

iACOS: Advancing Implicit Sentiment Extraction with Informative and Adaptive Negative Examples

Xiancai Xu* and Jia-Dong Zhang*[†] and Lei Xiong and Zhishang Liu

Brands & Consumers Research Institute, Enbrands Inc., Shenzhen, China

{essen, zhangjd.1, xiongl.1, liuzs.1}@enbrands.com

Abstract

Aspect-based sentiment analysis (ABSA) have been extensively studied, but little light has been shed on the quadruple extraction consisting of four fundamental elements: aspects, categories, opinions and sentiments, especially with implicit aspects and opinions. In this paper, we propose a new method iACOS for extracting **Implicit Aspects with Categories and Opinions with Sentiments**. **First**, iACOS appends two implicit tokens at the end of a text to capture the context-aware representation of all tokens including implicit aspects and opinions. **Second**, iACOS develops a sequence labeling model over the context-aware token representation to co-extract explicit and implicit aspects and opinions. **Third**, iACOS devises a multi-label classifier with a specialized multi-head attention for discovering aspect-opinion pairs and predicting their categories and sentiments simultaneously. **Fourth**, iACOS leverages informative and adaptive negative examples to jointly train the multi-label classifier and the other two classifiers on categories and sentiments by multi-task learning. **Finally**, the experimental results show that iACOS significantly outperforms other quadruple extraction baselines according to the F1 score on two public benchmark datasets.

1 Introduction

Aspect-based sentiment analysis (ABSA) has gained continuous attention during the last decade due to its broad application (Pontiki et al., 2014, 2015, 2016). ABSA aims to extract tuples consisting of closely related elements including the *aspect term*, *opinion term*, *aspect category* and *sentiment polarity*. The aspect term refers to a specific word or phrase in a text that is being evaluated, while the opinion term is a subjective statement in the text that expresses a personal sentiment on the aspect term. Both the aspect term and opinion term

are typically classified into a predefined category and sentiment polarity, respectively. Most of the existing works only extract explicit aspects and opinions but completely ignore the implicit ones that are absent from texts. Some works consider the extraction of implicit aspects (Cai et al., 2020; Wan et al., 2020; Zhang et al., 2021b,a; Mao et al., 2022), implicit opinions (Setiowati et al., 2022), or both (Cai et al., 2021; Peper and Wang, 2022; Xiong et al., 2023; Bao et al., 2023a,b; Hu et al., 2023).

In particular, the study (Cai et al., 2021) firstly attempts to extract implicit aspects and opinions simultaneously, because real textual reviews often contain a significant amount of implicit aspects and opinions. For example, in the product review “Looks nice and the surface is smooth, but certain apps take seconds to respond” (Cai et al., 2021), “surface” is an aspect term and classified into the “Design” category, “smooth” is the opinion term toward this aspect with the “Positive” sentiment. The four elements constitute an quadruple “surface-Design-smooth-Positive”. Obviously, there are two more quadruples: “null-Design-nice-Positive” and “apps-Software-null-Negative”, where “null” stands for an implicit aspect or opinion term that does not appear in the given text. Recently, the two studies (Peper and Wang, 2022; Xiong et al., 2023) have improved implicit quadruple extraction based on contrastive learning, a method that constructs positive and negative examples for each anchor (i.e., a training example). This method attempts to minimize the distance between the anchor and positive examples and maximize the distance between the anchor and negative examples in the latent representation space. Unfortunately, **most existing studies suffer from two limitations**: (1) **Uninformative negative examples**. In contrastive learning, it is crucial to sample informative negative examples that are difficult to distinguish from the positive examples (Schroff et al., 2015). However, exist-

*Equal contribution, alphabetical order of surnames.

[†]Corresponding author.

ing studies often fail to generate such informative negative examples due to the intrinsic nature of random perturbation methods. (2) **Non-adaptive negative examples.** The negative examples lack adaptiveness, as their sampling is not influenced by the current model parameters (Daghghi et al., 2021). As a result, there is significant scope to enhance performance in extracting aspect-category-opinion-sentiment quadruples from texts.

Therefore, this paper proposes a new method based on informative and adaptive negative examples, namely **iACOS**, for extracting **Implicit Aspects with Categories and Opinions with Sentiments**. **First**, iACOS employs the pre-trained encoder BERT (Devlin et al., 2019) to get the context-aware token representation of a text, by which a large amount of knowledge contained in BERT can be transferred into iACOS. Meanwhile, iACOS appends two implicit tokens at the end of texts to capture the semantic representation of implicit aspects and opinions, respectively. **Second**, iACOS builds a sequence labeling model over the context-aware token representation by extending the BIOES¹ tagging scheme to co-extract explicit and implicit aspects and opinions; the aspects-opinion co-extraction is preferentially executed, since it is a relatively simple task (Wang et al., 2017; Wang and Pan, 2018; Dai and Song, 2019) and our extended sequence labeling model can accurately generate aspect and opinion candidates for other subsequent tasks. **Third**, iACOS develops a multi-label classifier with a specialized multi-head attention to predict the category-sentiment combination label of each aspect-opinion candidate pair; this classifier is an end-to-end method for discovering aspect-opinion pairs and predicting their categories and sentiments at the same time to alleviate error propagation in the pipeline solution (Peng et al., 2020; Cai et al., 2020). **Fourth**, iACOS constructs negative examples based on the aspect-opinion co-extraction results to train the classifier. These negative examples are informative and adaptive to current model parameters due to two reasons. (1) They are carefully selected or constructed examples that closely resemble positive examples but are actually negative. Therefore, these examples help in refining the model’s ability to distinguish subtle differences between aspects, opinions, categories, and sentiments that are similar yet distinct. The

¹BIOES is a tagging scheme for sequence labeling and BIOES denotes Begin, Inside, Outside, End and Single, respectively.

informative nature of these examples stems from their relevance and challenge to the model, pushing it to learn more nuanced differentiations. (2) These examples are dynamically generated based on the current state of the model during training. Unlike static negative examples used in traditional models, adaptive examples evolve as the model learns, ensuring that the model is consistently challenged. This adaptiveness is critical in iACOS, as it allows the model to improve its understanding of complex sentiment relationships continuously. Additionally, the negative examples are used to jointly train two other classifiers: one for predicting aspect categories and another for opinion sentiments, using a multi-task learning approach. In this study, we address the critical shortage of labeled data impeding complex ABSA tasks by augmenting training data with negative examples, rather than employing contrastive learning.

The main contributions are listed below:

- We propose a new method iACOS for extracting aspect-category-opinion-sentiment quadruples. iACOS unifies the extraction of *explicit* and *implicit* aspects and opinions based on a sequence labeling model. We develop a multi-label classifier for integrating the prediction on categories, sentiments, and their matched pairs into one unified task to alleviate error propagation in the pipeline solution.
- We leverage informative and adaptive negative examples for jointly training multiple tasks, which significantly improves the effectiveness of quadruple extraction. To the best of our knowledge, this is the first attempt to construct informative and adaptive negative examples as data augmentation for ABSA tasks.
- We conduct extensive experiments to verify the effectiveness of iACOS on the two public benchmark datasets (Cai et al., 2021) for quadruple extraction. Experimental results show that iACOS improves the F1 score significantly, in comparison to other state-of-the-art quadruple extraction techniques for implicit aspects and opinions. Our source code is publicly released at <https://github.com/jiadongzh/iacos>.

The rest of this paper is organized as follows. We highlight related work in Section 2. Our iACOS is presented in Section 3, followed by experimental evaluation in Section 4. Finally, Section 5 concludes this paper.

2 Related Work

This section reviews the recent advances of sentiment quadruple extraction and contrastive learning.

Quadruple Extraction. Recently, Cai et al. (2021) introduce a new task, named aspect-category-opinion-sentiment quadruple extraction, construct two new datasets for this new task, and benchmark the task with four baseline systems. Later, most works (Zhang et al., 2021a; Bao et al., 2022; Mao et al., 2022; Bao et al., 2023a,b; Hu et al., 2023) apply the sequence-to-sequence model T5 to generate a list of quadruples for a given sentence and can be differentiated from one another in terms of the formats of quadruples. For instance, the literature (Zhang et al., 2021a) represents each quadruple as a paraphrase sentence, the reference (Bao et al., 2022, 2023a,b) formats all quadruples as an opinion tree with linearized order, and the research (Mao et al., 2022) denotes each quadruple as an independent path of a tree without linearized order. Other studies (Gao et al., 2022; Wang et al., 2022; Varia et al., 2022) develop a unified generative framework based on the T5 model with instructional prompts for a variety of ABSA tasks including quadruple extraction.

Implicit Aspects and Opinions. Although there are so many existing works on extracting aspects and opinions, most of them completely ignore the implicit aspects and opinions that do not appear in texts. Some recent studies pay attention on extracting implicit aspects (Cai et al., 2020; Wan et al., 2020; Zhang et al., 2021b,a; Mao et al., 2022). For example, the study (Cai et al., 2020) does not mine implicit aspects but directly derives their corresponding categories, the research (Wan et al., 2020) handles implicit aspects by classifying whether aspect terms exist in the sentence for a given category-sentiment pair, and other works (Zhang et al., 2021b,a; Mao et al., 2022) naturally represent implicit aspect terms as “null” in the output sequence based on the sequence-to-sequence model T5. In contrast, few work considers the extraction of implicit opinions. For instance, the work (Setiowati et al., 2022) infers implicit opinions via learning a co-occurrence matrix between aspects and opinions. More comprehensively, the study (Cai et al., 2021) is the first to manage implicit aspects and opinions simultaneously by predicting whether the implicit aspect or opinion exists in a given text. Later, Bao et al. (2023a,b) insert two fake tokens at the beginning of a sentence as the implicit aspect

and opinion term; other works (Peper and Wang, 2022; Xiong et al., 2023; Hu et al., 2023) naturally denote implicit aspect terms as “it” or “null” and implicit opinion terms as “null”.

Contrastive Learning. The works (Peper and Wang, 2022; Xiong et al., 2023) enhance implicit quadruple extraction through contrastively learning a sequence-to-sequence model. The contrastive learning utilizes positive and negative samples to produce better input representations by pushing a given anchor with its positive sample closer together while pulling the anchor with its negative sample farther apart in the latent space. It is essential to select negative samples that are challenging to differentiate from positive ones for effective contrastive learning (Schroff et al., 2015). Specifically, the work (Peper and Wang, 2022) perturbs each anchor representation with a random dropout probability to obtain positive and negative samples, while the work (Xiong et al., 2023) constructs negative samples by randomly replacing aspect and opinion words in positive samples. Nevertheless, existing works often fall short in producing informative and adaptive negative samples, owing to the inherent limitations of random techniques that are independent of current model parameters (Daghghi et al., 2021). In this study, we concentrate on generating more informative and adaptive samples for data augmentation instead of contrastive learning, due to a lack of labeled training data.

3 The Proposed iACOS

We define the research problem in Section 3.1, introduce the inference process in Sections 3.2-3.3, and present the training process in Sections 3.4-3.5.

3.1 Problem Statement

We first define basic concepts and the research problem for this paper.

Token. A text, e.g., a product review, is often segmented into a sequence of words or tokens $\langle w_1, w_2, \dots, w_l \rangle$. Both words and tokens are used interchangeably in this paper.

Aspect term. An aspect term a refers to a word span $\langle w_j, \dots, w_{j+m} \rangle$ ($1 \leq j \leq j+m \leq l$) in the text that represents an attribute or feature being evaluated by the corresponding opinion term(s). All aspects in the text constitute a set A .

Opinion term. An opinion term o refers to a word span $\langle w_k, \dots, w_{k+n} \rangle$ ($1 \leq k \leq k+n \leq l$) in the text that expresses a personal sentiment on

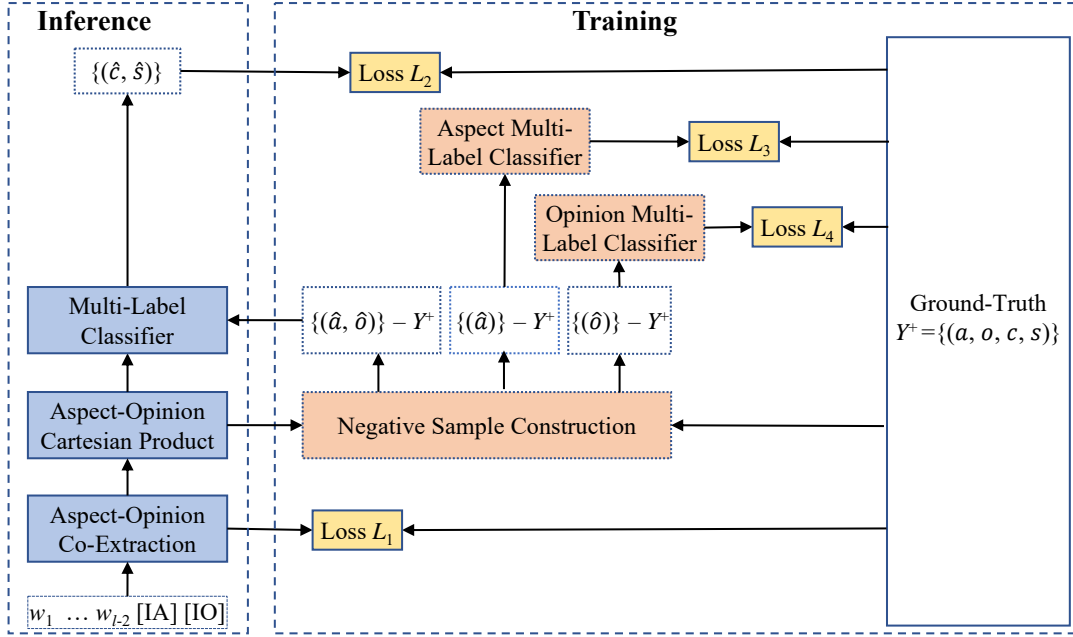


Figure 1: Framework of iACOS: left box for inference and right box for training with multi-tasking learning, in which negative sample construction is an important module.

the corresponding aspect term(s). All opinions in the text constitute a set O .

Category. A category $c \in C$ is a predefined label that is used to classify an aspect a .

Sentiment. A sentiment polarity, or simply sentiment $s \in S$, represents a predefined semantic orientation (e.g., positive, negative, or neutral) expressed by an opinion o .

Quadruple. A quadruple (a, o, c, s) represents the correlation among its four elements.

Research problem. Given a set of training texts, each containing l words $\langle w_1, w_2, \dots, w_l \rangle$ with ground-truth quadruples $Y^+ = \{(a, o, c, s)\}$, we aim to learn a model to extract a set of quadruples $\{(\hat{a}, \hat{o}, \hat{c}, \hat{s})\}$ from a new text.

3.2 Explicit and Implicit Aspect-Opinion Co-Extraction

Representation of implicit aspects and opinions.

People often do not explicitly express their opinions on aspects; it is common to observe implicit aspects and opinions which are absent from a given text. To handle these implicit aspects and opinions, iACOS designs two implicit tokens to capture their semantic representation as done for explicit tokens. As depicted in the left box of Figure 1, at first the two specialized tokens “[IA]” and “[IO]” are appended at the end of a given text. Then iACOS feeds the appended text into the pre-trained encoder BERT (Devlin et al., 2019) to learn the context-

aware representation for all tokens, denoted as:

$$\langle \mathbf{h}_1, \dots, \mathbf{h}_{l-2}, \mathbf{h}_{[IA]}, \mathbf{h}_{[IO]} \rangle = \text{BERT}(\langle w_1, \dots, w_{l-2}, [IA], [IO] \rangle), \quad (1)$$

where without loss of generality, the last two tokens $w_{l-1} = [IA]$ and $w_l = [IO]$ denote implicit aspects and opinions, respectively, and \mathbf{h} is the context-aware representation of a token. It is worth emphasizing that $\mathbf{h}_{[IA]}$ and $\mathbf{h}_{[IO]}$ are the semantic representations of implicit aspects and opinions which are learned from the whole text.

Aspects-opinion co-extraction. As we can see, obtaining implicit aspects and opinions is easy with the two specialized tokens [IA] and [IO]. However, a model is still required to co-extract explicit aspects and opinions. To this end, iACOS builds a sequence labeling model with the extended BIOES tagging scheme over the context-aware token representation. In particular, the extended BIOES tagging scheme consists of nine tags: $T = \{B-A, I-A, E-A, S-A, B-O, I-O, E-O, S-O, O\}$ with the suffix indicating the tag for aspects or opinions. Formally, we can predict the probability distribution $\mathbf{p}_i \in \mathbb{R}^9$ of each token $\mathbf{h}_i \in \mathbb{R}^d$ over nine tags via a linear layer with Softmax:

$$\mathbf{p}_i = \text{Softmax}(\mathbf{W}_1 \mathbf{h}_i + \mathbf{b}_1), \quad (2)$$

where $\mathbf{W}_1 \in \mathbb{R}^{9 \times d}$ and $\mathbf{b}_1 \in \mathbb{R}^9$ are the weight

matrix and bias vector, respectively. The tag is

$$\hat{y}_i = \arg \max_{t \in T} p_{i,t} \text{ and } p_{i,t} \in \mathbf{p}_i. \quad (3)$$

The predicted \hat{y}_i for a given text can be easily decoded to a set of aspects denoted as $A = \{\hat{a}\}$ and a set of opinions denoted as $O = \{\hat{o}\}$. It is worth noting that the tokens [IA] and [IO] for implicit aspects and opinions are always added into A and O , respectively. From now on, it is unified to process both *explicit* and *implicit* aspects and opinions.

3.3 Multi-label Classifier with Multi-head Attention for Quadruple Extraction

Given aspects A and opinions O from Equation (3), iACOS simultaneously performs category and sentiment prediction, and their pair matching process to alleviate the error propagation of the pipeline solution. iACOS considers the Cartesian product $C \times S$ as the combination label set, and predicts multiple combination labels for each aspect-opinion pair because implicit aspects and opinion may have multiple category and sentiment labels, respectively. Given a pair of aspect $\hat{a} = \langle w_j, \dots, w_{j+m} \rangle \in A$ and opinion $\hat{o} = \langle w_k, \dots, w_{k+n} \rangle \in O$, iACOS concatenates the vectors of all the tokens in the aspect and opinion by

$$\mathbf{H}^{(\hat{a}\hat{o})} = [\mathbf{h}_j, \dots, \mathbf{h}_{j+m}, \mathbf{h}_k, \dots, \mathbf{h}_{k+n}] \in \mathbb{R}^{(m+n+2) \times d}, \quad (4)$$

and exploits a multi-head attention (Vaswani et al., 2017) over them to get the attention vector

$$\mathbf{h}^{(\hat{a}\hat{o})} = \text{MultiHead}(\mathbf{q}, \mathbf{H}^{(\hat{a}\hat{o})}, \mathbf{H}^{(\hat{a}\hat{o})}) \in \mathbb{R}^d, \quad (5)$$

where $\mathbf{q} \in \mathbb{R}^d$ is the trainable query, $\mathbf{H}^{(\hat{a}\hat{o})}$ are the keys and values, and the head number is set to 8 by default. Further, the attention vector is fed into a linear layer with Sigmoid to obtain the probability of every combination label:

$$\mathbf{p}^{(\hat{a}\hat{o})} = \text{Sigmoid}(\mathbf{W}_2 \mathbf{h}^{(\hat{a}\hat{o})} + \mathbf{b}_2) \in \mathbb{R}^{|C \times S|}, \quad (6)$$

in which $\mathbf{W}_2 \in \mathbb{R}^{|C \times S| \times d}$ and $\mathbf{b}_2 \in \mathbb{R}^{|C \times S|}$ are the weight matrix and bias vector, respectively. Each entry in $\mathbf{p}^{(\hat{a}\hat{o})}$ from Equation (6) with the probability larger than 0.5 indicates the corresponding category c and sentiment s , i.e., one predicted quadruple $(\hat{a}, \hat{o}, \hat{c}, \hat{s})$.

3.4 Constructing Informative and Adaptive Negative Samples

Optimization objectives. As depicted in the right box of Figure 1, without loss of generality, considering a text with words $\langle w_1, w_2, \dots, w_l \rangle$ and ground-truth quadruples $Y^+ = \{(a, o, c, s)\}$, it is easy to obtain the BIOES tags of all tokens in terms of both ground-truth aspects $\{a\}$ and opinions $\{o\}$. Therefore, we can learn \mathbf{W}_1 and \mathbf{b}_1 in Equation (2) through minimizing the cross-entropy loss:

$$L_1 = \frac{1}{l} \sum_{i=1}^l \mathbf{y}_i \cdot \log \mathbf{p}_i, \quad (7)$$

where $\mathbf{y}_i \in \mathbb{R}^9$ is the one-hot vector of the ground-truth tag of token w_i . Moreover, $Y^+ = \{(a, o, c, s)\}$ may contain quadruples with equal (a, o) pair but different combinations $(c, s) \in C \times S$, that is, an aspect-opinion pair may have multiple combination labels. For instance, in the product review “so what you really end up paying for is the restaurant not the food”, the two quadruples “restaurant-Price-null-Negative” and “restaurant-Ambience-null-Neutral” suggest the restaurant-[IO] pair has two combination labels (Price, Negative) and (Ambience, Neutral) (Cai et al., 2021). Hence, $Y^+ = \{(a, o, c, s)\}$ can be reduced to $Y^+ = \{(a, o, \{(c, s)\})\} = \{(a, o, \mathbf{y}^{(ao)})\}$ in which $\mathbf{y}^{(ao)}$ is the multiple ground-truth combination labels. Accordingly, \mathbf{W}_2 and \mathbf{b}_2 in Equation (6) can be learned through minimizing the binary cross-entropy loss:

$$L_2^+ = \frac{1}{|Y^+| |C \times S|} \sum_{(a,o,\mathbf{y}^{(ao)}) \in Y^+} [\mathbf{y}^{(ao)} \cdot \log \mathbf{p}^{(ao)} + (1 - \mathbf{y}^{(ao)}) \cdot \log(1 - \mathbf{p}^{(ao)})], \quad (8)$$

where $\mathbf{p}^{(ao)}$ is the probability distribution on combination labels of aspect a and opinion o , computed from Equation (4) to Equation (6). Unfortunately, when minimizing the loss L_2^+ in Equation (8) with ground-truth quadruples Y^+ , one problem is that Y^+ is often insufficient to learn a unified model for predicting categories, sentiments, and their matched pairs at the same time.

Negative sample construction. To tackle this problem, iACOS exploits informative and adaptive negative samples to train the unified model. The negative samples are constructed based on the aspect-opinion co-extraction results and hard to be discriminated against ground-truth samples by the current unified model. Further, the method is adaptive since the negative samples are dependent on

the input data and current dynamically updated parameters. The two characteristics are the key to acquire high-quality samples (Daghaghi et al., 2021) and differentiate this method from existing static methods such as random sampling, frequency-based static sampling (Bengio and Senecal, 2008; Mikolov et al., 2013) or learning-based biased sampling (Bamler and Mandt, 2020; Gutmann and Hyvärinen, 2010). Specifically, given the aspect-opinion co-extraction results from a text, i.e., the sets of predicted aspects $A = \{\hat{a}\}$ and opinions $O = \{\hat{o}\}$ according to Equation (3), the Cartesian product $A \times O$ contains all (\hat{a}, \hat{o}) pair candidates that are used to derive quadruples based on Equation (6). In other words, the unified model must learn to tell these candidates apart: some pair candidates are present in the ground-truth quadruples $Y^+ = \{(a, o, \mathbf{y}^{(ao)})\}$ accompanied by their corresponding combination labels $\mathbf{y}^{(ao)}$, while others are not. To this end, iACOS subtracts the ground-truth quadruples Y^+ from the Cartesian product $A \times O$, simply denoted as $Y^- = A \times O - Y^+ = \{(\hat{a}, \hat{o})\} - Y^+$ and considers the remainder pair candidates in Y^- as negative samples. Accordingly, the binary cross-entropy loss is given by

$$L_2^- = \frac{1}{|Y^-||C \times S|} \sum_{(\hat{a}, \hat{o}, \mathbf{y}^{(\hat{a}\hat{o})}) \in Y^-} [\mathbf{y}^{(\hat{a}\hat{o})} \cdot \log \mathbf{p}^{(\hat{a}\hat{o})} + (1 - \mathbf{y}^{(\hat{a}\hat{o})}) \cdot \log(1 - \mathbf{p}^{(\hat{a}\hat{o})})], \quad (9)$$

where $\mathbf{y}^{(\hat{a}\hat{o})} \in \mathbf{0}^{|C \times S|}$ is a zero vector, i.e., the ground-truth label for the negative pair of aspect \hat{a} and opinion \hat{o} . Finally, iACOS adds negative samples Y^- into ground-truth quadruples Y^+ and then computes their loss together:

$$L_2 = \frac{1}{|Y^+ \cup Y^-||C \times S|} \sum_{(\hat{a}, \hat{o}, \mathbf{y}^{(\hat{a}\hat{o})}) \in Y^+ \cup Y^-} [\mathbf{y}^{(\hat{a}\hat{o})} \cdot \log \mathbf{p}^{(\hat{a}\hat{o})} + (1 - \mathbf{y}^{(\hat{a}\hat{o})}) \cdot \log(1 - \mathbf{p}^{(\hat{a}\hat{o})})]. \quad (10)$$

3.5 Multi-task Learning

It is straightforward to train the model for quadruple extraction based on minimizing the sum of losses L_1 and L_2 . Due to the lack of training data, iACOS also leverages both the ground-truth quadruples Y^+ and negative samples Y^- to jointly learn the other two related classifiers for predicting the category of aspects and sentiment of opinions, respectively. Similar to Equation (4), iACOS sepa-

rately concatenates the token vectors by

$$\begin{aligned} \mathbf{H}^{(\hat{a})} &= [\mathbf{h}_j, \dots, \mathbf{h}_{j+m}] \in \mathbb{R}^{(m+1) \times d} \text{ and} \\ \mathbf{H}^{(\hat{o})} &= [\mathbf{h}_k, \dots, \mathbf{h}_{k+n}] \in \mathbb{R}^{(n+1) \times d}, \end{aligned} \quad (11)$$

and get the corresponding attention vectors

$$\begin{aligned} \mathbf{h}^{(\hat{a})} &= \text{MultiHead}(\mathbf{q}_1, \mathbf{H}^{(\hat{a})}, \mathbf{H}^{(\hat{a})}) \text{ and} \\ \mathbf{h}^{(\hat{o})} &= \text{MultiHead}(\mathbf{q}_2, \mathbf{H}^{(\hat{o})}, \mathbf{H}^{(\hat{o})}), \end{aligned} \quad (12)$$

which are fed into a linear layer with Sigmoid to obtain the probability distributions

$$\begin{aligned} \mathbf{p}^{(\hat{a})} &= \text{Sigmoid}(\mathbf{W}_3 \mathbf{h}^{(\hat{a})} + \mathbf{b}_3) \in \mathbb{R}^{|C|} \text{ and} \\ \mathbf{p}^{(\hat{o})} &= \text{Sigmoid}(\mathbf{W}_4 \mathbf{h}^{(\hat{o})} + \mathbf{b}_4) \in \mathbb{R}^{|S|}. \end{aligned} \quad (13)$$

Further, similar to Equation (10), iACOS minimizes the two binary cross-entropy losses

$$L_3 = \frac{1}{|Y^+ \cup Y^-||C|} \sum_{(\tilde{a}, \mathbf{y}^{(\tilde{a})}) \in Y^+ \cup Y^-} [\mathbf{y}^{(\tilde{a})} \cdot \log \mathbf{p}^{(\tilde{a})} + (1 - \mathbf{y}^{(\tilde{a})}) \cdot \log(1 - \mathbf{p}^{(\tilde{a})})] \text{ and} \quad (14)$$

$$L_4 = \frac{1}{|Y^+ \cup Y^-||S|} \sum_{(\tilde{o}, \mathbf{y}^{(\tilde{o})}) \in Y^+ \cup Y^-} [\mathbf{y}^{(\tilde{o})} \cdot \log \mathbf{p}^{(\tilde{o})} + (1 - \mathbf{y}^{(\tilde{o})}) \cdot \log(1 - \mathbf{p}^{(\tilde{o})})], \quad (15)$$

where $(\tilde{a}, \mathbf{y}^{(\tilde{a})})$ or $(\tilde{o}, \mathbf{y}^{(\tilde{o})})$ denotes a projection of $Y^+ \cup Y^-$ on aspects or opinions for simplicity. Eventually, iACOS jointly trains all model parameters by minimizing the total loss with the Adam optimization algorithm on data batches:

$$L = L_1 + L_2 + L_3 + L_4. \quad (16)$$

The multi-task learning improves data efficiency and reduces overfitting because of shared context-aware representations \mathbf{h} among these tasks.

4 Experiments

We present the evaluation setup in Section 4.1 and experimental results in Section 4.2.

4.1 Experimental Setup

Datasets. We use two public benchmark datasets on the quadruple extraction task with implicit aspects and options from the work (Cai et al., 2021) which reports the basic statistics of the two datasets in Table 1. We adopt exactly the same splits on the two datasets for training, validation and testing as the original work.

Compared methods. We compare iACOS with the state-of-the-art baselines on quadruple extraction with implicit aspects and options listed below:

	<i>Restaurant</i>	<i>Laptop</i>
#Categories	13	121
#Sentences	2,286	4,076
#Quadruples	EA&EO	2,429
	IA&EO	530
	EA&IO	350
	IA&IO	349
	All	3,658

Table 1: Statistics of the two datasets from the work (Cai et al., 2021). E, I, A and O denote Explicit, Implicit, Aspect and Opinion, respectively.

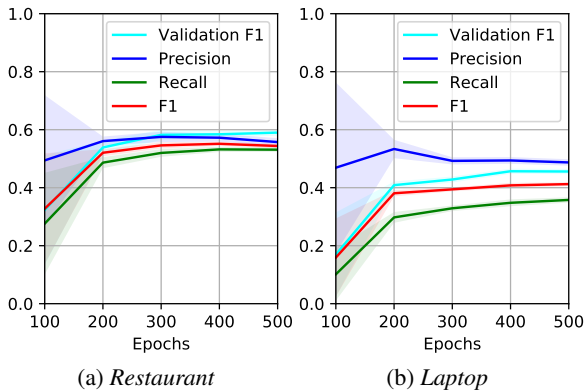


Figure 2: Convergence analysis on iACOS

- **TAS:** It adapts the input transformation strategy of the target-aspect-sentiment model (Wan et al., 2020) to perform category-sentiment conditional aspect-opinion co-extraction, following by filtering out the invalid aspect-opinion pairs to form the final quadruples.
- **Extract-Classify:** It performs aspect-opinion co-extraction and predicts the sentiment polarity of the extracted aspect-opinion pair candidates conditioned on each category (Cai et al., 2021).
- **PARAPHRASE:** It casts the quadruple extraction task to a paraphrase generation process that jointly detects all four elements (Zhang et al., 2021a) and has been adapted for implicit aspects and opinions (Xiong et al., 2023).
- **GEN-NAT-SCL:** It uses a contrastive learning objective to aid quadruple prediction by encouraging the model to produce input representations (Peper and Wang, 2022).
- **BART-CRN:** It is a BART-based contrastive and retrospective network (BART-CRN) that learns

the associations among all types of quadruples (Xiong et al., 2023).

To ensure fairness, we focus on models with BERT or BART backbones, excluding the larger and stronger T5 models (Bao et al., 2023a,b; Hu et al., 2023). Our work centers on a novel method of employing informative and adaptive negative examples for joint multi-task training, which could improve performance when applied to stronger backbones like T5. Thus, our approach could lead to surpassing current top results.

Evaluation metrics. In line with existing studies (Zhang et al., 2021a; Cai et al., 2021), the Precision, Recall, and F1 score are adopted as the main evaluation metrics. Moreover, we view a predicted quadruple as correct if and only if the four elements as well as their combination are exactly the same as those in the ground-truth quadruples.

Experimental settings. We adopt the pre-trained BERT as the backbone and use the AdamW optimizer to minimize the total loss. The hyperparameters are determined based on existing studies (Zhang et al., 2021a; Cai et al., 2021) and several trials on the validation data instead of exhausting grid search. By default, we respectively set the batch size, learning rate and attention head number to 32, 1e-5 and 8 for both datasets. All experiments are carried out with an RTX 3090 GPU and the results are obtained by averaging 10 trials with different random seeds on testing data. Following Guo et al. (2020), we train our model for 500 epochs due to two main reasons: (1) Multi-task learning often needs more epochs to converge due to its complex objectives and task balancing. (2) We notice continued performance improvement beyond the usual training duration.

4.2 Experimental Results

Convergence analysis. Our iACOS shows quite stable and consistent performance on different trials. Figure 2 depicts the convergent process with respect to the number of epochs on a trial, in which 500 epochs are equally divided into five bins and the mean performance is calculated for each bin, along with standard deviation boundaries. After 200 epochs, iACOS reaches relatively stable performance and the F1 score steadily and slowly increases on both validation data and testing data. After 300 epochs, the standard deviation is negligible, and although the validation F1 score remains increasing, the F1 score records the maximum value

Methods	<i>Restaurant</i>			<i>Laptop</i>		
	Precision	Recall	F1	Precision	Recall	F1
TAS	0.2629	0.4629	0.3353	0.4715	0.1922	0.2731
Extract-Classify	0.3854	0.5296	0.4461	0.4556	0.2948	0.3580
PARAPHRASE	0.4362	0.3619	0.3956	0.3636	0.2963	0.3265
GEN-NAT-SCL	0.4893	0.4051	0.4432	0.3713	0.3244	0.3463
BART-CRN	0.5084	0.4710	0.4890	0.4816	0.3183	0.3832
iACOS	0.5724	0.5321	0.5515	0.4959	0.3465	0.4080
std	± 0.0095	± 0.0079	± 0.0072	± 0.0121	± 0.0101	± 0.0082

Table 2: Performance comparison on the two datasets with implicit aspects and opinions. The results of compared methods are from the previous works (Cai et al., 2021; Xiong et al., 2023).

Methods	<i>Restaurant</i>				<i>Laptop</i>			
	EA&EO	IA&EO	EA&IO	IA&IO	EA&EO	IA&EO	EA&IO	IA&IO
TAS	0.3360	0.3184	0.1403	0.3976	0.2610	0.4154	0.1090	0.2115
Extract-Classify	0.4496	0.3466	0.2386	0.3370	0.3539	0.3900	0.1682	0.1858
PARAPHRASE	0.3852	0.3780	0.1667	0.3850	0.3130	0.3892	0.2111	0.3556
GEN-NAT-SCL	0.4692	0.3053	0.2051	0.3763	0.3593	0.407	0.2085	0.3022
BART-CRN	0.5413	0.5064	0.1893	0.4286	0.3891	0.5430	0.2450	0.4071
iACOS	0.6166	0.4778	0.2491	0.4345	0.4201	0.5808	0.2394	0.4124
std	± 0.0093	± 0.0139	± 0.0085	± 0.0108	± 0.0101	± 0.0078	± 0.0103	± 0.0105

Table 3: F1 score on testing subsets with different aspect & opinion types. E, I, A and O denote Explicit, Implicit, Aspect and Opinion, respectively. The results of compared methods are from the previous works (Cai et al., 2021; Xiong et al., 2023).

at 400 epochs on *Restaurant* testing data and at 500 epochs on *Laptop* testing data. Hereafter, unless otherwise specified, we report results at 400 epochs on testing data.

Overall comparison. Table 2 compares the performance of all evaluated methods. Our iACOS consistently achieves the best results averaged at ten random trials with negligible standard deviation on both datasets. Note that the original references (Cai et al., 2021; Xiong et al., 2023) have not reported the standard deviation for the other methods. Compared to the second best BART-CRN, iACOS relatively improves the F1 score by 12.78% and 6.47% on *Restaurant* and *Laptop* datasets, respectively.

Further, we conform to the approach outlined in the reference (Cai et al., 2021) by focusing on the four principal combinations: EA&EO, IA&EO, EA&IO, and IA&IO. Table 3 demonstrates the performance on different testing subsets. iACOS reaches the highest F1 score among all evaluated methods in most cases, especially on the two testing subsets: EA&EO and IA&IO. BART-CRN has a better performance on IA&EO of *Restaurant* and EA&IO of *Laptop*. One reasonable explanation is that the proportion of IA&EO in *Restaurant* or EA&IO in *Laptop* is higher than the other im-

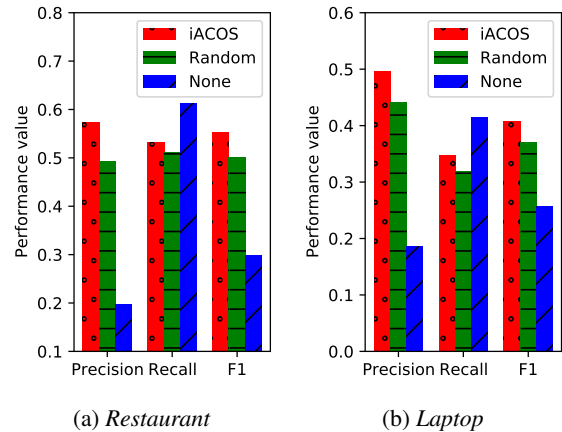


Figure 3: Effect of negative samples

PLICIT testing subsets as shown in Table 1, which helps BART-CRN to fully capture the input features (Xiong et al., 2023). These results indicate the effectiveness of iACOS with informative and adaptive negative samples.

Study on negative samples. Figure 3 depicts the effect of different negative sample construction methods with three findings. Firstly, “None” does not apply any negative samples, i.e., training with ground-truth quadruples Y^+ only. Its precision and F1 score decrease severely, even though it records

Methods	<i>Restaurant</i>				<i>Laptop</i>			
	EA&EO	IA&EO	EA&IO	IA&IO	EA&EO	IA&EO	EA&IO	IA&IO
iACOS	0.6166	0.4778	0.2491	0.4345	0.4201	0.5808	0.2394	0.4124
Random	0.5554	0.4508	0.2464	0.2715	0.3975	0.5452	0.1857	0.2190
None	0.4360	0.1751	0.1385	0.1931	0.3529	0.1992	0.1434	0.0817

Table 4: Effect of negative samples on the extraction of different aspect & opinion types.

the highest recall. One reason is that without negative samples, it is prone to extract more aspects and opinions from texts and results in proposing more quadruples. Secondly, the random method generates the sets of aspects and opinions randomly instead of employing the sequence labeling model presented in Section 3.2, and then follows the same remainder process as iACOS. The random method outperforms “None” in terms of the F1 scores on both datasets, which indicates that these negative samples are helpful to improve the model performance regardless of underlying sampling methods. Finally, iACOS constructs information and adaptive samples based on the aspect-opinion co-extraction results and increases the F1 score by 8% at least on both datasets in comparison to the random method. This indicates a better construction method can bring larger performance improvement.

Furthermore, Table 4 shows the effect of negative samples on the extraction of different aspect & opinion types. The “None” condition, which does not apply any negative samples, results in the lowest F1 scores across all cases. This observation allows us to conclude that negative samples have a positive effect on the extraction of all aspect and opinion types. This outcome occurs because our proposed method is not specifically tailored for IA&IO.

Ablation study. Figure 4 illustrates the influence of implicit tokens, multi-head attention and multi-task learning in iACOS. (1) Without adding implicit tokens using the [CLS] token of BERT for implicit aspects and opinions, the performance of iACOS degrades noticeably in both the *Restaurant* and *Laptop* domains. The reason is that this alternative method cannot differentiate between implicit aspects and implicit opinions, resulting in significantly lower performance on IA&EO and EA&IO, particularly on IA&IO. (2) Without multi-head attention by simply taking average of all vectors of \mathbf{H} in Equation (5), iACOS encounters underfitting and reports the lowest F1 score, especially on *Laptop* domain that has much more aspect categories

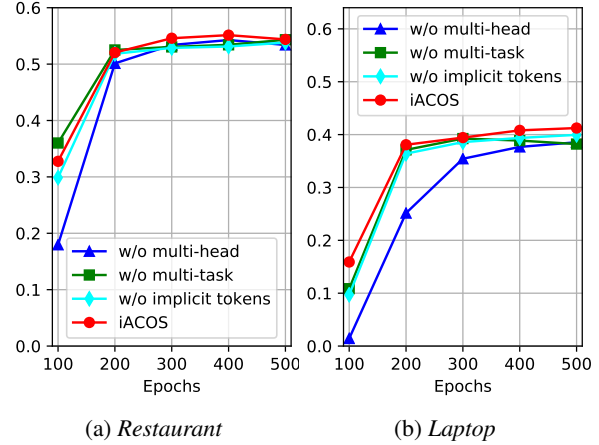


Figure 4: Ablation experiments

than *Restaurant* domain. This result justifies that multi-head attention plays an important role in augmenting model capacity by enabling the simultaneous capture of diverse features and relationships within the input data, leading to improved representation learning and overall performance gains. (3) Without multi-task learning by minimizing the sum of L_1 and L_2 rather than the total loss in Equation (16), iACOS quickly converges, suffers from overfitting, and deteriorates performance on *Laptop* domain. This result indicates that the multi-task learning enhances model generalization, improves predictive accuracy, and enables effective knowledge transfer across related tasks.

5 Conclusion

In this paper, we propose iACOS, a novel approach for extracting implicit sentiment quadruples with multi-label classifier and multi-head attention over the context-aware representation of implicit aspects and opinions. Furthermore, we devise an informative and adaptive sample construction method for generating negative examples to train multiple classifiers by multi-task learning. Experiment results have verified our method’s effectiveness and superiority in comparison to existing strong baselines.

Limitations

First, we have not provided theoretical justification for the proposed informative and adaptive sampling method. Second, our model is only evaluated on the quadruple extraction task and its effectiveness on other ABSA tasks is unknown. Third, we have not extensively investigated the effect of various hyper-parameters, e.g., the batch size, learning rate and attention head number. Lastly, we have not explored applying our negative sample construction method to large language models. Despite these limitations, our study provides valuable insights into the effectiveness of iACOS for extracting implicit sentiment quadruples and suggests areas for future research.

References

- Robert Bamler and Stephan Mandt. 2020. Extreme classification via adversarial softmax approximation. In *Proceedings of the 8th International Conference on Learning Representations*, Addis Ababa, Ethiopia.
- Xiaoyi Bao, Xiaotong Jiang, Zhongqing Wang, Yue Zhang, and Guodong Zhou. 2023a. Opinion tree parsing for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7971–7984, Toronto, Canada.
- Xiaoyi Bao, Zhongqing Wang, and Guodong Zhou. 2023b. Exploring graph pre-training for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3623–3634, Singapore.
- Xiaoyi Bao, Wang Zhongqing, Xiaotong Jiang, Rong Xiao, and Shoushan Li. 2022. Aspect-based sentiment analysis with opinion tree generation. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, pages 4044–4050, Vienna, Austria.
- Yoshua Bengio and Jean-SÉbastien Senécal. 2008. Adaptive importance sampling to accelerate training of a neural probabilistic language model. *IEEE Transactions on Neural Networks*, 19(4):713–722.
- 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*, pages 833–843, Barcelona, Spain.
- Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 340–350, Online.
- Shabnam Daghighi, Tharun Medini, Nicholas Meisburger, Beidi Chen, Mengnan Zhao, and Anshumali Shrivastava. 2021. A tale of two efficient and informative negative sampling distributions. In *Proceedings of the 38th International Conference on Machine Learning*, pages 2319–2329, Online.
- Hongliang Dai and Yangqiu Song. 2019. Neural aspect and opinion term extraction with mined rules as weak supervision. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5268–5277, Florence, Italy.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *The 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. 2022. LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7002–7012, Gyeongju, Korea.
- Pengsheng Guo, Chen-Yu Lee, and Daniel Ulbricht. 2020. Learning to branch for multi-task learning. In *Proceedings of the 37th International Conference on Machine Learning*, pages 3854–3863, Online.
- Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 297–304, Sardinia, Italy.
- Mengting Hu, Yin hao Bai, Yike Wu, Zhen Zhang, Liqi Zhang, Hang Gao, Shiwan Zhao, and Minlie Huang. 2023. Uncertainty-aware unlikelihood learning improves generative aspect sentiment quad prediction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13481–13494, Toronto, Canada.
- Yue Mao, Yi Shen, Jingchao Yang, Xiaoying Zhu, and Longjun Cai. 2022. Seq2path: Generating sentiment tuples as paths of a tree. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2215–2225, Dublin, Ireland.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, pages 3111–3119. Curran Associates, Inc., Lake Tahoe, NV.
- 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 AAAI Conference on Artificial Intelligence*, pages 8600–8607, New York City, NY.
- Joseph Peper and Lu Wang. 2022. Generative aspect-based sentiment analysis with contrastive learning and expressive structure. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6089–6095, Abu Dhabi.
- 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 Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, pages 19–30, San Diego, CA.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 486–495, Denver, CO.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Haris 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*, pages 27–35, Dublin, Ireland.
- Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, Boston, MA.
- Yuliana Setiowati, Arif Djunaidy, and Daniel Oranova Siahaan. 2022. Aspect-based extraction of implicit opinions using opinion co-occurrence algorithm. In *Proceedings of the 5th International Seminar on Research of Information Technology and Intelligent Systems*, pages 781–786, Yogyakarta, Indonesia.
- Siddharth Varia, Shuai Wang, Kishaloy Halder, Robert Vacareanu, Miguel Ballesteros, Yassine Benajiba, Neha Anna John, Rishita Anubhai, Smaranda Muresan, and Dan Roth. 2022. Instruction tuning for few-shot aspect-based sentiment analysis. *arXiv preprint arXiv:2210.06629*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, pages 5998–6008. Curran Associates, Inc.
- 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 AAAI Conference on Artificial Intelligence*, pages 9122–9129, New York City, NY.
- Wenya Wang and Sinno Jialin Pan. 2018. Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2171–2181, Melbourne, Australia.
- 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 AAAI Conference on Artificial Intelligence*, pages 3316–3322, San Francisco, CA.
- Zengzhi Wang, Rui Xia, and Jianfei Yu. 2022. UnifiedABSA: A unified ABSA framework based on multi-task instruction tuning. *arXiv preprint arXiv:2211.10986*.
- Haoliang Xiong, Zehao Yan, Chuhan Wu, Guojun Lu, Shiguan Pang, Yun Xue, and Qianhua Cai. 2023. BART-based contrastive and retrospective network for aspect-category-opinion-sentiment quadruple extraction. *International Journal of Machine Learning and Cybernetics*, 14(9):3243–3255.
- 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*, pages 9209–9219, Punta Cana, Dominican.
- 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*, pages 504–510, Online.