

Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions

Hongjie Cai* Rui Xia*† Jianfei Yu

School of Computer Science and Engineering,
Nanjing University of Science and Technology, China
{hjcai, rxia, jfyu}@njjust.edu.cn

Abstract

Product reviews contain a large number of implicit aspects and opinions. However, most of the existing studies in aspect-based sentiment analysis ignored this problem. In this work, we introduce a new task, named Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction, with the goal to extract all aspect-category-opinion-sentiment quadruples in a review sentence and provide full support for aspect-based sentiment analysis with implicit aspects and opinions. We further construct two new datasets Restaurant-ACOS and Laptop-ACOS for this new task. The former is an extension of the SemEval Restaurant dataset; the latter is a brand new Laptop dataset with much larger size than the SemEval Laptop dataset. Both contain the annotations of not only aspect-category-opinion-sentiment quadruples but also implicit aspects and opinions. We finally benchmark the task with four baseline systems. Experiments demonstrate the feasibility of the new task and its advantage in extracting and describing implicit aspects and implicit opinions in ABSA. The two datasets and source code of four systems are publicly released at <https://github.com/NUSTM/ACOS>.

1 Introduction

As a fine-grained sentiment analysis task, aspect-based sentiment analysis (ABSA) has received continuous attention. Its core task is to extract the opinion target described by an entity and its aspect (collectively referred to as aspect) from product reviews, and identify the sentiment toward the aspect (Liu, 2012). The standard aspect-based sentiment analysis task includes two basic subtasks: aspect extraction and aspect-based sentiment classification. By integrating the two subtasks, one can

*Equal contribution.

†Corresponding author.

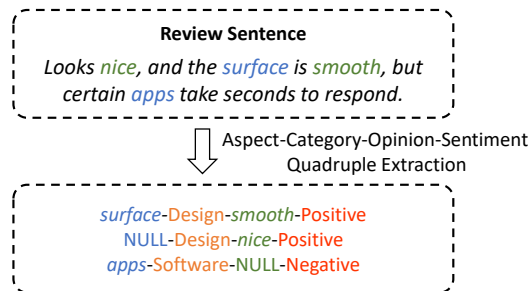


Figure 1: An example of the Aspect-Category-Opinion-Sentiment Quadruple Extraction task.

	Restaurant	Laptop
Explicit Aspect & Explicit Opinion	63.34%	56.06%
Implicit Aspect & Explicit Opinion	19.47%	17.54%
Explicit Aspect & Implicit Opinion	12.38%	27.55%
Implicit Aspect & Implicit Opinion	14.83%	8.24%

Table 1: The percentage of review sentences with explicit and implicit aspect/opinion.

identify an aspect-sentiment pair (g, s) , where g is an aspect term, and s is the sentiment polarity toward the aspect. (Hu and Liu, 2004; Qiu et al., 2011) pointed out that the correlation between the aspect term and the opinion term is helpful for better ABSA. The following studies in this direction includes aspect-opinion co-extraction (Wang et al., 2016a, 2017; Yu et al., 2018; Li et al., 2018; Dai and Song, 2019), aspect-opinion pair extraction (Chen et al., 2020a; Zhao et al., 2020), and aspect-opinion-sentiment triple extraction (Peng et al., 2020; Xu et al., 2020; Wu et al., 2020; Mao et al., 2021), etc.

However, most of the existing studies only considered the extraction of explicit aspects and opinions, while ignored the implicit ones. In fact, product reviews contain a large amount of implicit aspects and opinions. Table 1 summarizes the percentage of implicit aspects and opinions in the

SemEval Restaurant dataset and our new Laptop dataset. It can be seen that nearly 44% of the review sentences contain implicit aspects or implicit opinions in the Laptop domain, and the percentage of sentences containing both implicit aspects and implicit opinions also exceeds 8%. Similar percentages can be observed in the Restaurant domain. Although some studies have attempted to solve the implicit aspect problem (Liu et al., 2005; Poria et al., 2014; Chen and Chen, 2016; Wan et al., 2020) or the implicit opinion problem (Lazhar and Guiyassa, 2016) from respective perspectives, there is still a lack of a unified framework that fully discusses and solves the implicit aspect and implicit opinion problems.

In this work, we introduce a new task named **Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction**, with the goal to extract all aspect-category-opinion-sentiment quadruples in a review sentence, and provide full support for aspect-level sentiment analysis with implicit aspects and opinions. As shown in Figure 1, in the review sentence “*Looks nice and the surface is smooth, but certain apps take seconds to respond*”, *surface* is an aspect, *Design* is its category, *smooth* is the opinion toward this aspect, and *Positive* is the corresponding sentiment. The four elements are combined into an explicit quadruple *surface-Design-smooth-Positive*. In addition to that, there are two other quadruples that need to be extracted: *Null-Design-nice-Positive* which contains an implicit aspect, and *apps-Software-Null-Negative* which contains an implicit opinion.

The new ACOS Quadruple Extraction task has the following two challenges:

- In term of dataset, so far there was no available dataset that is fully annotated with aspect-category-opinion-sentiment quadruples including all implicit aspects and opinions;
- In terms of modeling complexity, the task includes two extraction problems (aspect extraction, opinion extraction) and two classification problems (category classification, sentiment classification). It is challenging to effectively model the four subtasks together to construct quadruples containing implicit aspects and implicit opinions.

To address these issues, we further construct two new datasets, **Restaurant-ACOS** and **Laptop-ACOS**, for the new task. The former is an extension of the existing SemEval Restaurant dataset,

based on which we add the annotation of implicit aspects, implicit opinions, and the quadruples. The latter is a brand new one collected from the Amazon Laptop domain. It has twice size of the SemEval Laptop dataset, and is annotated with quadruples containing all explicit/implicit aspects and opinions.

We finally benchmark the task by establishing four baseline systems, **Double-Propagation-ACOS**, **JET-ACOS**, **TAS-BERT-ACOS** and **Extract-Classify-ACOS**, by adapting the representative approaches in aspect-opinion pair extraction, aspect-category-opinion triple extraction or aspect-opinion-sentiment triple extraction to ACOS Quadruple Extraction. The experiments on the two ACOS datasets demonstrate the feasibility of the new ACOS Quadruple Extraction task and its effectiveness in extracting and describing implicit aspects and implicit opinions.

The contributions of this work can be summarized as follows:

- We introduce a new task named Aspect-Category-Opinion-Sentiment Quadruple Extraction, to address the implicit aspects/opinions issues in ABSA;
- We construct two new datasets for the task, with ACOS quadruple annotations including implicit aspects/opinions;
- We benchmark the task with four baseline systems. The experiments demonstrate the new task’s advantage in addressing the implicit aspect/opinion issues.

2 Task

We first define the four elements of the ACOS Quadruple Extraction task based on (Liu, 2012). (Peng et al., 2020; Mao et al., 2021) provided good summaries of recent tasks and terminology in ABSA. For simplicity, in this paper we use *aspect*, *category*, *opinion* and *sentiment* to denote *aspect term*, *aspect category*, *opinion term* and *sentiment polarity*, respectively. They are defined as follows:

- **Aspect** denotes an entity and its aspect indicating the opinion target, which is normally a word or phrase in the text;
- **Category** represents a unique predefined category for the aspect in a particular domain;
- **Opinion** refers the subjective statement on an aspect, which is normally a subjective word or phrase in the text;

- **Sentiment** is the predefined semantic orientation (e.g., Positive, Negative, or Neutral) toward the aspect.

Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction is then defined as a task to extract a set of aspect-category-opinion-sentiment quadruples described in a review sentence containing n words $r=[w_1, \dots, w_n]$:

$$\mathcal{S}_{ACOS} = \{\dots, a_i-c_j-o_k-s_l, \dots\}, \quad (1)$$

where $a_i-c_j-o_k-s_l$ denotes an aspect-category-opinion-sentiment quadruple, a_i is the extracted aspect, $c_j \in \mathcal{C}$ is its category, o_k is the extracted opinion, and $s_l \in \{\text{Positive, Neutral, Negative}\}$ is its corresponding sentiment.¹

Note that a review sentence usually contains multiple aspects and opinions. The ACOS Quadruple Extraction task does not only identify four elements, but also combine them into a set of valid quadruples, meanwhile considering implicit aspects/opinions. As the implicit aspect/opinion is not explicitly expressed as a word or phrase, in case of implicit aspect we set a as NULL and use category c to describe the opinion target, and in case of implicit opinion we set o as NULL and use sentiment s to describe the semantic orientation.

3 Datasets

We construct two new datasets, Restaurant-ACOS and Laptop-ACOS, for the ACOS Quadruple Extraction task.

3.1 Source

The Restaurant-ACOS dataset is constructed based on the SemEval 2016 Restaurant dataset (Pontiki et al., 2016) and its expansion datasets (Fan et al., 2019; Xu et al., 2020).

Laptop-ACOS is a brand new Laptop dataset collected from the Amazon platform at the years of 2017 and 2018 (covering ten types of laptops under six brands such as ASUS, acer, Samsung, Lenovo, MBP, MSI and so on). It contains 4,076 review sentences, much larger than the SemEval Laptop datasets.

¹Similarly, the previous representative tasks in ABSA can also be denoted by the combination of the above elements, e.g., aspect-sentiment (AS) pair extraction (Mitchell et al., 2013; Zhang et al., 2015), aspect-opinion (AO) pair extraction (Chen et al., 2020a; Zhao et al., 2020), aspect-opinion-sentiment (AOS) triple extraction (Peng et al., 2020; Xu et al., 2020; Wu et al., 2020; Mao et al., 2021; Chen et al., 2021), aspect-category-sentiment (ACS) triple extraction (Wan et al., 2020), etc.

		Restaurant-ACOS	Laptop-ACOS
#Categories		13	121
#Sentences		2286	4076
#Quadruples	EA & EO	2429 (66.40%)	3269 (56.77%)
	IA & EO	530 (14.49%)	910 (15.80%)
	EA & IO	350 (9.57%)	1237 (21.48%)
	IA & IO	349 (9.54%)	342 (5.94%)
	All	3658	5758
#Quadruples #Sentences		1.60	1.42

Table 2: Statistics of our two ACOS Quadruple datasets. EA, EO, IA and IO denote explicit aspect, explicit opinion, implicit aspect, and implicit opinion, respectively. #Categories represents the number of aspect categories which are consistent with that in (Pontiki et al., 2016).

3.2 Annotation

The SemEval 2016 Restaurant dataset (Pontiki et al., 2016) was annotated with explicit and implicit aspects, categories, and sentiment. (Fan et al., 2019; Xu et al., 2020) further added the opinion annotations. We integrate their annotations to construct aspect-category-opinion-sentiment quadruples and further annotate the implicit opinions.

For Laptop-ACOS, we annotate the four elements and their corresponding quadruples all by ourselves. We employ the aspect categories defined in the SemEval 2016 Laptop dataset. Two PhD students familiar with aspect-based sentiment analysis are selected as annotators for independent annotation with the annotation tool introduced by (Yang et al., 2017a). The strict quadruple matching F_1 score between two annotators is 75.86%, which indicates a substantial agreement between two annotators (Kim and Klinger, 2018). In case of disagreement, a third expert will be asked to make the final decision.

3.3 Statistics and Analysis

The basic statistics of the two datasets are reported in Table 2. The Restaurant-ACOS dataset contains 2286 sentences with 3658 quadruples, and the Laptop-ACOS dataset contains 4076 sentences with 5758 quadruples. As we have mentioned, a large percentage of the quadruples contain implicit aspects or implicit opinions. By comparing two datasets, it can be observed that Laptop-ACOS has higher percentage of implicit opinions than Restaurant-ACOS.

In Table 3, we further compare our two ACOS datasets with the existing representative datasets

	Sentence	Aspect	Category	Opinion	Sentiment	AS Pair	AO Pair	AOS Triple	ACS Triple	ACOS Quadruple
Restaurant-2014 (Pontiki et al., 2014)	3841	4827	4738	-	4534	4827	-	-	-	-
Laptop-2014 (Pontiki et al., 2014)	1910	3012	-	-	3012	3012	-	-	-	-
Restaurant-2016 (Pontiki et al., 2016)	2295	3122	3001	-	3122	3182	-	-	3364	-
Laptop-2016 (Pontiki et al., 2016)	2612	-	3705	-	3705	-	-	-	-	-
Restaurant-2014-AO (Fan et al., 2019)	2125	3503	-	3610	-	-	4092	-	-	-
Restaurant-2016-AO (Fan et al., 2019)	1407	1968	-	2146	-	-	2294	-	-	-
Restaurant-2014-AOS (Xu et al., 2020)	2068	3399	-	3443	3399	3399	3908	3908	-	-
Restaurant-2016-AOS (Xu et al., 2020)	1393	1946	-	2101	1946	1946	2247	2247	-	-
Restaurant-ACOS (ours)	2286	3110	2967	3335	3110	3155	3571	3575	3335	3658
Laptop-ACOS (ours)	4076	4958	4992	5378	4958	5035	5726	5731	5227	5758

Table 3: The comparison between the sizes of our two ACOS Quadruple datasets and existing representative ABSA datasets. AS, AO, AOS, and ACS denote Aspect-Sentiment, Aspect-Opinion, Aspect-Opinion-Sentiment, and Aspect-Category-Sentiment, respectively.

in ABSA. Restaurant 2014/2016 and Laptop 2014/2016 denote the SemEval 2014/2016 Restaurant and Laptop datasets, respectively. Restaurant 2014/2016 contains the annotations of aspect, category and sentiment. It should be noted the category definitions in two datasets are different. Laptop 2014 contains only the annotations of aspect and sentiment, while Laptop 2016 contains only the annotations of category and sentiment.

Restaurant-2014-AO and Restaurant-2016-AO are two aspect-opinion pair datasets proposed by (Fan et al., 2019), based on Restaurant 2014 and 2016, respectively. They removed the sentences with implicit aspects and added the opinion annotations. (Xu et al., 2020) further added sentiment which was originally included in Restaurant 2014/2016 to Restaurant-2014/2016-AO, and obtained two aspect-opinion-sentiment triple datasets: Restaurant-2014-AOS and Restaurant-2016-AOS.

For Restaurant-ACOS, we integrate the above annotations to construct ACOS quadruples. But it should be noted that we keep the sentences with implicit aspects in Restaurant-2016, and further annotate the implicit opinions. As a result, the size (including sentences, AO pairs and AOS triples) of Restaurant-ACOS is about 1.6 times that of Restaurant-2016-AO and Restaurant-2016-AOS.

The new Laptop-ACOS has 4076 review sentences. The numbers of annotations for aspect, category, opinion and sentiment are 4958, 4992, 5378 and 4958, respectively. By combining these elements, we construct 5035 AS pairs, 5726 AO pairs, 5731 AOS triples, 5227 ACS triples and 5758 ACOS quadruples, nearly twice the size of Restaurant-ACOS.²

²It is worth noting that the Restaurant-ACOS and Laptop-

4 Methods

We benchmark the ACOS Quadruple Extraction task with four baseline systems, namely, Double-Propagation-ACOS, JET-ACOS, TSBERT-ACOS and Extract-Classify-ACOS, by adapting the representative approaches in aspect-opinion pair extraction, aspect-category-opinion triple extraction or aspect-opinion-sentiment triple extraction to ACOS Quadruple Extraction.

4.1 Double-Propagation-ACOS

Since Double Propagation (DP) is one of the representative rule-based methods for aspect-opinion-sentiment triple extraction (Qiu et al., 2011), we propose to adapt it to our ACOS quadruple extraction task by first extracting all the aspect-opinion-sentiment triples, followed by assigning the aspect category for each extracted triple. We name the adapted approach as Double-Propagation-ACOS.

Specifically, we first follow the DP algorithm to extract the aspect-opinion-sentiment triples, where we utilize the syntactic relations between aspects and opinions to iteratively extract them in each review, and rely on the sentiment lexicon to assign sentiments (i.e., Positive, Negative, and Neutral) to aspects and opinions in a bootstrapping manner. Second, to identify the aspect category of each extracted triple, we use the following strategy: if the aspect in the triple is in the training set, we take its most co-occurred aspect category as the final aspect category; otherwise, we adopt the aspect

ACOS datasets are available for all subtasks in ABSA, including aspect-based sentiment classification, aspect-sentiment pair extraction, aspect-opinion pair extraction, aspect-opinion-sentiment triple extraction, aspect-category-sentiment triple extraction, etc.

category of the nearest aspect in the input review as the final aspect category.

Based on the two steps mentioned above, we can extract the ACOS quadruples in each review sentence.

4.2 JET-ACOS

As one of the state-of-the-art approaches for aspect-opinion-sentiment triple extraction, JET (Xu et al., 2020) introduced an end-to-end framework to this task, by combining the identification of aspects, their corresponding opinions, and their sentiment polarities with a position-aware tagging scheme³. Similar to Double-Propagation-ACOS, we adapt JET to our task by first extracting the triple with JET, followed by predicting the aspect category for each extracted triple.

Specifically, we first obtain the candidate aspect-opinion-sentiment triples based on JET, and then design a BERT-based model to get the aspect category of the extracted triples. Given the review sentence r , we first feed it to BERT to get the context-aware token representation H as follows:

$$H = [h_{[\text{CLS}]}, h_r, h_{[\text{SEP}]}, \dots], \quad (2)$$

where $h_r = [h_1, \dots, h_n]$ is the output representation for r . Next, given an extracted triple $a-o-s$, we can obtain the representation of the aspect and the opinion as $u_a = \text{avg}(h_a)$ and $u_o = \text{avg}(h_o)$, where $\text{avg}(h_a)$ and $\text{avg}(h_o)$ are the average vectors of words in the aspect h_a and the opinion h_o , respectively. We then concatenate u_a and u_o , and feed it to a fully-connected layer with the Sigmoid function for each category c :

$$y^c = \text{Sigmoid}(W_c^\top [u_a; u_o] + b_c). \quad (3)$$

Given $a-o-s$ and c , $y^c = 1$ indicates a valid quadruple, and $y^c = 0$ indicates an invalid quadruple.

In the training stage, we adopt the standard binary cross-entropy loss for optimization. In the inference stage, we combine the extracted aspect-opinion-sentiment triples from JET and our predicted aspect categories to get all the quadruples from each review sentence.

³JET contains two variants, i.e., JET^t and JET^o. JET^t aims to identify the aspects, the offset of their corresponding opinions, and their sentiment polarity; whereas JET^o aims to identify the opinions, the offset of their corresponding aspects, and their sentiment polarity. We employ JET^o to extract the aspect-opinion-sentiment triple, as it has been shown to obtain better performance than JET^t.

4.3 TAS-BERT-ACOS

TAS-BERT (Wan et al., 2020) is one of the state-of-the-art method for aspect-category-sentiment triple extraction, which integrates aspect category-based sentiment classification and aspect extraction in a unified framework by attaching the aspect category and the sentiment polarity to the review sentence and using it as the input of BERT. To adapt TAS-BERT to our ACOS extraction task, we propose to adopt the input transformation strategy in TAS-BERT to perform category-sentiment conditional aspect-opinion co-extraction, following by filtering out the invalid aspect-opinion pairs to form the final quadruples.

Specifically, given a review sentence r , an aspect category $c \in \mathcal{C}$, and a sentiment $s \in \mathcal{S}$, the input is constructed as follows:

$$x = [[\text{CLS}], r, [\text{SEP}], c, s, [\text{SEP}]], \quad (4)$$

We then feed x to BERT to get the context-aware token representation H :

$$H = [h_{[\text{CLS}]}, h_r, h_{[\text{SEP}]}, h_{cs}, h_{[\text{SEP}]}], \quad (5)$$

where $h_r = [h_1, \dots, h_n]$ is the output representation for r , h_{cs} is the output representation for the concatenation of c and s , and $h_{[\text{CLS}]}$ is used for category-sentiment verification.

We then perform aspect-opinion co-extraction over H by modeling it as a single sequence labeling task. Specifically, we employ a modified Begin-Inside-Outside (BIO) tagging scheme, which consists of five tags: {BA, IA, BO, IO, O}, indicating the beginning and inside of the aspect, the beginning and inside of the opinion, and others. We feed h_r to a CRF layer to extract the aspects and opinions in r with respect to the input category c and sentiment s as follows:

$$Y^{ao} = [y_1^{ao}, \dots, y_n^{ao}] = \text{CRF}(h_1, \dots, h_n); \quad (6)$$

Next, we perform Cartesian Product on the extracted aspects and opinions to obtain a set of candidate aspect-category-opinion-sentiment quadruples:

$$\mathcal{S}_{ACOS} = \{a_1-c_1-o_1-s_1, \dots, a_{|A|}-c_{|C|}-o_{|O|}-s_{|S|}\}, \quad (7)$$

where $|A|$ and $|O|$ are the number of extracted aspects and opinions, $|C|$ and $|S|$ are the number of detected categories and sentiment.

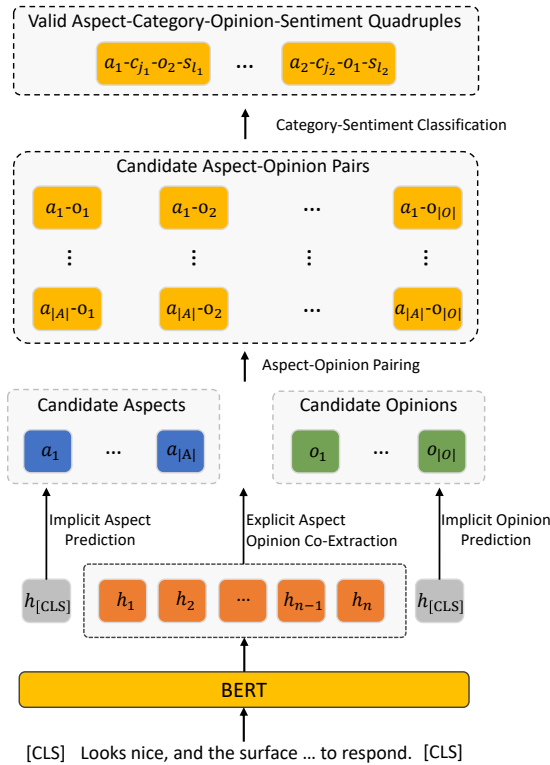


Figure 2: The Structure of Extract-Classify-ACOS.

On the basis of \mathcal{S}_{ACOS} , we average the vectors of tokens in the aspect and opinion, and then feed their concatenation $[\mathbf{u}_a; \mathbf{u}_o]$ to a quadruple filter:

$$y^{acos} = \text{Sigmoid}(\mathbf{W}^\top [\mathbf{u}_a; \mathbf{u}_o] + b), \quad (8)$$

where $y^{acos} = 1$ indicates a valid quadruple, and $y^{acos} = 0$ indicates an invalid quadruple.

4.4 Extract-Classify-ACOS

Finally, we propose Extract-Classify-ACOS by adapting one of the representative aspect-opinion co-extraction system (Wang et al., 2017) to our ACOS quadruple extraction task. Specifically, the first step performs aspect-opinion co-extraction, and the second step predicts category-sentiment given the extracted aspect-opinion pairs.

As shown in Figure 2, we first insert two [CLS] tokens at the beginning and the end of the review sentence r , and then feed the transformed input to BERT to obtain the context-aware token representations \mathbf{H} as follows:

$$\mathbf{H} = [h_{[\text{CLS}]}, h_r, h_{[\text{CLS}]}], \quad (9)$$

Similar to the method in TAS-BERT-ACOS, the explicit aspect-opinion co-extraction is based on a CRF layer with the modified BIO tagging scheme.

	Training	Validation	Testing
Restaurant-ACOS	1531	170	585
Laptop-ACOS	2934	326	816

Table 4: The division of training, validation, and testing sets.

We further apply two binary classification tasks on the [CLS] tokens to predict whether there is implicit aspect or implicit opinion. Thus, we can obtain the potential aspect set \mathcal{S}_A , opinion set \mathcal{S}_O , and perform Cartesian Product on \mathcal{S}_A and \mathcal{S}_O to obtain a set of candidate aspect-opinion pairs:

$$\mathcal{S}_{AO} = \{a_1-o_1, \dots, a_{|A|}-o_{|O|}\}. \quad (10)$$

Next, we model the category-sentiment classification as a multiple multi-class classification problem. Specifically, for each category c , we concatenate the average vectors of each aspect-opinion pair $a-o$, and feed them to a fully-connected layer with Softmax function as follows:

$$s^{aoc} = \text{Softmax}(\mathbf{W}_{aoc}^\top [\mathbf{u}_a; \mathbf{u}_o] + b_{aoc}), \quad (11)$$

where $s^{aoc} \in \{\text{Positive, Negative, Neutral, Invalid}\}$ denotes its sentiment given current $a-o$ and c , or indicates an invalid quadruple.

5 Experiments

We evaluate the performance of four baselines systems on two ACOS quadruple datasets.

5.1 Experimental Settings and Evaluation Metrics

In Extract-Classify-ACOS, we adopt BERT_{base} (Devlin et al., 2018) as the basic encoder, which consists of 12 stacked Transformer blocks. During training, we use the AdamW optimizer of BERT with weight decay fix. The maximum length of the review sentence is set to 128, covering all sentences in two datasets. We set the batch size and learning rates in aspect opinion co-extraction and category-sentiment classification as [32, 2e-5] and [16, 3e-5], respectively. The dropout rate is set as 0.1. The batch size and learning rate in the category classification of JET-ACOS and the aspect-opinion pair filtering in TAS-BERT-ACOS are all set as [8, 5e-5], other settings of these two modules are the same as Extract-Classify-ACOS.

We divide the original dataset into a training set, a validation set and a testing set according to Table 4.

Method	Restaurant-ACOS			Laptop-ACOS		
	P	R	F_1	P	R	F_1
Double-Propagation-ACOS	0.3467	0.1508	0.2104	0.1304	0.0057	0.0800
JET-ACOS	0.5981	0.2894	0.3901	0.4452	0.1625	0.2381
TAS-BERT-ACOS	0.2629	0.4629	0.3353	0.4715	0.1922	0.2731
Extract-Classify-ACOS	0.3854	0.5296	0.4461	0.4556	0.2948	0.3580

Table 5: Main results of the Aspect-Category-Opinion-Sentiment Quadruple Extraction task.

Method	Restaurant-ACOS				Laptop-ACOS			
	EA & EO	IA & EO	EA & IO	IA & IO	EA & EO	IA & EO	EA & IO	IA & IO
Double-Propagation-ACOS	0.2602	N/A	N/A	N/A	0.0980	N/A	N/A	N/A
JET-ACOS	0.5230	N/A	N/A	N/A	0.3570	N/A	N/A	N/A
TAS-BERT-ACOS	0.3360	0.3184	0.1403	0.3976	0.2610	0.4154	0.1090	0.2115
Extract-Classify-ACOS	0.4496	0.3466	0.2386	0.3370	0.3539	0.3900	0.1682	0.1858

Table 6: F_1 score on testing subsets with different aspect & opinion types. EA, EO, IA and IO denote explicit aspect, explicit opinion, implicit aspect and implicit opinion, respectively. N/A means the model can not deal with the corresponding type.

In evaluation, a quadruple is viewed as correct if and only if the four elements as well as their combination are exactly the same as those in the gold quadruple. On this basis, we calculate the Precision and Recall, and use F_1 score as the final evaluation metric for AOCs Quadruple Extraction.

5.2 Main Results

Table 5 reports the ACOS quadruple extraction performance of four different systems on the two datasets. It can be seen that Double-Propagation-ACOS gets the lowest performance. It is reasonable that only using rules is somehow difficult to identify multiple implicit elements and their complex combinations in reviews.

JET-ACOS and TAS-BERT-ACOS achieve comparable F_1 performance: the former is better on Restaurant-ACOS dataset and the latter is better on Laptop-ACOS.

Extract-Classify-ACOS achieves the best performance among four baseline systems. It outperforms JET-ACOS by 5.60 percentage points on Restaurant-ACOS and outperforms TAS-BERT-ACOS by 8.49 percentage points on Laptop-ACOS, respectively. The main advantage is that Extract-Classify-ACOS can achieve robustly higher recall score. In comparison, JET-ACOS has higher or comparable precision score but its recall is much lower.

It is also worth noting that the F_1 score of Extract-Classify-ACOS on both datasets are not high (0.4461 and 0.3580). It is reasonable because the evaluation metric is based on exact matching

and the ACOS Quadruple Extraction is a more complicated task than the traditional ABSA tasks.

5.3 Effectiveness of Modeling of Implicit Aspects/Opinions

As we have mentioned, a large percentage of review sentences contain implicit aspects/opinions. Therefore, efficient modeling of implicit aspects/opinions is of great importance.

To investigate the ability of different systems in addressing the implicit aspects/opinion problem, in Table 6 we split the testing set into four subsets and observe the performance on different subsets: 1) EA & EO denotes the subset with explicit aspects and explicit opinions; 2) IA & EO denotes the subset with implicit aspects and explicit opinions; 3) EA & IO denotes the subset with explicit aspects and implicit opinions; 4) IA & IO denotes the subset with both implicit aspects and implicit opinions.

Among four systems, Double-Propagation-ACOS and JET-ACOS can only address EA & EO, while TAS-BERT-ACOS and Extract-Classify-ACOS can support both implicit aspects and implicit opinions. They show comparable ability in modeling the implicit aspects/opinions. Extract-Classify-ACOS is better in case of IA & EO and EA & IO on Restaurant-ACOS, while TAS-BERT-ACOS is better in case of IA & EO and IA & IO on Laptop-ACOS. But Extract-Classify-ACOS performs significantly better in case of EA & EO on two datasets.

We further compare the performance on differ-

Aspect & Opinion Type		EA & EO	IA & EO	EA & IO	IA & IO
Review Sentence		<i>Keyboard is comfortable and screen is sharp.</i>	<i>Nice, I ordered this just for simple web browsing and personal use.</i>	<i>I noticed the battery went down to 67% for no reason.</i>	<i>We waited for an hour to be seated.</i>
AS Pair	RACL (Chen and Qian, 2020)	<i>screen-Pos</i> ✓ <i>Keyboard-Pos</i> ✓	N/A	✗	N/A
AO Pair	SDRN (Chen et al., 2020a)	<i>screen-sharp</i> ✓ <i>Keyboard-comfortable</i> ✓	N/A	N/A	N/A
ACS Triple	TAS-BERT (Wan et al., 2020)	<i>screen-Design&Feature-Pos</i> ✓ <i>Keyboard-Usability-Pos</i> ✓	✗	<i>battery-Performance-Neg</i> ✓	✗
AOS Triple	JET (Xu et al., 2020)	<i>screen-sharp-Pos</i> ✓ <i>Keyboard-comfortable-Pos</i> ✓	N/A	N/A	N/A
ACOS Quadruple	JET-ACOS	<i>screen-Performance-sharp-Pos</i> ✗ <i>Keyboard-Usability-comfortable-Pos</i> ✓	N/A	N/A	N/A
	TAS-BERT-ACOS	<i>screen-Design&Feature-sharp-Pos</i> ✓ <i>Keyboard-Usability-comfortable-Pos</i> ✓	✗	<i>battery-Performance-NULL-Neg</i> ✓	<i>NULL-Service-NULL-Neg</i> ✓
	Extract-Classify-ACOS	<i>screen-Design&Feature-sharp-Pos</i> ✓ <i>Keyboard-Usability-comfortable-Pos</i> ✓	<i>NULL-General-Nice-Pos</i> ✓	<i>battery-Performance-NULL-Neg</i> ✓	<i>NULL-Service-NULL-Neg</i> ✓

Table 7: The predictions of some representative approaches in five ABSA tasks on review sentences with different aspect & opinion types. EA, EO, IA and IO denote explicit aspect, explicit opinion, implicit aspect and implicit opinion, respectively. N/A stands for non-available; ✓ and ✗ denote correct and false predictions, respectively.

ent subsets. The result shows that the worst performance is obtained on EA & IO rather than IA & IO. One possible reason is that the categories corresponding to IA & IO are relatively regular than EA & IO, and is easier to predict.

5.4 Case study

In Table 7, we further conduct case study by comparing the predictions of some representative approaches on five ABSA tasks including Aspect-Sentiment (AS) Pair extraction, Aspect-Opinion (AO) Pair extraction, Aspect-Category-Sentiment (ACS) Triple extraction, Aspect-Opinion-Sentiment (AOS) Triple extraction, and ACOS extraction.

We choose four different sentences according to whether the aspect/opinion is explicit or implicit, and observe the predictions obtained by different approaches. It can be observed that: 1) RACL (Chen and Qian, 2020) accurately extracts the AS pairs in case of EA & EO, but it does not support implicit aspects and it fails to make predictions in case of EA & IO on our testing sentence; 2) SDRN (Chen et al., 2020a) is only capable of aspect-opinion pair extraction in case of EA & EO; 3) JET (Xu et al., 2020) can only extract aspect-opinion-sentiment triples in case of EA & EO; 4) Although TAS-BERT (Wan et al., 2020) supports aspect-category-sentiment triple extraction for either implicit aspect or implicit opinion, it fails to give accurate predictions in case of IA & EO and IA & IO on our testing sentences; 5) As for the three ACOS baseline systems, JET-ACOS is only capable of ACOS quadruple extraction in case of EA & EO, and has a false prediction. TAS-BERT-ACOS and Extract-Classify-ACOS support ACOS quadruple

extraction in case of both implicit aspects and implicit opinions. TAS-BERT-ACOS performs better than JET-ACOS but still fails in the case of IA & EO. Extract-Classify-ACOS performs generally the best and produces more accurate predictions in all cases.

6 Related Work

Aspect-based sentiment analysis (ABSA) has drawn wide attention during the last decade. As a core task of ABSA, **aspect-based sentiment classification** (ABSC) which aims to detect the sentiment of a given aspect has been extensively studied in the literature (Jiang et al., 2011; Vo and Zhang, 2015; Tang et al., 2015; Wang et al., 2016b; Tang et al., 2016; Zhang et al., 2016; Yang et al., 2017b; Ma et al., 2017; Zhang et al., 2018; Wang et al., 2018, 2019; Xu et al., 2019; Tang et al., 2020; Chen et al., 2020b).

In recent years, on the basis of traditional ABSC, a series of expansion tasks have appeared in this field. We divide these work into the following four categories:

Aspect-Sentiment Pair Extraction. It also can be viewed as joint aspect extraction and ABSC. (Mitchell et al., 2013) first explored the open-domain aspect-sentiment extraction task by designing a variety of conditional random field-based models based on traditional discrete features. With the recent trend of deep learning, researchers have proposed various neural pipeline approaches (Zhang et al., 2015; Hu et al., 2019) or joint learning approaches for this task (Li et al., 2019; Luo et al., 2019; He et al., 2019; Chen and Qian, 2020).

Aspect-Opinion Pair Extraction. (Hu and Liu, 2004) first addressed the task in a pipeline manner. (Chen et al., 2020a) proposed to extract aspect-opinion pairs with a double-channel recurrent network while taking the correlation between aspects and opinions into consideration. (Zhao et al., 2020) designed a span-based multi-task learning framework to extract aspect-opinion pairs jointly. The work on aspect-opinion co-extraction (Wang et al., 2016a, 2017; Yu et al., 2018) can be viewed as the first stage of aspect-opinion pair extraction.

Aspect-Opinion-Sentiment Triple Extraction. Considering the relation between aspect and opinion, (Hu and Liu, 2004) designed a feature-based opinion summary system, which identifies explicit aspect, opinion and sentiment, and integrates them into review opinion summaries. (Qiu et al., 2011) further proposed a Double Propagation method to utilize the syntactic relations between aspects and opinions to iteratively extract the aspect-opinion-sentiment triples. More recently, (Peng et al., 2020) proposed a two-stage framework to first extract aspect-sentiment pairs and opinions separately, followed by matching them to obtain aspect-opinion-sentiment triples. (Xu et al., 2020) further proposed an end-to-end position-aware tagging scheme to model the relations among aspect, opinion and sentiment. (Wu et al., 2020) proposed a Grid Tagging Scheme to address this problem. (Mao et al., 2021; Chen et al., 2021) transformed the triple extraction task into multi-turn machine reading comprehension task and achieved state-of-the-art performances.

Aspect-Category-Sentiment Triple Extraction. Previous two categories only focus on explicit aspect-based sentiment analysis, while ignoring the implicit aspects. To address this issue, (Liu et al., 2005) designed rule-based method to find the corresponding implicit aspects through the opinion existing in the review sentence. With the recent advances of pre-trained models, (Wan et al., 2020) proposed a BERT-based architecture to address this task in an end-to-end fashion.

Since the problem of implicit aspect and implicit opinion has not been systematically addressed in previous studies, in this work we introduce a new task for Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction with implicit aspects and opinions, construct two ACOS Quadruple datasets, and benchmark the task with four baseline systems.

7 Conclusions and Future Work

In this paper, we introduce a new task, Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction, aiming to systematically address the implicit aspect/opinion problem. We construct two new datasets for this task, with ACOS annotations including implicit aspects and implicit opinions. We finally benchmark the task with four baseline systems. Experiments demonstrate the advantages of the new task in aspect-based sentiment analysis with implicit aspects/opinions.

The focus of this paper is the introduction of the new task and datasets. The proposed four baseline systems are relatively simple and leave much room for further improvements. We welcome future work proposing stronger models on this task. We also welcome the usage of our datasets on the other ABSA tasks.

Acknowledgments

This work was supported by the Natural Science Foundation of China (No. 62076133 and 62006117), and the Natural Science Foundation of Jiangsu Province for Young Scholars (No. BK20200463) and Distinguished Young Scholars (No. BK20200018).

References

- Huan-Yuan Chen and Hsin-Hsi Chen. 2016. Implicit polarity and implicit aspect recognition in opinion mining. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 20–25.
- Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020a. Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 6515–6524.
- Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 12666–12674.
- Xiao Chen, Changlong Sun, Jingjing Wang, Shoushan Li, Luo Si, Min Zhang, and Guodong Zhou. 2020b. Aspect sentiment classification with document-level sentiment preference modeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3667–3677.

- Zhuang Chen and Tiejun 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)*, pages 3685–3694.
- Hongliang Dai and Yangqiu Song. 2019. Neural aspect and opinion term extraction with mined rules as weak supervision. *arXiv preprint arXiv:1907.03750*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Zhifang Fan, Zhen Wu, Xinyu Dai, Shujian Huang, and Jiajun Chen. 2019. Target-oriented opinion words extraction with target-fused neural sequence labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 2509–2518.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 504–515.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 537–546.
- 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*, pages 168–177.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual Meeting of the association for computational linguistics (ACL)*, pages 151–160.
- Evgeny Kim and Roman Klinger. 2018. Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*, pages 1345–1359.
- Farek Lazhar and Yamina Tlili Guiyassa. 2016. Mining explicit and implicit opinions from reviews. *Int. J. Data Min. Model. Manag.*, 8:75–92.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, pages 6714–6721.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. Aspect term extraction with history attention and selective transformation. *arXiv preprint arXiv:1805.00760*.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, pages 1–167.
- Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web (WWW)*, pages 342–351.
- Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019. Doer: Dual cross-shared rnn for aspect term-polarity co-extraction. *arXiv preprint arXiv:1906.01794*.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.
- Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. *arXiv preprint arXiv:2101.00816*.
- 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)*, pages 1643–1654.
- Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 8600–8607.
- Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, pages 19–30.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
- Soujanya Poria, Erik Cambria, Lun-Wei Ku, Chen Gui, and Alexander Gelbukh. 2014. A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*, pages 28–37.

- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.
- Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2015. Effective lstms for target-dependent sentiment classification. *arXiv preprint arXiv:1512.01100*.
- Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*.
- Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 6578–6588.
- Duy-Tin Vo and Yue Zhang. 2015. Target-dependent twitter sentiment classification with rich automatic features. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1347–1353.
- Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z Pan. 2020. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 9122–9129.
- Jingjing Wang, Changlong Sun, Shoushan Li, Xiaozhong Liu, Luo Si, Min Zhang, and Guodong Zhou. 2019. Aspect sentiment classification towards question-answering with reinforced bidirectional attention network. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3548–3557.
- Shuai Wang, Sahisnu Mazumder, Bing Liu, Mianwei Zhou, and Yi Chang. 2018. Target-sensitive memory networks for aspect sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 957–967.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. Recursive neural conditional random fields for aspect-based sentiment analysis. *arXiv preprint arXiv:1603.06679*.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 3316–3322.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016b. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing (EMNLP)*, pages 606–615.
- Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for end-to-end fine-grained opinion extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 2576–2585.
- Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 2324–2335.
- Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. *arXiv preprint arXiv:2010.02609*.
- Jie Yang, Yue Zhang, Linwei Li, and Xingxuan Li. 2017a. Yedda: A lightweight collaborative text span annotation tool. *arXiv preprint arXiv:1711.03759*.
- Min Yang, Wenting Tu, Jingxuan Wang, Fei Xu, and Xiaojun Chen. 2017b. Attention based lstm for target dependent sentiment classification. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, pages 5013–5014.
- Jianfei Yu, Jing Jiang, and Rui Xia. 2018. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1):168–177.
- Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.
- 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)*, pages 612–621.
- Meishan Zhang, Yue Zhang, and Duy-Tin Vo. 2016. Gated neural networks for targeted sentiment analysis. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 3087–3093.
- He Zhao, Longtao Huang, Rong Zhang, Quan Lu, et al. 2020. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3239–3248.