

Received 3 December 2022, accepted 10 December 2022, date of publication 12 December 2022,  
date of current version 20 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228835

## RESEARCH ARTICLE

# Gated Relational Encoder-Decoder Model for Target-Oriented Opinion Word Extraction

TAEGWAN KANG<sup>ID</sup>, SEGWANG KIM<sup>ID</sup>, HYEONGU YUN<sup>ID</sup>, HWANHEE LEE<sup>ID</sup>,  
AND KYOMIN JUNG<sup>ID</sup>

Department of Electrical and Computer Engineering, Automation and Systems Research Institute, Seoul National University, Seoul 08826, South Korea

Corresponding author: Kyomin Jung (kjung@snu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant through the Korea Government under Grant 2021R1A2C2008855; in part by the Brain Korea 21 (BK21) FOUR Program of the Education; and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant through the Korea Government (MSIT), Development and Study of AI Technologies to Inexpensively Conform to Evolving Policy on Ethics, under Grant 2022-0-00184.

**ABSTRACT** Target-Oriented Opinion Word Extraction (TOWE) is a challenging information extraction task that aims to find the *opinion words* corresponding to given *opinion targets* in text. To solve TOWE, it is important to consider the surrounding words of *opinion words* as well as the *opinion targets*. Although most existing works have captured the *opinion target* using Deep Neural Networks (DNNs), they cannot effectively utilize the local context, i.e. relationship among surrounding words of *opinion words*. In this work, we propose a novel and powerful model for TOWE, Gated Relational target-aware Encoder and local context-aware Decoder (GRED), which dynamically leverages the information of the *opinion target* and the local context. Intuitively, the target-aware encoder catches the *opinion target* information, and the local context-aware decoder obtains the local context information from the relationship among surrounding words. Then, GRED employs a gate mechanism to dynamically aggregate the outputs of the encoder and the decoder. In addition, we adopt a pretrained language model Bidirectional and Auto-Regressive Transformer (BART), as the structure of GRED to improve the implicit language knowledge. Extensive experiments on four benchmark datasets show that GRED surpasses all the baseline models and achieves state-of-the-art performance. Furthermore, our in-depth analysis demonstrates that GRED properly leverages the information of the *opinion target* and the local context for extracting the *opinion words*.

**INDEX TERMS** Target-oriented opinion words extraction, aspect-based sentiment analysis, neural network model, deep learning, sentiment analysis, information retrieval.

## I. INTRODUCTION

Target-Oriented Opinion Word Extraction (TOWE) [8] is a recently designed text retrieval task of aspect-based sentiment analysis (ABSA) [12], [20], [27]. In TOWE, entities or objects toward which users show their attitudes are regarded as the *opinion targets*. Correspondingly, those terms explicitly expressing attitudes are defined as *opinion words*. Given a sentence and *opinion targets*, the goal of TOWE is to extract the *opinion words* that reveal sentiment about the *opinion targets*. For example, in the sentence “The service is amazing and food is out of this world.”, TOWE identifies the

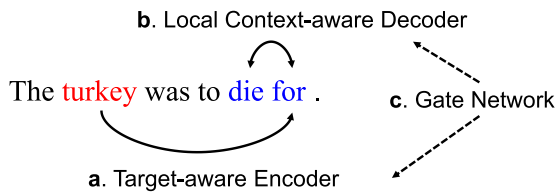
word “amazing” as the *opinion word* for “service”, and the terms “out of this world” as the *opinion words* for “food”. Table 1 shows more example pairs of *opinion targets* and *opinion words*.

Extracting *opinion targets* or *opinion words* has been widely used in various Natural Language Processing (NLP) tasks such as sentiment analysis [17], [26], [38], [46] and text mining [15], [30], [45]. From this trend, TOWE has become more important because it explicitly informs the correlations between *opinion targets* and *opinion words*. To address this task, [8] have constructed four benchmark datasets based on the ABSA datasets [27], [28], [29] from different domains, restaurant and laptop. Also, they have dealt with TOWE task by designing it as a BIO-tagged token

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li<sup>ID</sup>.

**TABLE 1.** Example of user reviews and their extracted pairs of *opinion targets* and *opinion words*. The red-colored words and the blue-colored words represent opinion targets and opinion words, respectively.

<b>User Reviews:</b>	
"The <i>service</i> is <i>amazing</i> and <i>food</i> is <i>out of this world</i> . I ordered the <i>Clams Oreganato appetizer</i> . It was <i>great</i> and <i>top notch</i> . The Parm lunch special came with pasta and choice of soup or salad. I had a <i>delicious meal</i> with <i>great service</i> . Awesome!"	
<b>Extracted opinion targets - opinion pairs:</b>	
1. <i>service</i> - <i>amazing</i>	
2. <i>food</i> - <i>out of this world</i>	
3. <i>Clams Oreganato appetizer</i> - <i>great, top notch</i>	
4. <i>meal</i> - <i>delicious</i>	
5. <i>service</i> - <i>great</i>	



**FIGURE 1.** The concept of GRED. The target-aware encoder finds target-dependent *opinion words*, and the local context-aware decoder seeks local context-dependent *opinion words*. Then, the gate network dynamically aggregates the outputs of these two networks.

classification problem [32] for a sentence with *opinion targets*.

To solve TOWE, previous works have primarily focused on how to incorporate *opinion target* information into sentence representations. In this literature, there have been various attempts to encode the *opinion target* such as using an additional BiLSTM-based target encoder [8] or adopting a target position embedding [36], [41]. With the development of pre-trained language models (PLMs) [6], [7], [16], recent works have added a special token indicating the *opinion targets* to exploit the power of PLMs, which have shown promising results [10], [14].

Nevertheless, these approaches have a critical limitation in that they cannot effectively utilize the surrounding words of the *opinion words*. Understanding such local context as well as the *opinion target* information is helpful in extracting the *opinion words*. For example, in the sentence “The turkey was to die for”, a human can easily identify the word “for” as the *opinion words* by considering the *opinion target* “turkey” and the surrounding word “die” together. Although [14] have attempted to incorporate this local context information into PLMs, they have merely combined the target information with the local context information, resulting in congested sentence representation. Consequently, their performance has shown limited improvements only where the local context information has less importance.

In this paper, we propose a **Gated Relational target-aware Encoder** and local context-aware **Decoder**-based sequence labeling model (**GRED**), which dynamically leverages the *opinion target* information and the local context information

for TOWE. Specifically, the target-aware encoder first obtains the *opinion target* information by using the target relation network [34]. Simultaneously, the local context-aware decoder captures the local context information from the relationships among surrounding words by using the local context relation network. Then, GRED employs the gate network to aggregate the outputs of the encoder and the decoder. The role of the gate network is to determine how much those outputs will impact the final prediction. Therefore, GRED can properly mix the *opinion target* information and the local context information rather than roughly combining them. The concept of GRED is illustrated in Fig. 1. In addition, to improve the language knowledge of both the encoder and the decoder, we adopt a pretrained language model Bidirectional and Auto-Regressive Transformer (BART) [16] as the structure of GRED.

We evaluate our GRED on the four public datasets, *14res*, *14lap* [29], *15res* [27] and *16res* [28]. Our extensive experiments demonstrate that GRED performs better than the baselines and achieves state-of-the-art performance. Additionally, further comprehensive analyses validate the effectiveness of the target-aware encoder, the local context-aware decoder, and the gate network of GRED. In summary, the contributions of our work are as follows:

- We propose a novel transformer-based sequence labeling model GRED for TOWE. To enhance the language knowledge of GRED, we adopt both the encoder and decoder of BART, a pretrained language model.
- GRED’s target-aware encoder and local context-aware decoder utilize the *opinion target* information and local context information, which are critical to solving TOWE task. Furthermore, the gate network of GRED can dynamically leverage a target-aware encoder and a local context-aware decoder.
- With our comprehensive analysis of GRED, we demonstrate the effectiveness of GRED from various perspectives. Our proposed methods not only enhance the performance of TOWE task but also help interpret predicted results as well as improve the model performance.

The remaining parts of our paper are organized into four sections. In section II, we briefly review the previous works on aspect-based sentiment analysis and TOWE tasks. Then we introduce our model GRED in section III. Experimental results are provided and discussed in section IV. Finally, we present the conclusion and future works in section V.

## II. RELATED WORKS

Extracting *opinion targets* and *opinion words* have been principal tasks for natural language processing. One line of this research has focused on *opinion target* extraction (OTE), which aims to seek the *opinion target* aspect terms in sentences [23], [30], [42], [44]. In other approaches, *opinion words* extraction (OWE) has attempted to find words expressing users’ attitudes [2], [5], [11], [33]. Recently, several works

**TABLE 2.** The examples of BIO-tagging for TOWE. The underlined terms indicate the *opinion targets*, and the bold terms represent the *opinion words*.

Targets	BIO-scheme sentence
duck confit	"The/O duck/O confit/ <u>O</u> is/O always/O <b>[amazing/B]</b> and/O the/O foie/O gras/O terrine/O with/O figs/O was/O out/O of/O this/O world/O ./O"
foie gras terrine	"The/O duck/O confit/O is/O always/O amazing/O and/O the/O foie/O gras/O <u>terrine/O</u> with/O figs/O was/O <b>[out/B of/I this/I world/I]</b> ./O"

have proposed a coextraction framework that extracts *opinion targets* and *opinion words* jointly. They have detected the targets and *opinion words* jointly by utilizing a word alignment model [21] or multitask learning [18], [39], [40]. These works have been able to obtain information on *opinion targets* and *opinion words*, respectively. However, none of them have considered the relationship between *opinion targets* and *opinion words*.

To study this, there have been researches conducted on the task of extracting corresponding *opinion words* for given *opinion targets*. Classical methods have been designed to seek corresponding opinion terms based on word distance [12] and dependency parsing tree [47]. However, these methods require external knowledge and show vulnerability to diverse patterns of data. Therefore, subsequent works have integrated *opinion target* information into the context and extracted the corresponding *opinion words* using deep neural networks such as RNN [8], [41] and GCN (Graph Convolutional Network) [13], [36]. In these works, [8] have first proposed an end-to-end neural network model using IOG (Inward-Outward LSTM + Global context) to fuse *opinion target* information with the global context but required high model complexity. Unlike the previous work, [13], [22], [36], [41] have employed position embedding of *opinion targets*, resulting in not increasing the model complexity excessively. Additionally, [22] have used various combinations of embedding architectures such as Transformer, GCN, and RNN. GRED has also utilized the transformer architecture to obtain sentence embeddings. However, while most studies have adopted only the encoder of the transformer, we have used both the encoder and the decoder.

Nevertheless, the above methods cannot fully utilize powerful pretrained language models to address TOWE task. Thus, recent works have adopted pretrained language models and achieved promising results [9], [10], [14], [43]. They have incorporated *opinion target* information into PLMs by modifying the input sentence with explicitly marking the *opinion targets*. In addition, [14] have shown that local context information is also important for solving TOWE task. Inspired by these previous works, our proposed method GRED has attempted to obtain the *opinion target* information and local context information. However, GRED is different from the pretrained language model-based methods

as GRED dynamically mixes these two important pieces of information. Furthermore, we have adopted pretrained language model BART, which can exploit both the encoder and decoder of the transformer. As a result, GRED has shown the most overwhelming performance in TOWE task.

### III. METHODS

#### A. TASK FORMALIZATION

TOWE task can be formalized as a sequence labeling problem for *opinion target* specified sentences [8]. For given a sentence  $s = \{w_1, w_2, \dots, w_n\}$  consisting of a sequence of  $n$  words and *opinion targets*, we use BIO tagging scheme [32] that classifies each word as  $y_i \in \{B, I, O\}$  (Beginning, Inside, and Others) to solve TOWE task. Table 2 illustrates the BIO-scheme sentence for given *opinion targets*.

#### B. OVERALL FRAMEWORK

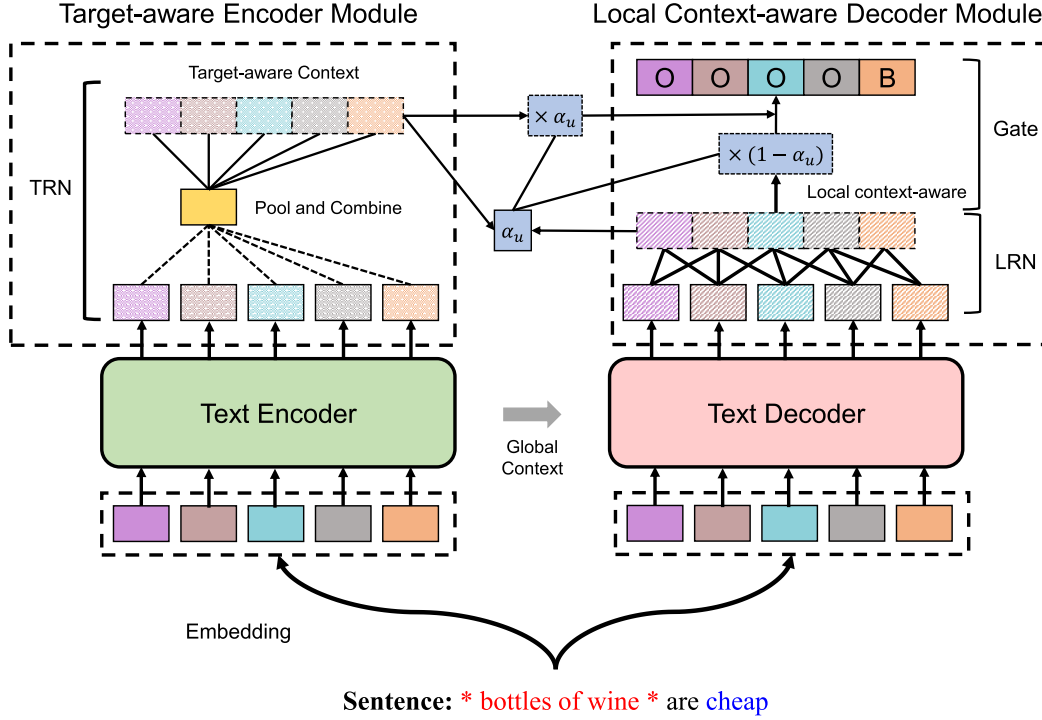
The structure of our Gated Relational target-aware Encoder and local context-aware Decoder (GRED) is illustrated in Fig. 2. GRED consists of a target-aware context encoder module and a local context-aware decoder module with a gate network. The overall framework of GRED is as follows. We first add a special token to indicate the *opinion targets* within the sentences. Next, these modified sentences are encoded into the target-aware encoder module and the local context-aware decoder module. These modules aim to capture *opinion target* information and local context information via the relation networks as in [34] and [14], respectively. The target-aware encoder module extracts the target-aware representation using multi-head self-attention layers and a target relation network. In the local context-aware decoder module, the local context relation network catches the local context representation from the surrounding words, and then the gate network outputs final representations by aggregating these two representations. Finally, the Conditional Random Field (CRF) layer determines the tags of the sequence based on the final representation as in [8], [14], and [10].

#### C. TARGET-AWARE INPUT PREPROCESSING

To fully utilize the pretrained language model, we use the symbol “\*” as a target indicating token instead of the separator token “[SEP]” used in the BERT-based models [10], [14]. We insert the symbol “\*” before and after *opinion target*  $t = \{w_i, \dots, w_{i+j}\}$ . Then, the sentence  $s$  is pre-processed as  $s^* = \{w_1, \dots, *, w_i, \dots, w_{i+j}, *, \dots, w_n\}$ . GRED also uses a BART-style sequence classification scheme, which processes the right-shifted input for the decoder. Thus, the input sentence of the decoder  $s_d^*$  is modified as  $s_d^* = \{<s>, w_1, \dots, *, w_i, \dots, w_{i+j}, *, \dots, w_n, </s>\}$ .

#### D. TARGET-AWARE ENCODER MODULE

The target-aware encoder module consists of a text encoder and a target relation network. The text encoder receives the modified sentence  $s^*$  and produces the embedding of each word of those sentences. For each word in the sentence, the



**FIGURE 2.** The overview of GRED. GRED comprises a target-aware context encoder module, a local context-aware decoder module, and a gate network. The target-aware context encoder module and the local context-aware decoder module contain the target relation module and the local context relation module, respectively. The gate network computes final representations based on outputs from both modules.

target relation network generates target-aware word representations based on the relationships between the *opinion targets* and each word in the sentence.

### 1) TEXT ENCODER SENTENCE EMBEDDINGS

The text encoder adopts the structure of the transformer encoder in Vaswani et al. [35], which has increased expressive power by using a multi-head self-attention mechanism. To exploit the power of the pretrained language model, we use the encoder part of BART for the text encoder. BART has the same architecture as the transformer but incorporated pretrained language knowledge. The text encoder takes the modified sentence  $s^*$  and generates context embedding of  $s^*$  (Figure 2 Text Encoder):

$$H^{enc} = \text{BART}_{enc}(s^*), \quad (1)$$

where  $\text{BART}_{enc}$  is the encoder layers of BART,  $h_i^e$  is the encoder representation of each token  $w_i$ , and  $H^{enc} = \{h_1^e, \dots, h_i^e, \dots, h_{i+j}^e, \dots, h_n^e\}$  is the sentence embedding of  $s^*$ .

### 2) OPINION TARGET EMBEDDING

To obtain the *opinion target* embedding, we first gather the parts that correspond to the *opinion target* from the sentence embedding  $H^{enc}$  and then apply a pooling layer to these parts. For the pooling layer, various pooling methods such as max-pooling, mean-pooling, and LogSumExp (LSE)-pooling are

used to compute the *opinion target* embeddings:

$$\begin{aligned} h_{target}^e &= \text{Max}(\{h_i^e, \dots, h_{i+j}^e\}), \quad (\text{Max-pool}), \\ h_{target}^e &= \frac{1}{j+1} \sum_{k=i}^{i+j} h_k^e \quad (\text{Mean-pool}), \\ h_{target}^e &= \text{LSE}(\{h_i^e, \dots, h_{i+j}^e\}) \quad (\text{LSE-pool}), \end{aligned} \quad (2)$$

where  $\text{LSE}(\{h_i^e, \dots, h_{i+j}^e\}) = \log \sum_{k=i}^{i+j} \exp(h_k)$  and  $h_{target}^e$  is an *opinion target* embedding.

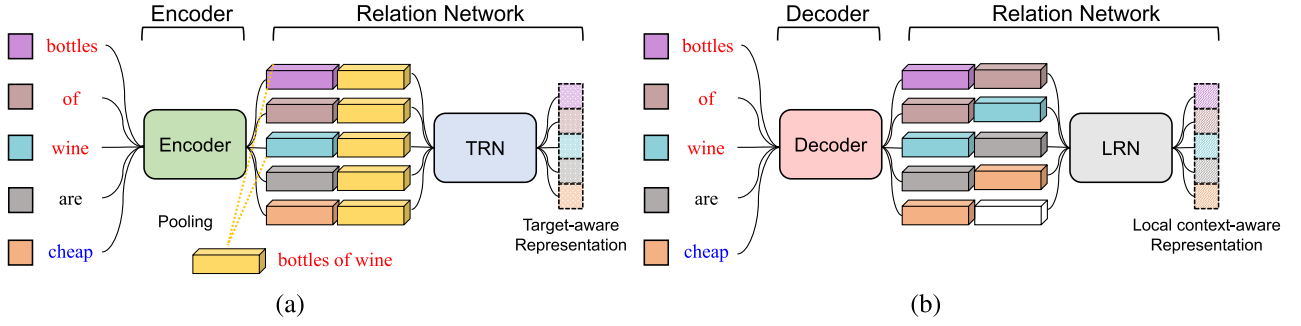
### 3) TARGET-AWARE REPRESENTATION

Target-aware representations are derived from the relationships between each word in the sentence embedding and the *opinion target* embedding. Thus, we employ the relation network [34] to compute these relationships. As shown in Fig. 3(a), the target relation network takes the *opinion target* and each word in the sentence and then produces the target-aware representation. In this work, we use **MLP** as a structure of the target relation network. Given the *opinion target* embedding  $h_{target}^e$  and the sentence embedding  $H^{enc}$ , the target-aware representation is calculated as follows:

$$\begin{aligned} r_u^{Tar} &= \text{TRN}(\mathcal{C}^{Tar}(h_u^e, h_{target}^e)), \quad u = 1, 2, \dots, n \\ R^{Tar} &= \{r_1^{Tar}, r_2^{Tar}, \dots, r_n^{Tar}\}, \end{aligned} \quad (3)$$

where **TRN** is the target relation network,  $\mathcal{C}^{Tar}$  is combined function,  $r_u^{Tar}$  is the target-aware representation of each token,





**FIGURE 3.** Structures of the two relation networks: (a) Target relation network and (b) Local context relation network. The target relation network compares the *opinion target* (“bottles of wine”) and each word (“bottles”, “of”, “wine”, “are”, and “cheap”) in the sentence (“bottles of wine are cheap”). The local context relation network captures relationship among the words (“bottles”, “of”, “wine”, “are”, and “cheap”) in the sentence (“bottles of wine are cheap”).

and  $R^{Tar}$  is the target-aware representation. We use the concatenation operator as  $C^{Tar}$ , but other methods such as elementwise sum and multiplication are possible.

### E. LOCAL CONTEXT-AWARE DECODER

A local context-aware decoder is composed of a text decoder and a local context relation network. The text decoder takes the right-shifted sentence  $s_d^*$  and generates causal sentence embedding based on the encoder output and the previous words. The local context relation network produces the local context-aware representation by exploring the relationships among surrounding words in the given sentence. Finally, the gate network dynamically fuses the local context-aware representation into a target-aware representation for predicting labels.

#### 1) TEXT DECODER SENTENCE EMBEDDING

The text decoder has the same structure as the BART decoder, which is a decoder of the transformer but incorporates language knowledge. Unlike the text encoder, the text decoder uses masked multi-head attention. Thus, the right-shifted sentence  $s_d^*$  is passed to the decoder. Given the encoder sentence embedding  $H^{enc}$  and the sentence  $s_d^*$ , then the text decoder computes the text decoder sentence embedding as follows (Figure 2 Text Decoder):

$$H^{dec} = \text{BART}_{dec}(s_d^*, H^{enc}), \quad (4)$$

where  $\text{BART}_{dec}$  is the decoder layers of BART,  $h_i^d$  is the decoder representation of each token  $w_i$ , and  $H^{dec} = \{h_1^d, \dots, h_i^d, \dots, h_{i+j}^d, \dots, h_n^d\}$  is the sentence embedding of  $s_d^*$ .

#### 2) LOCAL CONTEXT-AWARE REPRESENTATION

The local context-aware representations are obtained by contextualizing each word and its surrounding words in the sentence. We use the local context relation network to explore the relationship among surrounding words in the given sentence. As illustrated in Fig. 3(b), the local context relation network calculates these relationships based on the word, its left-side

**TABLE 3.** Statistics of the four datasets. #Sentences, #Pairs, and #Targets denote the numbers of sentences, pairs of *opinion targets* and corresponding *opinion words*, and *opinion targets*, respectively.

Datasets		#Sentences	#Pairs	#Targets
14res	Train	1627	2643	2638
	Test	500	865	864
14lap	Train	1158	1634	1626
	Test	343	482	481
15res	Train	754	1076	1076
	Test	325	436	436
16res	Train	1079	1512	1512
	Test	329	457	456

word, and its right-side word. In this work, we consider tri-grams as the local context (e.g.  $\{w_{u-1}, w_u, w_{u+1}\}$ ). Thus, the local context-aware representation is computed as follows:

$$r_u^{Loc} = \text{LRN}(C^{Loc}(h_{u-1}^d, h_u^d, h_{u+1}^d)), \quad u = 1, 2, \dots, n$$

$$R^{Loc} = \{r_1^{Loc}, r_2^{Loc}, \dots, r_n^{Loc}\}, \quad (5)$$

where  $\text{LRN}$  is the local context relation network,  $C^{Loc}$  is combined function,  $r_u^{Loc}$  is the local context-aware representation of each token, and  $R^{Loc}$  is local context-aware representation.  $C^{Loc}$  is composed of two different **MLP** and elementwise sum (i.e.,  $C^{Loc}(x, y, z) = A(x, y) + B(y, z)$ ;  $A$  and  $B$  are **MLP**).

#### 3) GATE NETWORK

The final representations aggregate the target-aware representation from the encoder and the local context-aware representation from the decoder. Instead of simply combining the two representations, GRED leverages the gate network to decide how both representations play a part in sequence labeling. Given the target-aware representation and the local context-aware representation, the gate network computes the aggregated representation as follows:

$$\alpha_u = \sigma(W_u^{Tar} r_u^{Tar} + W_u^{Loc} r_u^{Loc} + b_u),$$

$$r_u = \alpha_u r_u^{Tar} + (1 - \alpha_u) r_u^{Loc},$$

$$R = \{r_1, r_2, \dots, r_n\}, \quad (6)$$

where  $\alpha_u$  is the gate weight of each word,  $\sigma$  is the sigmoid function,  $W_u^{Tar}$  and  $W_u^{Loc}$  are weight matrices,  $b_u$  is the bias vector,  $r_u$  is the final representation of each token, and  $R$  is the final representation.

#### F. DECODING STRATEGY AND LOSS FUNCTION

Given the final representation  $R$ , we decode the sequence label  $Y = \{y_1, \dots, y_n\}$  based on the probability  $p(Y|R)$ . In this work, we adopt Conditional Random Field (**CRF**) as our decoding strategy because it can capture the word structural dependency of the sentence and correlations between labels. Specifically, the score function of **CRF** can be defined as:

$$S(R, Y) = \sum_{u=0}^n (A_{y_{u-1}, y_u} + Q_{u, y_u}),$$

$$Q = RW_q + b_q \quad (7)$$

where the  $A$  measures the transition score between two adjacent labels,  $W_q$  is the weight matrix,  $b_q$  is the bias, and the matrix  $Q$  is the emission score.

Then, we can compute the probability using the score function:

$$p(Y|R) = \frac{\exp(S(R, Y))}{\sum_{\hat{y} \in \bar{Y}} \exp(S(R, \hat{y}))}, \quad (8)$$

where  $\bar{Y}$  is the set of possible sequential labels. **CRF** uses the negative log-likelihood as the loss function. Thus, the given sentence loss is calculated as follows:

$$Loss(s) = -\log(p(Y|R)) \quad (9)$$

We minimize this loss function  $Loss(s)$  for training. For a decoding process, the model generates the label sequence, which maximizes  $p(Y|R)$  via the Viterbi algorithm [37].

## IV. EXPERIMENTS

### A. EXPERIMENTAL SETUP

#### 1) DATASETS

To verify the effectiveness of GRED, we conduct extensive experiments on four benchmark TOWE datasets: **14res**, **15res**, **16res**, and **14lap**. These datasets were built by [8] for TOWE task based on the SemEval Challenge 2014 Task 4 [29], SemEval Challenge 2015 task 12 [27] and SemEval Challenge 2016 task 5 [28] respectively. **14res**, **15res** and **16res** are collected from review sentences in restaurant domain. **14 lap** contains the review sentence in the lap-top domain. The statistics of these datasets are summarized in Table 3.

#### 2) BASELINES

For the comprehensive and comparative analysis of GRED, we compare it with the following methods:

##### a: RULE-BASED METHODS

We adopt Distance-rule and Dependency-rule as our baselines. The distance-rule method utilizes distance and

Part-Of-Speech (POS) tags to extract *opinion words*. The dependency-rule method uses the dependency tree of the sentence to determine the *opinion words*.

##### b: DEEP NEURAL NETWORK (DNNs)-BASED METHODS

We choose BiLSTM, TC-BiLSTM, IOG [8], PE-BiLSTM and LOTN [41] for our DNN-based baseline methods. These methods employ the Recurrent Neural Networks (RNNs) to capture the dependency between the *opinion target* and its corresponding *opinion words*. IOG uses an Inward-Outward BiLSTM to learn *opinion target*-aware representation and global context representation and then fuses these representations to predict labels. LOTN utilizes the additional position embeddings for indicating *opinion targets* and transfers the latent opinion knowledge from resource-rich datasets to TOWE task model.

##### c: PRE-TRAINED LANGUAGE MODEL (PLM) BASED METHODS

We also adopt SDRN [4], ONG [36], ARGCN [13], TSMSA [10], RABERT [14] and UNI-GEN [43] for baselines. SDRN employs a BERT-based encoder with a target entity extraction network, an entity relation detection network, and a synchronization network for the Aspect Opinion Pair Extraction (AOPE) task. ONG incorporates the syntactic structures of the sentence into deep learning models using Graph Convolution Networks (GCNs). ARGCN consists of BiLSTM-based sequential layers and attention-based relational GCN layers to capture semantic and syntactic relation between words simultaneously. TSMSA uses a multi-head self-attention mechanism to specify *opinion target* in the sentence. RABERT integrates the relation network into BERT layer to capture the relationship between words. UNI-GEN converts all ABSA subtasks into a unified generative formulation and exploits BART to solve all these tasks.

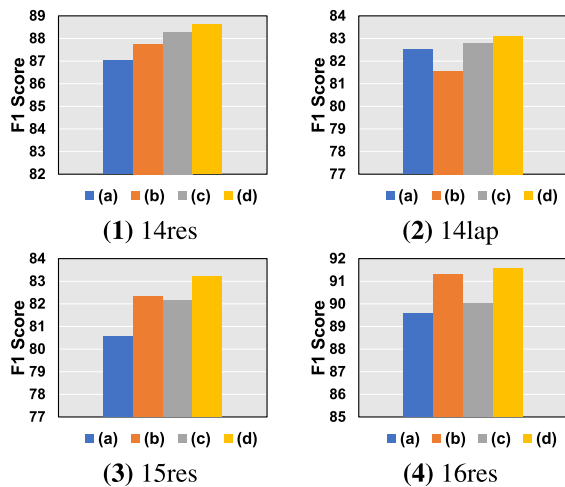
#### 3) HYPER PARAMETER SETTING

We implement our proposed GRED with the Pytorch library [25] and Hugging Face Transformers.<sup>1</sup> In our experiments, the batch size is set to 8, and the maximum sequence length is set to 128. We train the model using the *Adam* optimizer and learning rate decay strategy with  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . We also set the warmup steps to 100 steps. The dropout rate is selected from {0.3, 0.4, 0.5} based on the performances of the validation sets. The learning rate of the encoder layer and decoder layer is set to  $5 \times 10^{-5}$ . We set the learning rate of **TRN** and **LRN** to  $1 \times 10^{-4}$ . We adopt pretrained BART from [16], which consists of 12 layers for the encoder and the decoder, respectively. **TRN** and the gate network consist of 1-layer Feed-Forward Networks (FFNs). And the **LRN** is composed of two parallel 1-layer FFNs and one additional FFN. GRED could fit in i7-6850 64G

<sup>1</sup><https://huggingface.co/transformers>

**TABLE 4.** Main experimental results on the four benchmark datasets. Performance comparison with the baselines. The results of baselines are from their original papers and implementations [4], [8], [10], [13], [14], [36], [41], [43].

Models	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Distance-rule	58.39	43.59	49.92	50.13	33.86	40.42	54.12	39.96	45.97	61.90	44.57	51.83
Dependency-rule	64.57	52.72	58.04	45.09	31.57	37.14	65.49	48.88	55.98	76.03	56.19	64.62
LSTM	52.64	65.47	58.34	55.71	57.53	56.52	57.27	60.69	58.93	62.46	68.72	65.33
BiLSTM	58.34	61.73	59.95	64.52	61.45	62.71	60.46	63.65	62.00	68.68	70.51	69.57
Pipeline	77.72	62.33	69.18	72.58	56.97	63.83	74.75	60.65	66.97	81.46	67.81	74.01
TC-BiLSTM	67.65	67.67	67.61	62.45	60.14	61.21	66.06	60.16	62.94	73.46	72.88	73.10
I0G	82.38	78.25	80.23	73.43	68.74	70.99	72.19	71.76	71.91	84.36	79.08	81.60
PE-BiLSTM	80.10	76.51	78.26	72.01	64.20	67.83	70.36	65.73	67.96	82.27	74.95	78.43
LOTN	84.00	80.52	82.21	77.08	67.62	72.02	76.61	70.29	73.29	86.57	80.89	83.62
SDRN	<b>91.14</b>	76.37	83.10	84.37	65.42	73.69	83.57	70.33	76.38	91.13	80.34	85.40
ONG	83.23	81.46	82.33	73.87	77.78	75.77	76.63	81.14	78.81	87.72	84.38	86.01
ARGCN	87.32	83.59	85.42	75.83	76.90	76.36	78.81	77.69	78.24	88.49	84.95	86.69
TSMSA	86.67	86.50	86.59	83.76	80.07	81.88	81.60	82.76	82.18	88.72	89.90	89.31
RABERT	87.82	88.93	88.37	83.33	79.37	81.30	<b>85.68</b>	80.12	82.81	90.49	90.67	90.58
UNI-GEN	86.01	84.76	85.38	83.11	78.13	80.55	80.12	80.93	80.52	89.22	86.67	87.92
GRED	88.10	<b>89.13</b>	<b>88.61</b>	<b>84.24</b>	<b>82.01</b>	<b>83.11</b>	82.05	<b>84.38</b>	<b>83.20</b>	<b>91.29</b>	<b>91.81</b>	<b>91.55</b>

**FIGURE 4.** Ablation study of GRED. We report the F1 score of the variants of GRED on the four benchmark datasets.

CPU and a single NVIDIA GTX 1080ti GPU. For a fair comparison with the baselines, we randomly sample 20% of the train set as the dev set using the same random seeds as in [8].

#### 4) EVALUATION METRICS

To maintain consistency with previous works [8], [10], [14], [41], we adopt precision, recall, and F1 score as the evaluation metrics to compare the performance of the models. As in [8], we consider the predicted *opinion words* spans to be correct when their starting and ending points are equal to those of the golden spans.

**TABLE 5.** The F1 scores of various pooling methods of *opinion targets* on the four benchmark datasets.

Methods	Datasets				
	14res	14lap	15res	16res	Avg
Max-pool	87.83	81.25	81.50	91.27	85.46
Mean-pool	<b>88.97</b>	81.94	<b>83.33</b>	90.65	86.22
LSE-pool	88.61	<b>83.11</b>	83.20	<b>91.55</b>	<b>86.62</b>

## B. EXPERIMENTATION RESULTS

### 1) PERFORMANCE COMPARISON ON THE BENCHMARK DATASETS

Here, we focus on the TOWE task performance comparison between the proposed GRED and the existing models on the four benchmark datasets. All the experimental results are reported in Table 4. The best scores are highlighted in bold style. First, compared with the other baselines, the proposed GRED obtains superior performance and achieves new state-of-the-art performance on all of the datasets. In detail, we find that GRED outperforms the state-of-the-art scores by 0.24%, 1.71%, 0.39%, and 0.97% for F1 scores on the four datasets. These results validate the effectiveness of the proposed GRED for TOWE task.

For the results of rule-based methods, the dependency rule performs better than the distance rule, which indicates that word dependency is critical to solving TOWE task. And we note that both rule-based methods (Distance-rule and Dependency-rule) show poor performance across all the scores. This reveals that the rules cannot handle the diverse patterns of TOWE task. On the other hand, DNN-based methods achieve relatively better performance than rule-based methods because of their model

**TABLE 6.** Case study: the prediction of Encoder+Decoder model and GRED. The red-colored words and blue-colored words represent *opinion targets* and *opinion words*, respectively.

<b>Sentence 1:</b> They have <b>great rolls</b> , the triple color and norwegetan rolls , are awesome and filling .
<b>Encoder+Decoder:</b> They have <b>great rolls</b> , the triple color and norwegetan rolls , are awesome and filling . ✓
<b>GRED:</b> They have <b>great rolls</b> , the triple color and norwegetan rolls , are awesome and filling . ✓
<b>Sentence 2:</b> I did swap out the hard drive for a <b>Samsung 830 SSD</b> which I highly <b>recommend</b> .
<b>Encoder+Decoder:</b> I did <b>swap out</b> the hard drive for a <b>Samsung 830 SSD</b> which I <b>highly recommend</b> . ✗
<b>GRED:</b> I did swap out the hard drive for a <b>Samsung 830 SSD</b> which I highly <b>recommend</b> . ✓
<b>Sentence 3:</b> My daughter 's wedding reception at <b>Water 's Edge</b> received the <b>highest compliments</b> from our guests .
<b>Encoder+Decoder:</b> My daughter 's wedding reception at <b>Water 's Edge</b> received the <b>highest compliments</b> from our guests . ✗
<b>GRED:</b> TMy daughter 's wedding reception at <b>Water 's Edge</b> received the <b>highest compliments</b> from our guests . ✓
<b>Sentence 4:</b> Then the <b>system</b> would many times <b>not power down</b> without a forced power-off .
<b>Encoder+Decoder:</b> Then the <b>system</b> would many times <b>not power down without</b> a forced power-off . ✗
<b>GRED:</b> Then the <b>system</b> would many times <b>not power down</b> without a forced power-off . ✓
<b>Sentence 5:</b> The <b>veal</b> and the mushrooms were cooked <b>perfectly</b> .
<b>Encoder+Decoder:</b> The <b>veal</b> and the mushrooms were <b>cooked perfectly</b> . ✗
<b>GRED:</b> The <b>veal</b> and the mushrooms were <b>cooked perfectly</b> . ✗
<b>Sentence 6:</b> this Mac Mini does not have a built-in mic , and it would seem that its <b>Mac OS 10.9</b> does <b>not handle</b> external microphones properly .
<b>Encoder+Decoder:</b> this Mac Mini does not have a built-in mic , and it would seem that its <b>Mac OS 10.9</b> does <b>not handle</b> external microphones <b>properly</b> . ✗
<b>GRED:</b> this Mac Mini does not have a built-in mic , and it would seem that its <b>Mac OS 10.9</b> does <b>not handle</b> external microphones <b>properly</b> . ✗

The hot dogs are **top notch** , and they 're Slamwich is **amazing** !

The hot dogs are **top notch** , and they 're Slamwich is **amazing** !

The have a **great** cocktail with Citrus Vodka and lemon and lime juice and mint leaves that is to **die for** !

**Not** much of a selection of bottled beer either , we went with **Brahma** .

**FIGURE 5.** Visualization of the gate weight  $\alpha_u$  of each word. The *opinion targets* of each sentence are "hot dogs", "Slamwich", "cocktail" and "selection of bottled beer", respectively. The extracted *opinion words* of each sentence are "top notch", "amazing", {"great", "die", "for"}, and {"Not", "much"}. The stronger the red words, the greater the value of  $\alpha_u$ .

expressive power. Next, we observe that IOG, PE-BiLSTM, and LOTN obtain an F1-score with an approximately 25% improvement over the LSTM and BiLSTM methods, which demonstrates the effectiveness of using *opinion target* information. Thus, all of these results reveal that capturing both word dependency and *opinion target* is important for TOWE.

Finally, pretrained language model-based methods achieve great performance on all the scores. In particular, RABERT and TSMSA show relatively better performance than other pretrained language model-based methods such as SDRN, ONG, ARGCN, and UNI-GEN. These experimental results demonstrate that the target-indicating token can be effective in fully utilizing the pre-trained language model for TOWE task. Additionally, RABERT achieves the previous state-of-the-art performance. This validates the effectiveness of capturing target information and local context information together.

## 2) ABLATION STUDY

To investigate the effectiveness of each part of GRED, we evaluate the variants of GRED: (a) **Encoder only**: only using the target-aware encoder to predict labels, (b) **Decoder only**: only feeding the output of the decoder to compute the final representation, (c) **Encoder+Decoder**: uniformly combining the outputs of the encoder and the decoder without the gate network, and (d) **GRED**.

The results of the ablation study are depicted in Fig. 4. First, Encoder only model is inferior to the other models on most of the datasets. This inferiority demonstrates that using only *opinion target* information cannot extract the diverse patterns of *opinion words* in TOWE task. Second, naive aggregation of the encoder and the decoder does not ensure performance improvement. The performance drops of Encoder+Decoder at *15res* and *16res* confirm the suitability of our proposed gate mechanism. Overall, GRED shows the best performance compared to all the other models on all



the datasets. This high performance reveals that all three components of GRED are critical to solving TOWE: 1) capturing the *opinion words*, 2) utilizing the local context, and 3) dynamically leveraging by our gate mechanism.

Next, we also report the performance of various *opinion target* pooling methods in Table 5: Max-pool, Mean-pool, and LSE-pool from Equation 2. We observe that the Max-pooling method shows relatively poor performance, but the Mean-pool and the LSE-pool perform with almost identical performance, which indicates that the Max-pool can lose more target information than the other methods. Comparing the Mean-pool and the LSE-pool, the performance of the LSE-pool is more robust in the four benchmark datasets. In particular, the performance of LSE-pool is good at 14lap, which uses more complicated words than the other datasets. Thus, we adopt the LSE-pool for the target pooling method of GRED in this work.

### 3) CASE STUDY

In order to validate the effectiveness of our proposed GRED, we extract some TOWE examples of GRED and Encoder+Decoder model from the two different domains, restaurant and laptop (Table 6). In a simple case such as Sentence 1, we can observe that both GRED and Encoder+Decoder model provide the correct prediction for the given sentence and *opinion targets*. However, in Sentence 2, Sentence 3, and Sentence 4 which are more complicated than Sentence 1, only GRED can successfully extract the correct opinion terms. In sentence 3, Encoder+Decoder cannot extract the multiple *opinion words*. We also note that, even in the wrong prediction examples, Sentence 5 and Sentence 6, GRED gives the prediction closer to the gold label answers than Encoder+Decoder model. In sentence 5, Encoder+Decoder only extracts *cooked* as the *opinion words* but GRED identifies *perfectly* as the *opinion words*. These results demonstrate that the proposed GRED can leverage the gate network to improve the effectiveness of complicated patterns of TOWE task.

We also visualize the weight of the gate network  $\alpha_u$  to investigate its role in predicting labels in Fig. 5. In the first example, the value of  $\alpha_u$  increases at the word “*top notch*”, which corresponds to *opinion words* for “*hot dogs*”. However, in the second example, the highest value of  $\alpha_u$  is at the word “*amazing*”. These results indicate that the target-aware encoder mainly focuses on *opinion target*-dependent words. In the third example, we can observe that the value of  $\alpha_u$  increases in the words, “*great*”, “*die*”, and “*for*”, which indicates that the gate network can capture multiple *opinion words* simultaneously. In particular, comparing the words “*die for*” and “*Not much*” in the third and fourth examples, and the words “*top notch*” in the first example, the values of  $\alpha_u$  at “*much*” and “*for*” are lower than those at “*die*”, “*Not*” and “*top notch*”. Since the *opinion words* “*for*” and “*much*” are dependent on the surrounding words “*die*” and “*Not*” respectively as well as their corresponding *opinion targets*, the gate network reduces the influence of the target-aware

encoder. Therefore, these results demonstrate that the gate network can effectively regulate the target-aware encoder and the local context-aware decoder.

## V. CONCLUSION

In this paper, we propose a novel and powerful transformer-based encoder-decoder model named GRED for target-oriented *opinion words* extraction. In GRED, its target-aware encoder and local context-aware decoder can capture the *opinion target* information and the local context information, respectively, which is crucial for addressing TOWE task. Then, the gate network dynamically combines the outputs of the encoder and the decoder to predict the label sequences. Therefore, GRED can flexibly utilize the *opinion target* and local context information based on the gate mechanism. Furthermore, GRED can enhance its performance by incorporating language knowledge of the pretrained language model, BART. To validate our proposed GRED, we conduct extensive experiments on the four widely used benchmark datasets. Our GRED outperforms all the baseline methods and achieves state-of-the-art performance. Additionally, in-depth analyses and qualitative studies demonstrate that the gate network appropriately adjusts the influence of the target-aware encoder and local context-aware decoder to identify *opinion words*. As a result, GRED effectively solves the TOWE task by dynamically utilizing *opinion target* and local context information.

In future work, we plan to design more efficient modules to capture the invaluable information in a sentence and investigate more informative relationships for TOWE task. In addition, we desire to use large-scale language models such as T5 [31] and GPT3 [3] for GRED. Also, we would like to apply GRED to the other datasets that can be more helpful to the public interests, like COVID-19 datasets [1], [24] Furthermore, we attempt to use GRED for the entire space as in [19] to validate the generalization performance of GRED.

## REFERENCES

- [1] A. M. Almars, E.-S. Atlam, T. H. Noor, G. Elmarhomy, R. Alagamy, and I. Gad, “Users opinion and emotion understanding in social media regarding COVID-19 vaccine,” *Computing*, vol. 104, pp. 1–16, Feb. 2022.
- [2] S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky, “Automatic extraction of opinion propositions and their holders,” in *Proc. AAAI*, vol. 2224, 2004, pp. 1–8.
- [3] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, and S. Agarwal, “Language models are few-shot learners,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.
- [4] S. Chen, J. Liu, Y. Wang, W. Zhang, and Z. Chi, “Synchronous double-channel recurrent network for aspect-opinion pair extraction,” in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 6515–6524.
- [5] I. Chetviorkin and N. Loukachevitch, “Extraction of domain-specific opinion words for similar domains,” in *Proc. Workshop Inf. Extraction Knowl. Acquisition*, 2011, pp. 7–12.
- [6] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, “ELECTRA: Pre-training text encoders as discriminators rather than generators,” 2020, *arXiv:2003.10555*.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2019, pp. 4171–4186.

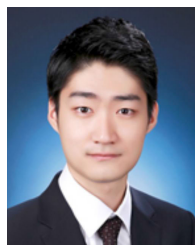
- [8] Z. Fan, Z. Wu, X.-Y. Dai, S. Huang, and J. Chen, "Target-oriented opinion words extraction with target-fused neural sequence labeling," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2019, pp. 2509–2518.
- [9] H. Fei, F. Li, C. Li, S. Wu, J. Li, and D. Ji, "Inheriting the wisdom of predecessors: A multiplex cascade framework for unified aspect-based sentiment analysis," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, 2022, pp. 4096–4103.
- [10] Y. Feng, Y. Rao, Y. Tang, N. Wang, and H. Liu, "Target-specified sequence labeling with multi-head self-attention for target-oriented opinion words extraction," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2021, pp. 1805–1815.
- [11] S. S. Htay and K. T. Lynn, "Extracting product features and opinion words using pattern knowledge in customer reviews," *Sci. World J.*, vol. 2013, pp. 1–5, Dec. 2013.
- [12] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 168–177, doi: [10.1145/1014052.1014073](https://doi.org/10.1145/1014052.1014073).
- [13] J. Jiang, A. Wang, and A. Aizawa, "Attention-based relational graph convolutional network for target-oriented opinion words extraction," in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics, Main Volume*, 2021, pp. 1986–1997.
- [14] T. Kang, M. Lee, N. Yang, and K. Jung, *RABERT: Relation-Aware BERT for Target-Oriented Opinion Words Extraction*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 3127–3131, doi: [10.1145/3459637.3482165](https://doi.org/10.1145/3459637.3482165).
- [15] A. Kumar, A. S. Veerubhotla, V. T. Narapareddy, V. Aruru, L. B. M. Neti, and A. Malapati, "Aspect term extraction for opinion mining using a hierarchical self-attention network," *Neurocomputing*, vol. 465, pp. 195–204, Nov. 2021.
- [16] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," 2019, *arXiv:1910.13461*.
- [17] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Jul. 2018, pp. 946–956.
- [18] X. Li and W. Lam, "Deep multi-task learning for aspect term extraction with memory interaction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2886–2892.
- [19] Y. Li, F. Wang, and S.-H. Zhong, "Training entire-space models for target-oriented opinion words extraction," 2022, *arXiv:2204.07337*.
- [20] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [21] K. Liu, H. L. Xu, Y. Liu, and J. Zhao, "Opinion target extraction using partially-supervised word alignment model," in *Proc. IJCAI*, vol. 13, Aug. 2013, pp. 2134–2140.
- [22] S. Mensah, K. Sun, and N. Aletras, "An empirical study on leveraging position embeddings for target-oriented opinion words extraction," 2021, *arXiv:2109.01238*.
- [23] A. Mukherjee and B. Liu, "Aspect extraction through semi-supervised modeling," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Long Papers*, vol. 1, Jul. 2012, pp. 339–348.
- [24] T. H. Noor, A. Almars, I. Gad, E.-S. Atlam, and M. Elmezain, "Spatial impressions monitoring during COVID-19 pandemic using machine learning techniques," *Computers*, vol. 11, no. 4, p. 52, Mar. 2022.
- [25] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. NIPS*, 2017, pp. 1–4.
- [26] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 5, pp. 8600–8607.
- [27] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 Task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Eval.*, 2015, pp. 486–495. [Online]. Available: <https://www.aclweb.org/anthology/S15-2082>
- [28] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clecq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. Jiménez-Zafra, and G. Eryigit, "SemEval-2016 Task 5: Aspect based sentiment analysis," in *Proc. Int. Workshop Semantic Eval.*, Jun. 2016, pp. 19–30.
- [29] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval.* Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 27–35. [Online]. Available: <https://www.aclweb.org/anthology/S14-2004>
- [30] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016.
- [31] K. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2020.
- [32] L. Ramshaw and M. Marcus, "Text chunking using transformation-based learning," in *Proc. 3rd Workshop Very Large Corpora*, 1995, pp. 157–176. [Online]. Available: <https://aclanthology.org/W95-0107>
- [33] I. Shamsurhin, "Extracting domain-specific opinion words for sentiment analysis," in *Proc. Mexican Int. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2012, pp. 58–68.
- [34] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.
- [35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [36] A. P. B. Veyseh, N. Nouri, F. Dernoncourt, D. Dou, and T. H. Nguyen, "Introducing syntactic structures into target opinion word extraction with deep learning," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 8947–8956.
- [37] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.
- [38] H. Wan, Y. Yang, J. Du, Y. Liu, K. Qi, and J. Z. Pan, "Target-aspect-sentiment joint detection for aspect-based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 5, pp. 9122–9129.
- [39] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Recursive neural conditional random fields for aspect-based sentiment analysis," 2016, *arXiv:1603.06679*.
- [40] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1–7.
- [41] Z. Wu, F. Zhao, X.-Y. Dai, S. Huang, and J. Chen, "Latent opinions transfer network for target-oriented opinion words extraction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 5, pp. 9298–9305.
- [42] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Double embeddings and CNN-based sequence labeling for aspect extraction," 2018, *arXiv:1805.04601*.
- [43] H. Yan, J. Dai, X. Qiu, and Z. Zhang, "A unified generative framework for aspect-based sentiment analysis," 2021, *arXiv:2106.04300*.
- [44] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang, and M. Zhou, "Unsupervised word and dependency path embeddings for aspect term extraction," 2016, *arXiv:1605.07843*.
- [45] L. Zhang and B. Liu, "Aspect and entity extraction for opinion mining," in *Data Mining and Knowledge Discovery for Big Data*. Berlin, Germany: Springer, 2014, pp. 1–40.
- [46] M. Zhang, Y. Zhang, and D.-T. Vo, "Gated neural networks for targeted sentiment analysis," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
- [47] L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in *Proc. 15th ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*, 2006, pp. 43–50, doi: [10.1145/1183614.1183625](https://doi.org/10.1145/1183614.1183625).



**TAEGWAN KANG** received the B.S. degree in electrical and computer engineering from Seoul University, South Korea, in 2016. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Seoul National University, South Korea. His research interest includes natural language processing.



**SEGWANG KIM** received the B.S. degree in mathematics from Seoul National University, Seoul, South Korea, in 2016, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interest includes systematic generalization.



**HWANHEE LEE** received the B.S. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2017, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. His research interests include text generation, text evaluation, and language grounding with vision.



**HYEONGU YUN** received the B.S. degree in electrical and computer engineering from Seoul University, South Korea, in 2015. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Seoul National University, South Korea. His research interests include neural sequence generation and question answering systems.



**KYOMIN JUNG** received the graduate degree from the Seoul Science High School, in February 1996, the B.Sc. degree from Seoul National University, in August 2003, and the Ph.D. degree from MIT, in June 2009. He worked at KAIST Computer Science Department, from 2009 to 2013, where he had joint appointments at the Department of Electrical Engineering and the Department of Mathematics. He is a Professor with the Electrical and Computer Engineering Department and an Adjunct Professor with the Department of Mathematical Sciences, Seoul National University. Currently, he is the Vice-Chair of the ECE Department for student affairs. His research interests include natural language processing, deep learning and applications, data analysis, and web services.

...