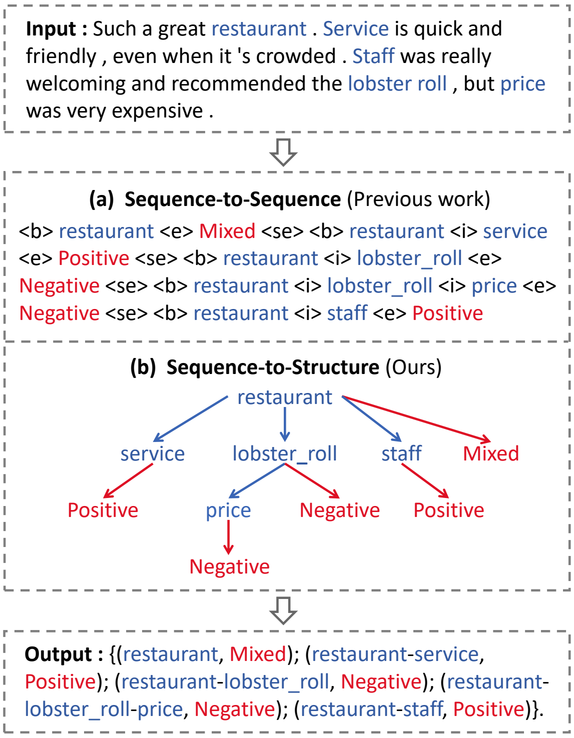
# A Sequence-to-Structure Approach to Document-level Targeted Sentiment Analysis

**Nan Song [1*], Hongjie Cai [1*], Rui Xia [1†], Jianfei Yu [1], Zhen Wu [2], Xinyu Dai [2]**
[1] School of Computer Science and Engineering,
Nanjing University of Science and Technology, China
[2] National Key Laboratory for Novel Software Technology, Nanjing University, China
{nsong, hjcai, rxia, jfyu}@njust.edu.cn
{wuz, daixinyu}@nju.edu.cn

## Abstract

Most previous studies on aspect-based sentiment analysis (ABSA) were carried out at the sentence level, while the research of document-level ABSA has not received enough attention. In this work, we focus on the document-level targeted sentiment analysis task, which aims to extract the opinion targets consisting of multi-level entities from a review document and predict their sentiments. We propose a Sequence-to-Structure (Seq2Struct) approach to address the task, which is able to explicitly model the hierarchical structure among multiple opinion targets in a document, and capture the long-distance dependencies among affiliated entities across sentences. In addition to the existing Seq2Seq approach, we further construct four strong baselines with different pretrained models. Experimental results on six domains show that our Seq2Struct approach outperforms all the baselines significantly. Aside from the performance advantage in outputting the multi-level target-sentiment pairs, our approach has another significant advantage - it can explicitly display the hierarchical structure of the opinion targets within a document. Our source code is publicly released at https://github.com/NUSTM/Doc-TSA-Seq2Struct.



**Input :** Such a great restaurant . Service is quick and friendly , even when it 's crowded . Staff was really welcoming and recommended the lobster roll , but price was very expensive .

**(a) Sequence-to-Sequence** (Previous work)

<b> restaurant <e> Mixed <se> <b> restaurant <i> service <e> Positive <se> <b> restaurant <i> lobster_roll <e> Negative <se> <b> restaurant <i> lobster_roll <i> price <e> Negative <se> <b> restaurant <i> staff <e> Positive

**(b) Sequence-to-Structure** (Ours)

**Output :** {(restaurant, Mixed); (restaurant-service, Positive); (restaurant-lobster_roll, Negative); (restaurant-lobster_roll-price, Negative); (restaurant-staff, Positive)}.

Figure 1: Comparison of two different approaches for the document-level TSA task. Text chunks in blue represent flat entities, and multi-level entities are connected with "-" to form an opinion target. Text chunks in red indicate the sentiment polarities of opinion targets.

## 1 Introduction

Aspect-based sentiment analysis (ABSA) has received wide attention in NLP for nearly two decades (Hu and Liu, 2004). Most of the previous studies have focused on sentence-level ABSA. However, a review text often consists of multiple sentences, and the opinion targets expressed in these sentences are often interrelated. Conducting sentence-level ABSA on each individual sentence cannot capture the interrelated opinion targets in the entire document. In comparison, document-level ABSA is more suitable for practical applications, yet it has not received enough attention. Only a limited amount of work attempted to identify the sentiments towards all aspect categories in a document (e.g., the Food Quality category in a Restaurant domain), without involving explicit opinion target terms (e.g., entities or aspects) (Titov and McDonald, 2008; Pontiki et al., 2016; Li et al., 2018; Wang et al., 2019; Bu et al., 2021).

Recently, Luo et al. (2022) introduced a new task called document-level Targeted Sentiment Analysis (document-level TSA), aiming to discover the opinion target consisting of multi-level entities (or aspects) in a review document, and predict the sentiment polarity label (Positive, Negative

[*] Equal contribution.
[†] Corresponding author.

or Mixed) towards the target. In sentence-level ABSA, an opinion target is usually a single entity or aspect. While in document-level TSA, it often involves multiple entities and aspects throughout the review document, and their relation is affiliated rather than flat. As shown in Figure 1, "*restaurant - lobster_roll - price*" is an opinion target consisting of three-level entities[1], indicating the price of the lobster roll sold in the restaurant. Luo et al. (2022) accordingly proposed a sequence-to-sequence (Seq2Seq) framework to solve the task. By using BART as the backbone, they took the review document as the input, and output a sequence indicating a set of target and sentiment tuples. For example, the tuple "<*b*> *restaurant* <*i*> *lobster_roll* <*i*> *price* <*e*> *Negative* <*se*>" denotes that the multi-level opinion target is "*restaurant - lobster_roll - price*" and its corresponding sentiment is Negative.

As shown in Figure 1(b), there is actually a hierarchical structure among multiple opinion targets in the document: the first layer is the "*restaurant*" entity, the second layer contains three entities (or aspects) affiliated to "*restaurant*" ("*service*", "*lobster_roll*" and "*staff*"), and the third layer is the "*price*" aspect of "*lobster_roll*". Although the Seq2Seq method appears simple and straightforward, it is imperfect to model such complex hierarchical structure. On one hand, it outputs the structural information by a sequence of tuples, where the previous tuples affects the generation of subsequent ones. On the other hand, the inherent encoder-decoder architecture is less flexible and effective to model the long-distance dependencies among affiliated entities/aspects across sentences.

To address the aforementioned issues, we propose a Sequence-to-Structure (Seq2Struct) approach in this work for the document-level TSA task. Our approach still takes a document as the input, but the output is no longer a sequence, but a structure as shown in Figure 1(b). It is a hierarchical structure with multiple layers of related entities, where each entity is assigned a predicted sentiment polarity. Seq2Struct contains four mains steps. We firstly identify flat opinion entities and their sentiments from the document. Secondly, we propose a multi-grain graphical model based on graph convolutional network (GCN), to better learn the semantic relations between document, sentences and entities. Thirdly, we employ a table-filling method to identify the affiliation relations in flat entities and consequently get the hierarchical opinion target structure. We finally incorporate the sentiments of the flat entities into the hierarchical structure and parse out the target-sentiment pairs as defined in (Luo et al., 2022).

We evaluate our approach on the document-level TSA dataset containing six domains. In addition to the existing approach, we further construct four strong baselines with different pretrained models. The experimental results show that our Seq2Struct approach outperforms all the baselines significantly on average F1 of the target-sentiment pairs. Aside from the advantage of performance, our approach can further explicitly display the hierarchical structure of the opinion targets in a document. We also make in-depth discussions from the perspectives of document length, the number of levels in opinion target, etc., verifying the effectiveness of our approach in capturing the long-distance dependency among across-sentence entities in document-level reviews.

## 2 Task Description

In document-level TSA task, an opinion target often consists of multi-level entities with affiliated relations (e.g., "*restaurant - lobster_roll - price*"). We call it "multi-level opinion target" in contrast to the opinion target at the sentence level. A document normally contain a set of multi-level opinion targets, which constitute a hierarchical structure as shown in Figure 1(b).

Similar as (Luo et al., 2022), we formulate document-level TSA as a task to detect a set of target-sentiment pairs from a document $\mathcal{D} = [x_1, x_2, ..., x_N]$ with $N$ tokens:

$$\mathcal{P} = \{(t, s)_i\}_{i=1}^{|\mathcal{P}|}, \qquad (1)$$

where $t = e_1$-$e_2$-$\ldots$-$e_m$ is a multi-level opinion target with $m$ denoting the number of its levels, and $s \in \{$Positive, Negative, Mixed$\}$ is the corresponding sentiment.

The document-level TSA task is challenging, as the opinion target consisting of multi-level entities, and a predicted opinion target is considered to be correct if and only if entities at all levels match the ground-truth exactly. For instance, "*restaurant-lobster-price*" is the correct target only if the first, second, and third levels are predicted as "*restaurant*","*lobster*", and "*price*", respectively.

[1] For the convenience of description, in the following we collectively refer to both "entity" (e.g. "*lobster_roll*") and "aspect" (e.g. "*price*") as "entity".

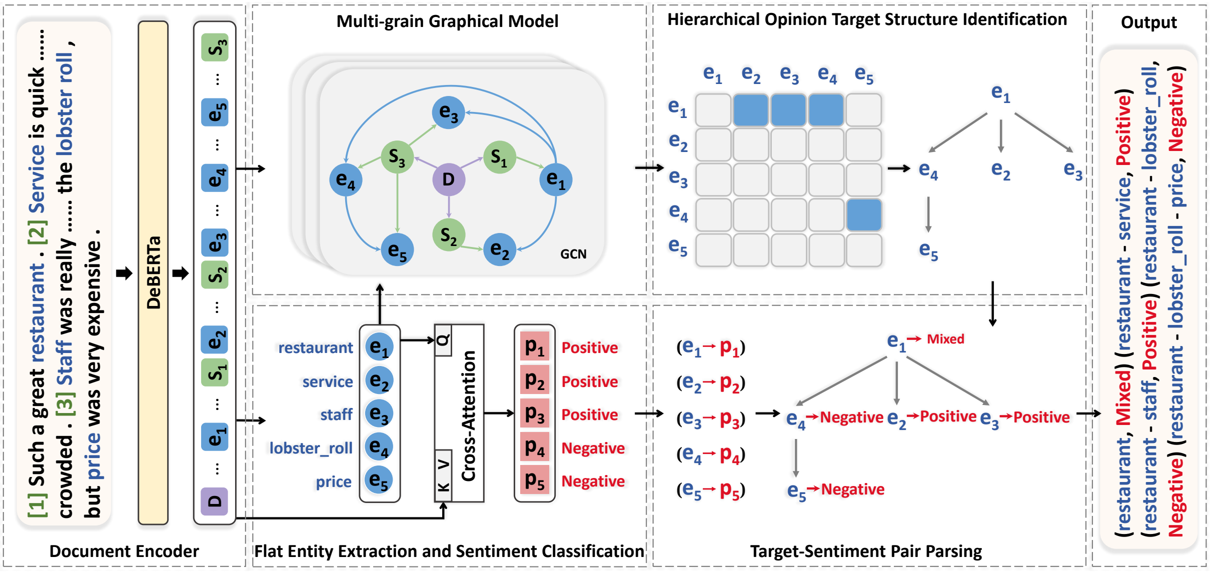

[1] Such a great restaurant . [2] Service is quick ..... crowded . [3] Staff was really ...... the lobster roll , but price was very expensive .
DeBERTa
D
e1
S1
e2
S2
e3
e4
e5
S3
Document Encoder
Multi-grain Graphical Model
GCN
restaurant
service
staff
lobster_roll
price
Q
K V
Cross-Attention
p1 Positive
p2 Positive
p3 Positive
p4 Negative
p5 Negative
Flat Entity Extraction and Sentiment Classification
Hierarchical Opinion Target Structure Identification
(e1→ p1)
(e2→ p2)
(e3→ p3)
(e4→ p4)
(e5→ p5)
e1→ Mixed
e4→Negative
e2→Positive
e3→Positive
e5→Negative
Target-Sentiment Pair Parsing
Output
(restaurant, Mixed) (restaurant - service, Positive)
(restaurant - staff, Positive) (restaurant - lobster_roll, Negative)
(restaurant - lobster_roll - price, Negative)

Figure 2: The overall architecture of our Seq2Struct approach.

## 3 Approach

As shown in Figure 2, we propose a Sequence-to-Structure (Seq2Struct) approach to address the document-level TSA task, which consists of four main modules.

### 3.1 Flat Entity Extraction and Sentiment Classification

We adopt DeBERTa (He et al., 2021) as the encoder of the input document $\mathcal{D} = \{x_1, x_2, ..., x_N\}$:

$$\boldsymbol{H} = \text{DeBERTa}(x_1, ..., x_N), \tag{2}$$

where $\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, .., \boldsymbol{h}_N\}$ is context representation of $\mathcal{D}$.

We then employ the BIO (Begin, Inside, Outside) tagging scheme to extract flat entities [2] from $\mathcal{D}$:

$$p(y_i^e|x_i) = \text{Softmax}(\boldsymbol{h}_i \boldsymbol{W}_e), \tag{3}$$

where $y_i^e \in \{B, I, O\}$, $\boldsymbol{W}_e \in \mathbb{R}^{d_{model} \times 3}$ is model parameter, and $d_{model}$ is the dimension of the hidden representation of each token.

Let $\mathcal{E} = \{e_i\}_{i=1}^{|\mathcal{E}|}$ represent the extracted entity set, where $e_i = (x_{start}, ..., x_{end})$. The representation of $e_i$ is the mean pooling of its tokens $\boldsymbol{h}_i^e = \text{MeanPooling}(\boldsymbol{h}_{start}, ..., \boldsymbol{h}_{end})$. Furthermore, we perform sentiment prediction on the extracted flat entities. Specifically, for an entity $e_i$ in $\mathcal{E}$ , we utilize an entity-context cross-attention module to capture the context information:

$$\hat{\boldsymbol{h}}_i^e = \text{MultiHeadCrossAttention}(Q, K, V), \tag{4}$$

| Dataset | #Total | #Multi | #Cross | #Cross/#Multi |
|---|---|---|---|---|
| Books | 2470 | 1005 | 804 | 80.00% |
| Clothing | 1554 | 389 | 278 | 72.46% |
| Restaurant | 4739 | 2796 | 2466 | 88.19% |
| Hotel | 3436 | 2028 | 1572 | 77.51% |

Table 1: Statistics of opinion targets in four domains of Luo et al. (2022), where #Total means the total number of opinion targets, #Multi means the number of opinion targets with more than one level, and #Cross means the number of cross-sentence opinion targets.

where $Q = \boldsymbol{h}_i^e$, $K = \boldsymbol{H}$, $V = \boldsymbol{H}$. Then, $\hat{\boldsymbol{h}}_i^e$ is fed into the softmax layer to predict the sentiment towards $e_i$:

$$p(y_i^s|e_i) = \text{Softmax}(\hat{\boldsymbol{h}}_i^e \boldsymbol{W}_s), \tag{5}$$

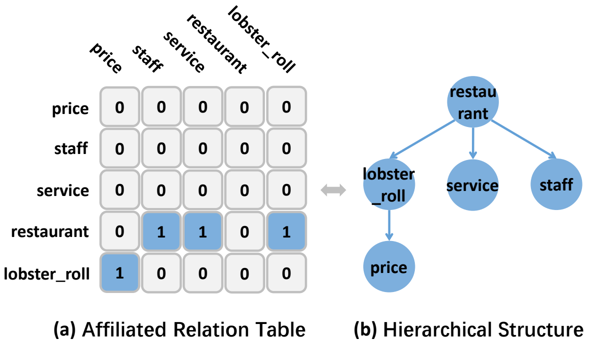
where $y_i^s \in \{\text{Positive, Negative, Mixed}\}$ and $\boldsymbol{W}_s \in \mathbb{R}^{d_{model} \times 3}$ is the model parameter.

We use $\mathcal{S} = \{(e, s)_i\}_{i=1}^{|\mathcal{S}|}$ to represent the predicted flat entity-sentiment pair set. For the example shown in Figure 2, $\mathcal{S}$ = {(“*restaurant*”, Positive), (“*price*”, Negative), (“*service*”, Positive), (“*staff*”, Positive), (“*lobster_roll*”, Negative)}.

### 3.2 Multi-grain Graphical Model

The affiliated entities in a multi-level opinion target often exist in multiple sentences. In Table 1, we report the number and proportion of across-sentence opinion targets in four domains of the dataset (Luo et al., 2022). It can be seen that the proportion of across-sentence opinion targets to all multi-level opinion targets reaches 80.0%, 72.5%, 88.2% and 77.5% in four domains respectively.

[2] Flat entities in this paper refer to all individual entities in the document, regardless of the hierarchy.



|  | price | staff | service | restaurant | lobster_roll |
|---|---|---|---|---|---|
| price | 0 | 0 | 0 | 0 | 0 |
| staff | 0 | 0 | 0 | 0 | 0 |
| service | 0 | 0 | 0 | 0 | 0 |
| restaurant | 0 | 1 | 1 | 0 | 1 |
| lobster_roll | 1 | 0 | 0 | 0 | 0 |

(a) Affiliated Relation Table (b) Hierarchical Structure

Figure 3: An illustration of Hierarchical Opinion Target Structure Identification.

To better model the affiliated relations and long-distance dependencies between different entities, we construct a multi-grain graphical model, to enhance entity representation learning.

The graph contains three different types of nodes: 1) Document node, 2) Sentence nodes, and 3) Entity nodes. These nodes are linked with three types of edges: 1) Document-to-Sentence Edges: The document node is linked to all sentence nodes; 2) Sentence-to-Entity edges: Each sentence node is linked to all entity nodes it contains; 3) Entity-to-Entity edges: We maintain a *global entity relation map* to capture affiliated relations between entities. As long as entity $e_k$ and entity $e_l$ were annotated as adjacent upper-level and lower-level entities (i.e., $e_l$ is affiliated with $e_k$) in any document of the training and validation corpus, $e_k$ is linked to $e_l$.

On this basis, we construct the adjacency matrix $\boldsymbol{A}$ where $A_{ij} = 1$ if the $i$-th node and $j$-th node have an edge, otherwise $A_{ij} = 0$. We then employ the Graph Convolutional Network (Kipf and Welling, 2017) to update the representation of nodes:

$$\boldsymbol{H}^e_{l+1} = \sigma(\boldsymbol{A}\boldsymbol{H}^e_l\boldsymbol{W}_l + \boldsymbol{b}_l), \tag{6}$$

where $l$ is the index of GCN sub-layers, $\boldsymbol{W}_l \in \mathbb{R}^{d_{model} \times d_{model}}$ and $\boldsymbol{b}_l \in \mathbb{R}^{d_{model}}$ are model parameters, $\sigma(\cdot)$ is an activation function, i.e., RELU.

### 3.3 Hierarchical Opinion Target Structure Identification

Till now, we have extracted the flat entities $\mathcal{E}$ and learned better entity representations $\tilde{\boldsymbol{H}}^e$. In this subsection, we further propose a table filling based method to identify the hierarchical structure among multiple opinion targets in a document.

Firstly, we construct a *Affiliated Relation Table* $\boldsymbol{T}$, as shown in Figure 3(a), based on the extracted flat entities from the input document. The row $e_i$ represents the upper-level entity, and the column $e_j$ represents the lower-level entity. We concatenate the representations of $e_i$ and $e_j$ as the representation of the cell $T_{ij}$: $\boldsymbol{h}^r_{ij} = [\tilde{\boldsymbol{h}}^e_i; \tilde{\boldsymbol{h}}^e_j]$, and then send it to a binary classifier to predict the affiliated relation:

$$p(y^r_{ij}|(e_i, e_j)) = \text{Softmax}(\boldsymbol{h}^r_{ij}\boldsymbol{W}_r), \tag{7}$$

where $\boldsymbol{W}_r \in \mathbb{R}^{2d_{model} \times 2}$ is the model parameter. $y^r_{ij} \in \{1, 0\}$ indicates the affiliated relation between $e_i$ and $e_j$. $T_{ij} = 1$ means that $e_j$ is affiliated with $e_i$.

Secondly, based on the predictions on each cell of $T$, we can finally obtain a hierarchical opinion target structure (a directed acyclic graph $\boldsymbol{G}$), as shown in Figure 3(b). The cell whose value is 1 constitutes an affiliated two-level entity (e.g., "*lobster_roll - price*"), and after decoding all cells on the entire table, we get the hierarchical structure.

Note that when the hierarchical structure contains a self-loop, we delete the edge with the smallest value of $p(y^t_{ij} = 1|(e_i, e_j))$ to ensure the output is a directed acyclic graph.

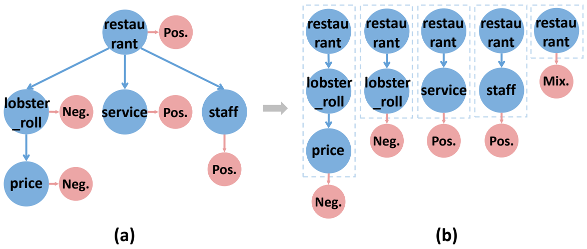


Figure 4: The process of Target-Sentiment Pair Parsing. For the left figure, each connected pair of blue and red nodes represents a flat entity and its corresponding sentiment, respectively. For the right figure, multi-level entities within the dashed box constitute an opinion target, and the red node indicates the updated sentiment polarities of opinion target.

### 3.4 Target-Sentiment Pair Parsing

In this section, we introduce a set of rules to parse out the target-sentiment pairs based on the predicted sentiments of flat entities and the hierarchical opinion target structure.

As shown in Figure 4(a), we firstly incorporate the flat entity sentiments obtained in Equation (5) to the hierarchical structure $\boldsymbol{G}$. Considering that the sentiment of lower-level entity should be embodied in the upper-level one, as the lower-level entities are part of the upper-level entities, we then introduce

| Dataset | Train |  | Dev |  | Test |  | #Sentence | #1-T | #2-T | #3-T |
|---|---|---|---|---|---|---|---|---|---|---|
|  | #D | #P | #D | #P | #D | #P |  |  |  |  |
| Books | 690 | 1682 | 99 | 287 | 197 | 501 | 5.97 | 1465 | 988 | 17 |
| Clothing | 649 | 1100 | 92 | 150 | 186 | 304 | 3.29 | 1165 | 385 | 4 |
| Restaurant | 658 | 3220 | 94 | 527 | 188 | 992 | 7.91 | 1943 | 2566 | 221 |
| Hotel | 720 | 2448 | 103 | 315 | 206 | 673 | 4.19 | 1408 | 1795 | 231 |
| News | 656 | 2334 | 93 | 351 | 187 | 644 | 7.25 | 2599 | 675 | 54 |
| PhraseBank | 835 | 1415 | 119 | 202 | 240 | 434 | 1.02 | 1413 | 589 | 49 |

Table 2: Dataset statistics. #D, #P and #Sentence respectively denote the number of documents, target-sentiment pairs and average number of sentences in each domain. #1-T, #2-T, #3-T denote the number of single-level targets, two-level targets and three-level targets respectively.

Algorithm 1 to update the sentiments of entities in the hierarchical structure. Specifically, as shown in line 1, we traverse the hierarchical opinion target structure $\boldsymbol{G}$ to obtain the path set $\boldsymbol{P}$ from the root node entity to the leaf node entity. For each path $p_i$, if the sentiment of the upper entity is conflict with that of its lower entity, the sentiment of the upper entity will be updated to "Mixed", as shown in lines 2 to 10. Finally, we traverse this structure to output a set of (multi-level) target-sentiment pairs, as shown in Figure 4(b).

**Algorithm 1** Multi-level Entity Sentiment Updating

```
Input: The predicted flat entity-sentiment pair set
    $\mathcal{S}$ and hierarchical opinion target structure $\boldsymbol{G}$
Output: The updated sentiment set for each entity
    in $\boldsymbol{G}$, denoted as $\hat{\mathcal{S}}$
 1: Traverse $\boldsymbol{G}$ to obtain the path set $\boldsymbol{P}$
 2: for path $p_i = \{e_u, ..., e_j, ..., e_{j+k}, ..., e_l\} \in \boldsymbol{P}$
    do
 3:   for the upper-level entity $e_j \in p_i$ do
 4:     for the lower-level entity $e_{j+k} \in p_i$ do
 5:       if $\mathcal{S}(e_j)$ is not equal to $\mathcal{S}(e_{j+k})$ then
 6:         $\hat{\mathcal{S}}(e_j) =$ Mixed
 7:       end if
 8:     end for
 9:   end for
10: end for
11: return $\hat{\mathcal{S}}$
```

### 3.5 Model Training

Our approach is a multi-task learning framework of three components. We employ cross-entropy of the ground truth and the prediction as our loss function for each component, and learn them jointly.

The loss functions for 1) flat entity extraction, 2) flat entity sentiment classification, and 3) entity affiliated relation prediction are:

$$\mathcal{L}_e = -\sum_{t}^{|\mathcal{T}|}\sum_{i}^{N}\hat{p}(y_i^e|x_i)\log p(y_i^e|x_i), \quad (8)$$

$$\mathcal{L}_s = -\sum_{t}^{|\mathcal{T}|}\sum_{i}^{|\mathcal{E}|}\hat{p}(y_i^s|e_i)\log p(y_i^s|e_i), \quad (9)$$

$$\mathcal{L}_r = -\sum_{t}^{|\mathcal{T}|}\sum_{i}^{|\mathcal{E}|}\sum_{j}^{|\mathcal{E}|}\hat{p}(y_{ij}^r|r_{ij})\log p(y_{ij}^r|r_{ij}), \quad (10)$$

where $\hat{p}$ is the golden one-hot distribution, $p$ is the predicted distribution, and $\mathcal{T}$ and $\mathcal{E}$ denote the example set and entity set, respectively.

The joint training loss is the sum of three parts:

$$\mathcal{L} = \mathcal{L}_e + \mathcal{L}_s + \mathcal{L}_r. \quad (11)$$

## 4 Experiments

### 4.1 Experimental Setup

#### 4.1.1 Dataset

We evaluated our model on the document-level TSA dataset, which includes product reviews, namely Book, Clothing, Restaurant, Hotel, as well as financial and social media reviews, namely News, PhraseBank. Note that we have slightly updated the annotations on PhraseBank and News, by dealing with the absence of sentiment annotations for upper-level entities. Table 2 presents detailed statistics for all six domains.

#### 4.1.2 Baselines and Evaluation Metrics

In addition to the Seq2Seq approach proposed by Luo et al. (2022), we further set up four strong baseline systems by adapting the newly proposed sentence-level ASBA approaches to document-level TSA.

BART-Extraction and T5-Extraction are adapted from the Extraction-based seq2seq approach in

| **Methods** | Books | Clothing | Restaurant | Hotel | News | PhraseBank | Average |
|---|---|---|---|---|---|---|---|
| Seq2Seq (Luo et al., 2022) | 34.76 | 49.40 | 19.08 | 34.17 | 12.91 | 55.27 | 34.27 |
| BART-Extraction | 33.83 | 55.42 | 33.05 | 58.90 | 21.80 | 63.15 | 44.36 |
| BART-Paraphrase | 32.90 | 55.18 | 33.21 | 59.71 | 21.47 | 63.08 | 44.26 |
| T5-Extraction | 32.66 | 52.49 | 32.85 | 57.92 | 22.31 | **65.48** | 43.95 |
| T5-Paraphrase | 32.64 | 53.47 | 33.36 | 57.95 | 22.96 | 64.85 | 44.21 |
| Seq2Struct | **38.41** | **57.36** | **36.41** | **60.10** | **23.47** | 63.18 | **46.49** |

Table 3: The main experimental results of our approach and five baselines on the six domains. Seq2Seq (Luo et al., 2022) represents the state-of-the-art approach in the document-level TSA task. We report the results in their paper. In addition, we construct the other four strong baselines as described in Subsection 4.1.2.

sentence-level ABSA (Zhang et al., 2021b), using BART and T5 as backbones respectively. BART-Paraphrase and T5-Paraphrase are adapted from the Paraphrase-based seq2seq approach in sentence-level ABSA (Zhang et al., 2021a), using BART and T5 as backbones respectively.

Following (Luo et al., 2022), we evaluate the document-level TSA task based on the output of a set of target-sentiment pairs given the input document. The precision and recall scores are calculated based on exact match of the predicted target-sentiment pairs and the ground-truth. The F1 score is taken as the final evaluation metric.

### 4.1.3 Implementation Details

We employ $\mathrm{DeBERTaV3}_{\mathrm{base}}$(He et al., 2021) as the backbone encoder of our approach, whose hidden size is 768 and maximum length of the input is 512. With a commitment to equitable model parameters, we have chose $\mathrm{T5}_{\mathrm{small}}$(Raffel et al., 2020) as the backbone of the baselines we designed in our paper. During training, the learning rate for fine-tuning the pre-trained language model is set to 3e-5, other learning rates are set to 5e-5, and the dropout rate is 0.1. We set batch size to 8 and training epochs to 30. We save the model parameters with the highest F1 value on the validation set. During testing, we report F1 score for each domain that are averaged over five different random seeds.

## 4.2 Main Results

In Table 3, we report the results of our approach and five baselines on the six domains. It can be observed that our method outperforms all the baselines on average F1. Furthermore, in comparison with the four strong baselines we designed, our approach can still achieve an average F1 score improvement larger than 2.1%. Specifically, the improvements are 4.58%, 1.94%, 3.05%, 0.39%, 0.51% on the Books, Clothing, Restaurant, Hotel and News, respectively, compared with the best results for all baselines. All the improvements are significant based on a paired $t$-test.

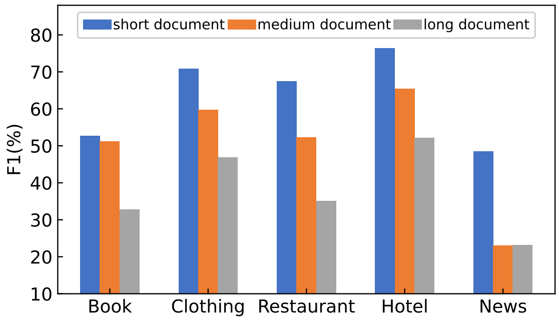


Figure 5: The performance of our approach on three types of document length.

Furthermore, it can be seen that our approach gains more improvement in the Books and Restaurant domains, which have longer document length and more across-sentence targets. This indicates the strength of our approach in identifying complex opinion targets from long documents.

An exception is that our approach on the PhraseBank domain is not optimal (slightly lower than T5-Extraction and T5-Paraphrase). According to Table 2, the document length of PhraseBank is the shortest in six domains and its average number of sentences is only 1.02. It is reasonable our approach does not show a significant advantage in this case.

## 4.3 The Impact of Document Length

We further investigate the performance of our approach on test subsets with different document lengths. Based on the number of sentences in the document, we divide the test set of each domain into subsets of short document (1 or 2 sentences), medium document (3 or 4 sentences), and long document (5 or more sentences).

As shown in Figure 5, in the Books, Clothing,

| Dataset | BART-Extraction | | | Seq2Struct | | |
|---|---|---|---|---|---|---|
| | 1-T | 2-T | 3-T | 1-T | 2-T | 3-T |
| Books | 47.92 | 24.68 | – | 54.61 | 35.23 | – |
| Clothing | 58.02 | 48.06 | – | 61.10 | 57.70 | – |
| Restaurant | 42.24 | 34.78 | 19.58 | 50.98 | 39.07 | 24.23 |
| Hotel | 68.54 | 61.48 | 42.20 | 78.57 | 64.45 | 45.32 |
| **Average** | 54.18 | 42.25 | 30.89 | 61.31 | 49.11 | 34.77 |

Table 4: The performance on different levels of targets, where "−" means that the corresponding domain does not have corresponding targets.

| **Dataset** | Seq2Struct | w/o GCN | Diff. |
|---|---|---|---|
| Books | 38.41 | 36.63 | 1.78↓ |
| Clothing | 57.36 | 56.74 | 0.62↓ |
| Restaurant | 36.41 | 35.84 | 0.58↓ |
| Hotel | 60.10 | 59.32 | 0.78↓ |
| News | 23.47 | 22.39 | 1.08↓ |
| PhraseBank | 63.18 | 62.98 | 0.20↓ |
| **Average** | 46.49 | 45.65 | 0.84↓ |

Table 5: The performance of our approach with and without the Multi-grain Graphical Model (on the entire test set).

Restaurant, and Hotel domains, the corresponding F1 scores consistently decrease when the document length increases. This illustrates the challenge of the Document-level TSA task, the longer the document length, the more difficult it is to accurately extract the multi-level opinion target and sentiment from it.

## 4.4 The Impact of Levels in Opinion Target

Table 4 reports the performance of our approach in extracting opinion targets with different levels. The same as Table 2, 1-T, 2-T, and 3-T represent the number of levels in an opinion target. We report the results on test subsets divided into 1-T, 2-T, and 3-T, respectively. We have not reported the results on News and PhraseBank as they contain relatively fewer multi-level opinion targets.

It can be seen that, when the number of levels increases, the F1 score decreases significantly. It indicates the challenge of the Document-level TSA task from another aspect. The more levels of entities the opinion target has, the more difficult the task will be.

In comparison with the BART-Extraction seq2seq method, our approach achieves consistent and stable improvements at different levels. The average improvements are 7.13%, 6.86%, and 3.88% on 1-T, 2-T, and 3-T, respectively.

| **Dataset** | **Within-Sentence** | | | **Across-Sentence** | | |
|---|---|---|---|---|---|---|
| | Seq2Str. | w/o GCN | Diff. | Seq2Str. | w/o GCN | Diff. |
| Books | 46.60 | 41.50 | 5.10↓ | 28.18 | 19.32 | 8.86↓ |
| Clothing | 61.12 | 59.92 | 1.20↓ | 54.70 | 46.51 | 8.19↓ |
| Restaurant | 41.76 | 40.64 | 1.12↓ | 35.90 | 32.50 | 3.40↓ |
| Hotel | 69.10 | 66.14 | 2.96↓ | 51.28 | 47.07 | 4.21↓ |
| **Average** | 54.64 | 52.05 | 2.59↓ | 42.51 | 36.35 | 6.16↓ |

Table 6: The performance of our approach with and without the Multi-grain Graphical Model (on opinion target with multi-level entities).

## 4.5 The Effect of the Multi-grain Graphical Model

In this part, we conduct ablation study on GCN to examine the effect of the multi-grain graphical model.

Firstly, in Table 5 we report the performance of our approach with and without the Multi-grain Graphical Model on the entire test set. It shows that removing the multi-grain graphical model from our approach causes an average of 0.84% decrease across six domains. The decrease is significant according to paired $t$-test.

Secondly, to analyze the effect of GCN on opinion target with multi-level entities, we divide multi-level opinion targets in the test set into a Within-Sentence subset and a Across-Sentence subset, where Within-Sentence denotes that multi-level entities are within a sentence, and Across-Sentence denotes that are across multiple sentences. The results in Table 6 shows that removing GCN causes a 2.59% and 6.16% drop in Within-Sentence and Across-Sentence respectively. It confirms the effectiveness of the advantage of multi-grain graphical model in capturing long-distance dependencies among affiliated entities, especially the across-sentence ones.

Finally, in Figure 6 we display the performance of our approach with and without GCN as the num-

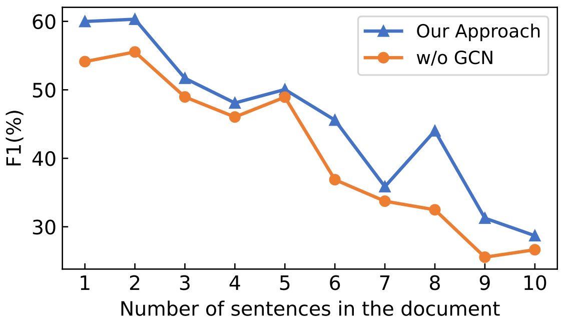


Figure 6: The average performance of six domains on different document lengths.

| **Dataset** | Seq2Struct | Seq2Struct$_{\texttt{variant}}$ | Diff. |
|---|---|---|---|
| Books | 38.41 | 37.08 | 1.33↓ |
| Clothing | 57.36 | 55.97 | 1.39↓ |
| Restaurant | 36.41 | 35.34 | 1.07↓ |
| Hotel | 60.10 | 59.58 | 0.52↓ |
| News | 23.47 | 22.85 | 0.62↓ |
| PhraseBank | 63.18 | 63.91 | 0.73↑ |
| **Average** | 46.49 | 45.79 | 0.70↓ |

Table 7: The performance comparison between Seq2Struct and Seq2Struct$_{\texttt{variant}}$ in document-level TSA task.

ber of sentences in the document increases. It can observed that the performance of our approach with GCN is insistently higher than that without GCN. As the number of sentences increases, the improvement becomes larger in general. Both suggest the effectiveness of the multi-grain graphical model of our approach in modeling long documents.

### 4.6 Discussion on the Place of Sentiment Classification

In our approach, we perform sentiment classification at the stage of flat entity extraction. A corresponding question is then raised: *Which place is the most suitable for sentiment classification?*

To answer this question, we design a variant of our approach Seq2Struct$_{\texttt{variant}}$, which predicts the sentiment after obtaining the hierarchical opinion target structure, and report its performance in Table 7. It can be seen that the F1 score of Seq2Struct$_{\texttt{variant}}$ has an average decrease of 0.7%. We speculate that the possible reason is that the opinion expression often appears near the entity, and has little relation with the structure of opinion targets. It may hence be more effective to perform sentiment classification towards the flat entities.

### 4.7 Case Study

In Figure 7, we conduct the case study by displaying the outputs of our approach (Seq2Struct) and Seq2Seq. In comparison, Seq2Struct can more explicitly display the hierarchical structure of opinion target in a document and more accurately predict the corresponding sentiments, across different document length. For example, in short document 1, Seq2Struct can predict the upper entity "*Danskin*" of "*quality*". In document 2, which is slightly longer, Seq2Struct can predict what Seq2Seq cannot predict ("*read-personalities*", Positive). In the longer document 3, Seq2Seq predicts the wrong pair ("*tights-size B*", Negative), while Seq2Struct predicts them all correctly.

Furthermore, Seq2Struct can accurately predict distant hierarchical entities. For example, in document 5, Seq2Struct predicts the pair ("*Alexander Cipher-character*", Positive), where "*Alexander Cipher*" and "*character*" are separated by four long sentences.

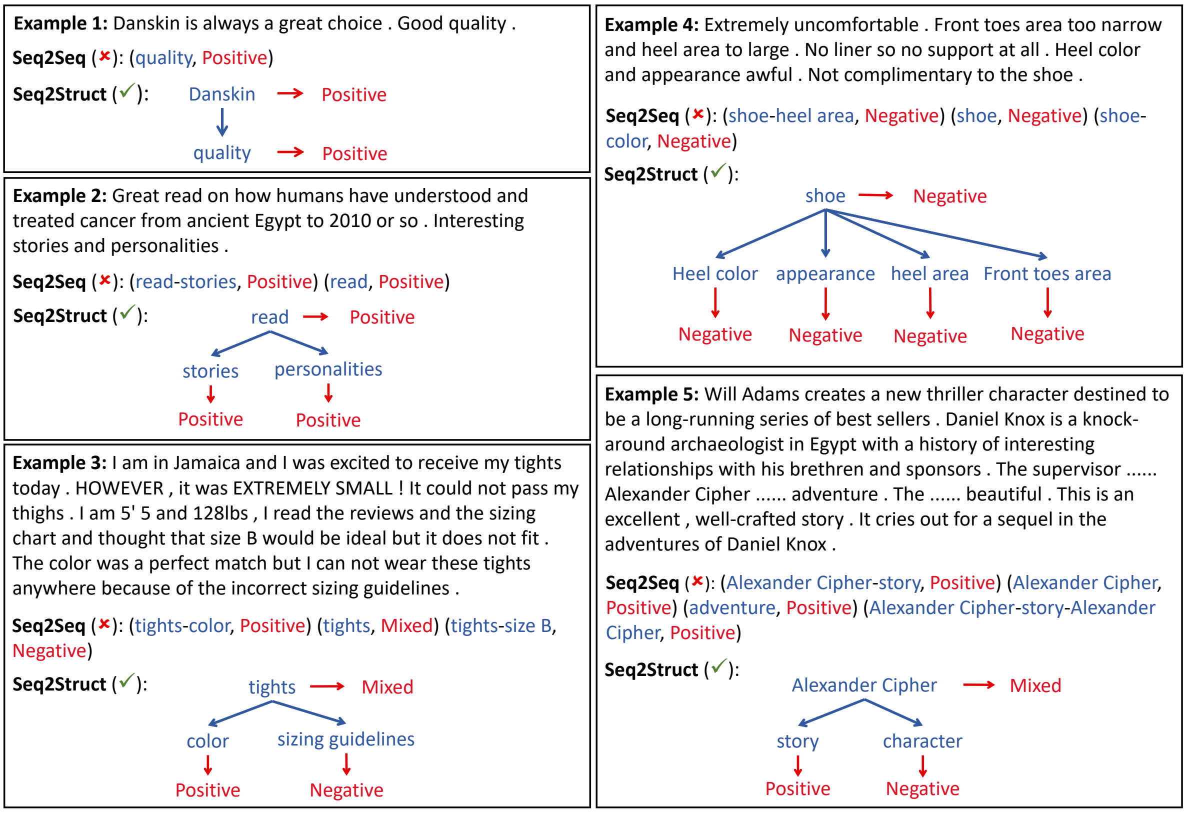
In addition, Seq2Struct can recall more entities. For example, in document 4, the predicted entities of Seq2Struct are "*Heel color*", "*Front toes area*", and "*appearance*".

## 5 Related Work

ABSA is a broad research area which includes various tasks. Schouten and Frasincar (2016); Zhang et al. (2022a) provided comprehensive survey to these subtasks. In this paper, due to space limitation, we only review the related tasks.

End-to-end ABSA, the task of joint aspect extraction and aspect-based sentiment classification (also called targeted sentiment analysis in some references), has received wide attention (Mitchell et al., 2013; Zhang et al., 2015; Poria et al., 2016; Hu et al., 2019; Li et al., 2019a,b; Jiang et al., 2019; Chen and Qian, 2020; Yu et al., 2021b; Hamborg and Donnay, 2021). However, all these studies were performed at the sentence level. In this paper, we focused on targeted sentiment analysis at the document level. Unlike the opinion target at the sentence level, which is normally an entity or aspect, the opinion target studied in this work often contains multi-level entities.

Among massive ABSA studies, only a few focused on the document level. Titov and McDonald (2008) proposed a statistical model to extract textual evidence for aspect category and predict sentiment rating for different categories in a review document. Lei et al. (2016) proposed an encoder-generator framework to extract rationales for aspect category and predict aspect category sentiment rating . Yin et al. (2017) modelled aspect category sentiment rating as a machine comprehension problem. Li et al. (2018) designed a hierarchical network for aspect category sentiment considering both user preference and overall rating. Wang et al. (2019) proposed a hierarchical reinforcement learning approach to interpretably predict aspect category sentiment rating. However, all these studies focused on identifying the sentiments of aspect categories in a document. By contrast, we extract the explicit opinion entity terms in a document and organize them in a hierarchical structure.



**Example 1:** Danskin is always a great choice . Good quality .

**Seq2Seq** (✗): (quality, Positive)

**Seq2Struct** (✓):

**Example 2:** Great read on how humans have understood and treated cancer from ancient Egypt to 2010 or so . Interesting stories and personalities .

**Seq2Seq** (✗): (read-stories, Positive) (read, Positive)

**Seq2Struct** (✓):

**Example 3:** I am in Jamaica and I was excited to receive my tights today . HOWEVER , it was EXTREMELY SMALL ! It could not pass my thighs . I am 5' 5 and 128lbs , I read the reviews and the sizing chart and thought that size B would be ideal but it does not fit . The color was a perfect match but I can not wear these tights anywhere because of the incorrect sizing guidelines .

**Seq2Seq** (✗): (tights-color, Positive) (tights, Mixed) (tights-size B, Negative)

**Seq2Struct** (✓):

**Example 4:** Extremely uncomfortable . Front toes area too narrow and heel area to large . No liner so no support at all . Heel color and appearance awful . Not complimentary to the shoe .

**Seq2Seq** (✗): (shoe-heel area, Negative) (shoe, Negative) (shoe-color, Negative)

**Seq2Struct** (✓):

**Example 5:** Will Adams creates a new thriller character destined to be a long-running series of best sellers . Daniel Knox is a knock-around archaeologist in Egypt with a history of interesting relationships with his brethren and sponsors . The supervisor ...... Alexander Cipher ...... adventure . The ...... beautiful . This is an excellent , well-crafted story . It cries out for a sequel in the adventures of Daniel Knox .

**Seq2Seq** (✗): (Alexander Cipher-story, Positive) (Alexander Cipher, Positive) (adventure, Positive) (Alexander Cipher-story-Alexander Cipher, Positive)

**Seq2Struct** (✓):

Figure 7: The case study by comparing the Seq2Seq approach and our Seq2Struct approach.

From the perspective of methodology, graph Convolutional network (GCN) has been widely used in ABSA (Zhang et al., 2019; Sun et al., 2019; Cai et al., 2020; Liang et al., 2021; Hou et al., 2021; Tian et al., 2021; Zhang et al., 2022b; Chen et al., 2022). However, most of these studies employed GCN to model the relations between different entities within a single sentence. Different from that, in this work a multi-grain graphical model is proposed to learn the affiliated relations among entities across multiple sentences in a document.

Table filling, the method to predict the relation between any two targets by filling a table, has received much attention on entity and relation extraction task (Miwa and Sasaki, 2014; Gupta et al., 2016; Wang and Lu, 2020) and open information extraction task (Yu et al., 2021a). In the ABSA task, Wu et al. (2020); Jing et al. (2021) proposed to use table filling to tag aspect terms, opinion terms and the relations between them. In contrast, in this work we use table filling to model the affiliated relation between two entities.

## 6 Conclusion

In this work, we focus on a document-level ABSA task, called document-level TSA, which aims to extract the opinion targets consisting of multi-level entities from a review document and predict the corresponding sentiments. Different from the existing Seq2Seq mythology, we propose a Sequence-to-Structure (Seq2Struct) approach to address this task, to model the hierarchical structure among multiple opinion targets and capture the long-distance dependencies among affiliated entities. Experiments have verified the advantages of our Seq2Struct approach in more accurately extracting multi-level opinion targets and predicting their sentiments, and more explicitly displaying the hierarchical structure of the opinion targets in a document.

## Limitations

This paper focuses on addressing the task of document-level TSA, which, along with its dataset, has been recently introduced. Our approach is primarily designed to tackle the challenge of extracting the affiliated relations among entities over an in-domain setting. Nevertheless, this task remains challenges, particularly in aspects such as long document encoding, coreference problem, and open-domain setting. We welcome more researchers to explore this task.

## Ethics Statement

We conduct experiments on the publicly available document-level TSA dataset, which includes Books, Clothing, Restaurant, Hotel, News and PhraseBank. The dataset do not share personal information and do not contain sensitive content that can be harmful to any individual or community.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their insightful comments. This work was supported by the Natural Science Foundation of China (62076133 and 62006117), and the Natural Science Foundation of Jiangsu Province for Young Scholars (BK20200463) and Distinguished Young Scholars (BK20200018).

## References

Jiahao Bu, Lei Ren, Shuang Zheng, Yang Yang, Jingang Wang, Fuzheng Zhang, and Wei Wu. 2021. ASAP: A chinese review dataset towards aspect category sentiment analysis and rating prediction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2069–2079. Association for Computational Linguistics.

Hongjie Cai, Yaofeng Tu, Xiangsheng Zhou, Jianfei Yu, and Rui Xia. 2020. Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 833–843. International Committee on Computational Linguistics.

Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 2974–2985. Association for Computational Linguistics.

Zhuang Chen and Tieyun Qian. 2020. Relation-aware collaborative learning for unified aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3685–3694. Association for Computational Linguistics.

Pankaj Gupta, Hinrich Schütze, and Bernt Andrassy. 2016. Table filling multi-task recurrent neural network for joint entity and relation extraction. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2537–2547. ACL.

Felix Hamborg and Karsten Donnay. 2021. Newsmtsc: A dataset for (multi-)target-dependent sentiment classification in political news articles. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 1663–1675. Association for Computational Linguistics.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: decoding-enhanced bert with disentangled attention. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Xiaochen Hou, Peng Qi, Guangtao Wang, Rex Ying, Jing Huang, Xiaodong He, and Bowen Zhou. 2021. Graph ensemble learning over multiple dependency trees for aspect-level sentiment classification. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2884–2894. Association for Computational Linguistics.

Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 537–546. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, pages 168–177. ACM.

Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 6279–6284. Association for Computational Linguistics.

Hongjiang Jing, Zuchao Li, Hai Zhao, and Shu Jiang. 2021. Seeking common but distinguishing difference, A joint aspect-based sentiment analysis model. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 3910–3922. Association for Computational Linguistics.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Tao Lei, Regina Barzilay, and Tommi S. Jaakkola. 2016. Rationalizing neural predictions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 107–117. The Association for Computational Linguistics.

Junjie Li, Haitong Yang, and Chengqing Zong. 2018. Document-level multi-aspect sentiment classification by jointly modeling users, aspects, and overall ratings. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 925–936. Association for Computational Linguistics.

Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019a. A unified model for opinion target extraction and target sentiment prediction. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6714–6721. AAAI Press.

Zheng Li, Xin Li, Ying Wei, Lidong Bing, Yu Zhang, and Qiang Yang. 2019b. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4589–4599. Association for Computational Linguistics.

Bin Liang, Hang Su, Rongdi Yin, Lin Gui, Min Yang, Qin Zhao, Xiaoqi Yu, and Ruifeng Xu. 2021. Beta distribution guided aspect-aware graph for aspect category sentiment analysis with affective knowledge. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 208–218. Association for Computational Linguistics.

Yun Luo, Hongjie Cai, Linyi Yang, Yanxia Qin, Rui Xia, and Yue Zhang. 2022. Challenges for open-domain targeted sentiment analysis. *CoRR*, abs/2204.06893.

Margaret Mitchell, Jacqui Aguilar, Theresa Wilson, and Benjamin Van Durme. 2013. Open domain targeted sentiment. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1643–1654. ACL.

Makoto Miwa and Yutaka Sasaki. 2014. Modeling joint entity and relation extraction with table representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1858–1869. ACL.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia V. Loukachevitch, Evgeniy V. Kotelnikov, Núria Bel, Salud María Jiménez Zafra, and Gülsen Eryigit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, pages 19–30. The Association for Computer Linguistics.

Soujanya Poria, Iti Chaturvedi, Erik Cambria, and Federica Bisio. 2016. Sentic LDA: improving on LDA with semantic similarity for aspect-based sentiment analysis. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 4465–4473. IEEE.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21:140:1–140:67.

Kim Schouten and Flavius Frasincar. 2016. Survey on aspect-level sentiment analysis. *IEEE Trans. Knowl. Data Eng.*, 28(3):813–830.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5678–5687. Association for Computational Linguistics.

Yuanhe Tian, Guimin Chen, and Yan Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2910–2922. Association for Computational Linguistics.

Ivan Titov and Ryan T. McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *ACL 2008, Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, June 15-20, 2008, Columbus, Ohio, USA*, pages 308–316. The Association for Computer Linguistics.

Jingjing Wang, Changlong Sun, Shoushan Li, Jiancheng Wang, Luo Si, Min Zhang, Xiaozhong Liu, and Guodong Zhou. 2019. Human-like decision making: Document-level aspect sentiment classification via hierarchical reinforcement learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5580–5589. Association for Computational Linguistics.

Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 1706–1721. Association for Computational Linguistics.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585.

Yichun Yin, Yangqiu Song, and Ming Zhang. 2017. Document-level multi-aspect sentiment classification as machine comprehension. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 2044–2054. Association for Computational Linguistics.

Bowen Yu, Yucheng Wang, Tingwen Liu, Hongsong Zhu, Limin Sun, and Bin Wang. 2021a. Maximal clique based non-autoregressive open information extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 9696–9706. Association for Computational Linguistics.

Guoxin Yu, Xiang Ao, Ling Luo, Min Yang, Xiaofei Sun, Jiwei Li, and Qing He. 2021b. Making flexible use of subtasks: A multiplex interaction network for unified aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 2695–2705. Association for Computational Linguistics.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4567–4577. Association for Computational Linguistics.

Meishan Zhang, Yue Zhang, and Duy-Tin Vo. 2015. Neural networks for open domain targeted sentiment. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 612–621. The Association for Computational Linguistics.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021a. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 9209–9219. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021b. Towards generative aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 2: Short Papers), Virtual Event, August 1-6, 2021*, pages 504–510. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022a. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *CoRR*, abs/2203.01054.

Zheng Zhang, Zili Zhou, and Yanna Wang. 2022b. SSEGCN: syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 4916–4925. Association for Computational Linguistics.