTAdam method.

### 3.5 Inference Layer

Because SDRN synchronously processes opinion entity extraction and relation detection, an inference layer is introduced to generate aspect-opinion pairs based on the results of the two channels.

With the label sequence  $Y^T$  predicted by the opinion entity extraction unit at the last recurrent step, we can obtain the aspect set  $A = \{a_1, a_2, \dots, a_{l_A}\}$  with  $l_A$  aspects and the opinion set  $O = \{o_1, o_2, \dots, o_{l_O}\}$  with  $l_O$  opinion expressions. Then, the relations between aspects and opinion expressions can be calculated according to the weight matrix  $G^T$  from the relation detection unit. For instance, given an aspect  $a = \{x_{i_S^a}, \dots, x_{i_E^a}\}$  and an opinion expression  $o = \{x_{i_S^o}, \dots, x_{i_E^o}\}$ , the correlated degree  $\delta$  between them can be calculated as follows:

$$\delta = \frac{1}{2} \left( \frac{1}{|a|} \sum_{k=i_S^a}^{i_E^a} \sum_{l=i_S^o}^{i_E^o} g_{k,l} + \frac{1}{|o|} \sum_{l=i_S^o}^{i_E^o} \sum_{k=i_S^a}^{i_E^a} g_{l,k} \right), \quad (17)$$

where  $|a|$  and  $|o|$  denote the length of aspect and opinion expression. The pair  $\langle a, o \rangle$  is extracted only if  $\delta$  is higher than a given threshold  $\hat{\delta}$ .



Dataset		#Sent	#A	#O	#R
SemEval-14	Train	3041	3693	3512	2809
	Restaurant	800	1134	1014	936
SemEval-14	Train	3045	2359	2500	1535
	Laptop	800	653	677	380
SemEval-15	Train	1315	1205	1217	1231
	Restaurant	685	542	516	516
JDPA	Camera	3125	6107	4557	4144
	Car	6501	8272	11123	8709
MPQA		9471	4676	5849	4823

Table 1: Statistics of datasets. #Sent, #A, #O, and #R represent the number of sentences, aspects, opinion expressions, and relations, respectively.

## 4 Experiments

### 4.1 Datasets

To evaluate the effectiveness of SDRN, we conduct extensive experiments on five benchmark datasets from SemEval 2014<sup>4</sup> (Pontiki et al., 2014), SemEval 2015<sup>5</sup> (Pontiki et al., 2015), MPQA version 2.0 corpus<sup>6</sup> (Wiebe et al., 2005), and J.D. Power and Associates Sentiment Corpora<sup>7</sup> (JDPA) (Kessler et al., 2010). The statistics of these benchmark datasets are shown in Table 1. For SemEval 2014 and 2015 datasets, we manually build relations between aspects and opinion expressions because the original datasets only contain the gold standard annotation for aspects. Note that we follow the annotations for opinion expressions provided by Wang et al. (2016) and Wang et al. (2017).

### 4.2 Experimental Setting

We adopt the BERT<sub>BASE</sub><sup>8</sup> model, which consists of 12 Transformer blocks with 12 self-attention heads, as the encoding layer of SDRN. The dimensions of both the embeddings and the context representation in BERT<sub>BASE</sub> are 768. To enhance the information interaction between the double channels, we set the recurrent step to 2. During training, we use the BERTAdam optimizer with 0.1 warmup rate. The learning rate is set to 2e-5 and 0.001 for fine-tuning BERT and training our model, respectively. Meanwhile, we set the batch size to 10 and the dropout rate to 0.5. With the cross-validation, other hyper-parameters are set as follows:  $d_o = 250$ ,  $d_r = 250$ ,  $\beta = 0.1$ , and  $\hat{\delta} = 0.5$ .

<sup>4</sup><http://alt.qcri.org/semeval2014/task4/>

<sup>5</sup><http://alt.qcri.org/semeval2015/task12/>

<sup>6</sup><http://www.cs.pitt.edu/mpqa/>

<sup>7</sup><http://verbs.colorado.edu/jdpacorpus/>

<sup>8</sup><https://github.com/google-research/bert>

### 4.3 Evaluation

We use  $F_1$ -score to evaluate the performance of SDRN. We consider a predicted aspect-opinion pair is correct if the gold standard annotations contain a pair the same as the prediction. Besides, following Katiyar and Cardie (2016), we report Binary Overlap  $F_1$ -score for MPQA dataset.

### 4.4 Baselines

To achieve the comprehensive and comparative analysis of SDRN, we compare it with two kinds of models, including *Pipeline* methods<sup>9</sup> and *Joint* methods.

#### 4.4.1 Pipeline method

For *Pipeline* methods, we first select five advanced extraction models to recognize opinion entities. Then, we train the relation detection unit (RD) separated from SDRN with BERT to detect relations. The details about RD are described in Section 3.2.2. The outputs of the extraction models are fed into the RD model to predict relations and obtain aspect-opinion pairs. The details of the five extraction models are described as follows:

- **HAST** (Li et al., 2018) exploits two useful clues, namely opinion summary and aspect detection history, to extract the aspects with the help of opinion information. Note that HAST can also extract aspects and opinion expressions simultaneously.
- **DE-CNN** (Xu et al., 2018) is a simple but outstanding CNN model employing two types of pre-trained embeddings, including general-purpose and domain-specific embeddings. We trained two DE-CNN models for aspect and opinion expression extraction, respectively.
- **IMN** (He et al., 2019) is an interactive multi-task learning network which jointly learns multiple tasks, including aspect and opinion term co-extraction, aspect-level sentiment classification, etc.
- **SPAN** (Hu et al., 2019) is a span-based extraction framework based on BERT. We trained two SPAN models for aspect and opinion expression extraction, respectively.
- **RINANTE** (Dai and Song, 2019) is a weak supervised opinion entity extraction model

<sup>9</sup>The *Pipeline* models are expressed in the form of ‘{\*}+{#}’, where ‘\*’ means the opinion entity extraction method and ‘#’ is the relation detection method.

trained with human-labeled data and rule labeled auxiliary data.

#### 4.4.2 Joint method

To sufficiently verify the performance of SDRN, we also compare it with *Joint* models: IDF (Klinger and Cimiano, 2013b), CRF+ILP (Yang and Cardie, 2013), and LSTM+SLL+RLL (Katiyar and Cardie, 2016). The details can be found in Section 2.

### 4.5 Experimental Results

We demonstrate and analyze the experimental results to answer the following research questions:

- How does SDRN perform compared with the baselines on AOPE task?
- Can the performance of opinion entity extraction subtask be improved by the joint learning with relation detection?
- Does the synchronization unit promote the information interaction and further enhance the joint learning?

#### 4.5.1 Pair Extraction

The comparison results of aspect-opinion pair extraction are shown in Table 2 and Table 3. According to the results, SDRN consistently obtains the state-of-the-art performances on five datasets. Compared to the best pipeline model, SDRN outperforms SPAN+RD by 2.31%, 1.14% and 3.39% on 14-Res, 14-Lap and 15-Res, respectively. This indicates that the joint model can effectively avoid the error propagation led by pipeline models. Furthermore, SPAN+RD outperforms other baselines, which shows that BERT can capture rich context representations. Besides, HAST+RD, IMN+RD and RINANTE+RD, which utilize the aspect and opinion term co-extraction models, achieve better performances than DE-CNN+RD. This shows that it is helpful to detect relations with considering latent relations between aspects and opinion expressions during the extraction phase.

We also compare SDRN with joint models on JDPA and MPQA datasets, and the results are reported using 10-fold cross validation. According to Table 3, our model brings significant improvements without any hand-crafted features. Particularly, for pair extraction, the results of IDF Joint are 7.4% and 10.5% inferior to IDF Pipeline on JDPA Camera and JDPA Car datasets. This illustrates that joint models may worse than pipeline models without adequate information interaction between opinion entity extraction and relation detection.

Models		14-Res	14-Lap	15-Res
Pipeline	HAST+RD	73.55	64.05	65.20
	DE-CNN+RD	71.02	61.11	64.19
	IMN+RD	73.69	62.98	65.56
	SPAN+RD	74.17	65.99	67.55
	RINANTE+RD	74.34	64.17	65.42
Joint	SDRN w/o ESM	74.60	66.57	69.28
	SDRN w/o RSM	75.01	66.43	69.33
	SDRN w/o ESM&RSM	74.28	65.74	67.67
	SDRN	<b>76.48</b>	<b>67.13</b>	<b>70.94</b>

Table 2: Experimental results of the aspect-opinion pair extraction compared on three SemEval datasets ( $F_1$  score, %). Note that the improvements over the baselines are significant ( $p < 0.05$ ).

Models	JDPA Camera	JDPA Car	MPQA
IDF Pipeline	21.5	26.6	N/A
IDF Joint	14.1	16.1	N/A
CRF+ILP	N/A	N/A	57.04
LSTM+SLL+RLL	N/A	N/A	54.98
SDRN	<b>48.63</b>	<b>47.85</b>	<b>63.95</b>

Table 3: Experimental results of aspect-opinion pair extraction compared on JDPA and MPQA datasets ( $F_1$  score, %). Note that the improvements are significant ( $p < 0.05$ ).

#### 4.5.2 Opinion Entity Extraction

Although our task aims to identify the aspect-opinion pairs, it is interesting to investigate the performance of opinion entity extraction. Hence, we compare SDRN with representative aspect and opinion expression extraction methods. The results are shown in Table 4. It is clearly shown that SDRN achieves state-of-the-art results on three datasets, which proves that the opinion entity extraction can be significantly improved by joint training with relation detection. Besides, the aspect and opinion term co-extraction models generally superior to aspect term extraction models, which demonstrates that joint extracting aspects and opinion expressions can benefits each other. HAST and SPAN are special cases of aspect term extraction models, because HAST extracts aspects with the help of opinion semantics, and SPAN adopts BERT as the backbone model.

#### 4.5.3 Synchronization Unit

To investigate the efficacy of the synchronization unit composed of ESM and RSM, we perform ablation study and list the results in the second block of Table 2. Concretely, for ‘SDRN w/o ESM’, we drop ESM and simply update the relation hidden representation  $H_t^r$  via a fully-connection layer. Similarly, ‘SDRN w/o RSM’ drops RSM and adopts a fully-connection layer to update the entity hidden representation  $H_t^o$ . For ‘SDRN w/o ESM&RSM’, we simultaneously do the above two operations.

Models	14-Res		14-Lap		15-Res	
	A	O	A	O	A	O
WDEmb (Yin et al., 2016)	84.97	N/A	75.16	N/A	69.73	N/A
RNCRF <sup>†</sup> (Wang et al., 2016)	84.93	84.11	78.42	79.44	67.74	67.62
CMLA <sup>†</sup> (Wang et al., 2017)	85.29	83.18	77.80	80.17	70.73	73.68
HAST (Li et al., 2018)	85.61	85.46*	79.52	78.58*	71.46	70.77*
DE-CNN (Xu et al., 2018)	85.20	81.99*	81.59	76.34*	68.28	68.56*
IMN <sup>†</sup> (He et al., 2019)	83.33	85.61	77.96	77.51	70.04	71.94
SPAN (Hu et al., 2019)	86.20*	86.52*	80.67*	82.07*	73.65*	79.13*
GMTCMLA <sup>†</sup> (Yu et al., 2019)	84.50	85.20	78.69	79.89	70.53	72.78
RINANTE <sup>†</sup> (Dai and Song, 2019)	86.45	85.67	80.16	81.96	69.90	72.09
SDRN	<b>89.49</b>	<b>87.84</b>	<b>83.67</b>	<b>82.25</b>	<b>74.05</b>	<b>79.65</b>

Table 4: Experimental results of opinion entity extraction ( $F_1$  score, %). **A** and **O** represent the aspect extraction and the opinion expression extraction, respectively. The methods with ‘<sup>†</sup>’ are aspect and opinion term co-extraction models, and others are aspect term extraction models. The results with ‘\*’ are reproduced by us, and others are copied from the released paper. Note that the improvements over baselines are significant ( $p < 0.05$ ).

Reviews	SPAN+RD	SDRN w/o ESM&RSM	SDRN
1. The receiver was full of [superlatives] <sub>1,2</sub> for the [quality] <sub>1</sub> and [performance] <sub>2</sub> .	(quality, superlatives) (performance, superlatives)	(receiver, superlatives) ✗ (quality, superlatives) (performance, superlatives)	(quality, superlatives) (performance, superlatives)
2. The [selection of food] <sub>1</sub> is [excellent] <sub>1</sub> , and the [atmosphere] <sub>2</sub> is [great] <sub>2</sub> .	(selection, excellent) ✗ (food, excellent) ✗ (atmosphere, great)	(selection of food, excellent) (atmosphere, great)	(selection of food, excellent) (atmosphere, great)
3. The [bartenders] <sub>1</sub> and the [managers] <sub>2</sub> are really [nice] <sub>1,2</sub> and the [decor] <sub>3,4,5</sub> is very [comfy] <sub>3</sub> and [laid-back] <sub>4</sub> , all the while being [trendy] <sub>5</sub> .	(bartenders, nice) (managers, nice) (decor, comfy) (decor, trendy)	(bartenders, nice) (managers, nice) (decor, comfy) (-, laid-back) ✗ (decor, trendy)	(bartenders, nice) (managers, nice) (decor, comfy) (decor, laid-back) (decor, trendy)

Table 5: Case Study. The gold standard aspects and opinion expressions are in red and blue, respectively. The gold standard relations are indexed by subscripts, where aspect and opinion expression in a pair have the same subscript.

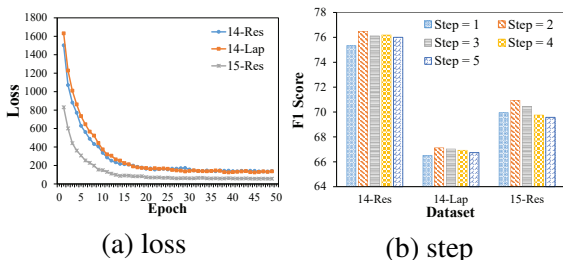


Figure 3: (a) The analysis of convergence. (b) The comparisons under varying number of recurrent steps.

Compared with *Pipeline* models, ‘SDRN w/o ESM&RSM’ is less competitive, which demonstrates that merely joint learning is not superior to the pipeline manner. By utilizing ESM or RSM, the performance is improved, which shows that either ESM or RSM is helpful. Specially, the contribution of ESM is slightly larger than RSM. Moreover, with the two synchronization mechanisms, SDRN surpasses all the baselines.

#### 4.6 Convergence and Sensitivity Study

In Figure 3(a), we verify the convergence of SDRN. The result shows that our model generally achieves convergence around 15 epochs. Besides, we present

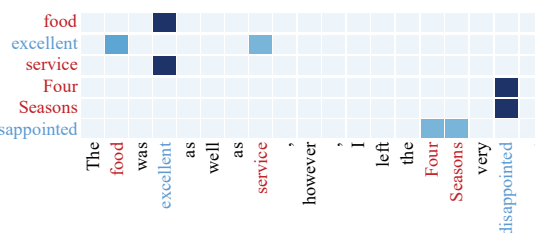


Figure 4: Visualization of attention scores. The aspects and the opinion expressions are marked with red and blue, respectively. (Best viewed in color.)

the effect of the number of recurrent steps in Figure 3(b). It can be observed that the performance of SDRN increases first and then becomes steady or slightly declining as the step number increases. For 15-Res, the limitation of training data may be the cause of performance decline. And the best results are generally obtained with two steps on three datasets, indicating that SDRN with two steps is enough to exploit the interaction information.

#### 4.7 Visualization and Case Study

In order to verify the relation detection capability of SDRN, we visualize the attention scores in Figure 4. It is shown that SDRN can accurately capture the



relations between aspects and opinion expressions, even with complex reviews.

To clearly analyze the effect of the joint learning and the synchronization unit, some predictions of SDRN, ‘SDRN w/o ESM&RSM’ and SPAN+RD are listed in Table 5. It can be concluded that SPAN+RD suffers the problem of error propagation. For example, it divides ‘*selection of food*’ into ‘*selection*’ and ‘*food*’ in Review #2, and misses ‘*laid-back*’ in Review #3. With the pipeline way, it is impossible to obtain a correct pair once there is an incorrect extraction of entities at the first step. Due to the lack of information interaction, ‘SDRN w/o ESM&RSM’ is generally faced with relation detection errors when relations are complex. For example, it extracts error pair (*receiver*, *superlatives*) in Review #1, and fails to detect the relations between ‘*decor*’ and ‘*laid-back*’ in Review #3. In contrast, our model can effectively avoid the above problems.

## 5 Conclusion

In this paper, we explored Aspect-Opinion Pair Extraction (AOPE) task and proposed Synchronous Double-channel Recurrent Network (SDRN). Specifically, the opinion entity extraction unit and the relation detection unit are designed to extract aspects, opinion expressions and their relations simultaneously. The two units update themselves in a recurrent manner and form two channels, respectively. Meanwhile, the synchronization unit is devised to integrate high-level interaction information and enable the mutual benefit on opinion entity extraction and relation detection. Extensive experiments showed that our model achieves state-of-the-art performances.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China under grant No. 61976119, the Natural Science Foundation of Tianjin under grant No. 18JCYBJC15800, and the Major Program of Science and Technology of Tianjin under grant No. 18ZXZNGX00310.

## References

- Hongliang Dai and Yangqiu Song. 2019. [Neural aspect and opinion term extraction with mined rules as weak supervision](#). In *ACL 2019*, pages 5268–5277.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). In *NAACL 2019*, pages 4171–4186.
- Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. [Target-oriented opinion words extraction with target-fused neural sequence labeling](#). In *NAACL 2019*, pages 2509–2518.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. [Deep residual learning for image recognition](#). In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778.
- 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 2019*, 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 *ACL 2019*, pages 537–546.
- Minqing Hu and Bing Liu. 2004. [Mining and summarizing customer reviews](#). In *SIGKDD 2004*, pages 168–177.
- Arzoo Katiyar and Claire Cardie. 2016. [Investigating lstms for joint extraction of opinion entities and relations](#). In *ACL 2016*.
- Jason S. Kessler, Miriam Eckert, Lyndsie Clark, and Nicolas Nicolov. 2010. [The 2010 icwsm jdpa sentiment corpus for the automotive domain](#). In *4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)*.
- Roman Klinger and Philipp Cimiano. 2013a. [Bi-directional inter-dependencies of subjective expressions and targets and their value for a joint model](#). In *ACL 2013*, pages 848–854.
- Roman Klinger and Philipp Cimiano. 2013b. [Joint and pipeline probabilistic models for fine-grained sentiment analysis: Extracting aspects, subjective phrases and their relations](#). In *ICDM 2013*, pages 937–944.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. [Conditional random fields: Probabilistic models for segmenting and labeling sequence data](#). In *ICML*, pages 282–289.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. [Aspect term extraction with history attention and selective transformation](#). In *IJCAI 2018*, pages 4194–4200.
- Xin Li and Wai Lam. 2017. [Deep multi-task learning for aspect term extraction with memory interaction](#). In *EMNLP 2017*, pages 2886–2892.

- Kang Liu, Liheng Xu, and Jun Zhao. 2012. [Opinion target extraction using word-based translation model](#). In *EMNLP 2012*, pages 1346–1356.
- Kang Liu, Liheng Xu, and Jun Zhao. 2015. [Co-extracting opinion targets and opinion words from online reviews based on the word alignment model](#). *IEEE Trans. Knowl. Data Eng.*, 27(3):636–650.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. [Semeval-2015 task 12: Aspect based sentiment analysis](#). In *NAACL 2015*, pages 486–495.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. [Semeval-2014 task 4: Aspect based sentiment analysis](#). In *COLING 2014*, pages 27–35.
- Ana-Maria Popescu and Oren Etzioni. 2005. [Extracting product features and opinions from reviews](#). In *EMNLP 2005*, pages 339–346.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *NIPS 2017*, pages 5998–6008.
- Wenya Wang and Sinno Jialin Pan. 2019. [Transferable interactive memory network for domain adaptation in fine-grained opinion extraction](#). In *AAAI 2019*, pages 7192–7199.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. [Recursive neural conditional random fields for aspect-based sentiment analysis](#). In *EMNLP 2016*, pages 616–626.
- 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 2017*, pages 3316–3322.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. [Annotating expressions of opinions and emotions in language](#). *Language Resources and Evaluation*, 39(2-3):165–210.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, and et al. 2016. [Google’s neural machine translation system: Bridging the gap between human and machine translation](#). *CoRR*, abs/1609.08144.
- Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. [Double embeddings and cnn-based sequence labeling for aspect extraction](#). In *ACL 2018*, pages 592–598.
- Liheng Xu, Kang Liu, Siwei Lai, Yubo Chen, and Jun Zhao. 2013. [Mining opinion words and opinion targets in a two-stage framework](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pages 1764–1773. The Association for Computer Linguistics.
- Bishan Yang and Claire Cardie. 2013. [Joint inference for fine-grained opinion extraction](#). In *ACL 2013*, pages 1640–1649.
- Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. [Unsupervised word and dependency path embeddings for aspect term extraction](#). In *IJCAI 2016*, pages 2979–2985.
- Jianfei Yu, Jing Jiang, and Rui Xia. 2019. [Global inference for aspect and opinion terms co-extraction based on multi-task neural networks](#). *IEEE/ACM Trans. Audio, Speech & Language Processing*, 27(1):168–177.
- Li Zhuang, Feng Jing, and Xiaoyan Zhu. 2006. [Movie review mining and summarization](#). In *CIKM 2006*, pages 43–50.