

## Article

# A Novel Cascade Model for End-to-End Aspect-Based Social Comment Sentiment Analysis

Hengbing Ding<sup>1</sup>, Shan Huang<sup>2</sup>, Weiqiang Jin<sup>3</sup> , Yuan Shan<sup>4</sup> and Hang Yu<sup>3,\*</sup> 

- <sup>1</sup> School of Electrical Engineering, Shanghai Dianji University, Shanghai 201306, China; dinghb@sdju.edu.cn  
<sup>2</sup> Faculty of arts and social science, University of Technology Sydney, Sydney 2002, Australia; shanhuang0206@gmail.com  
<sup>3</sup> School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; weiqiangjin@shu.edu.cn  
<sup>4</sup> ACRE Coking & Refractory Engineering Consulting Coporation (Dalian), Metallurgical Corporation Of China, Dalian 116085, China; shanyuan@acre.com.cn  
\* Correspondence: yuhang@shu.edu.cn

**Abstract:** The end-to-end aspect-based social comment sentiment analysis (E2E-ABSA) task aims to discover human's fine-grained sentimental polarity, which can be refined to determine the attitude in response to an object revealed in a social user's textual description. The E2E-ABSA problem includes two sub-tasks, i.e., opinion target extraction and target sentiment identification. However, most previous methods always tend to model these two tasks independently, which inevitably hinders the overall practical performance. This paper investigates the critical collaborative signals between these two sub-tasks and thus proposes a novel cascade social comment sentiment analysis model for jointly tackling the E2E-ABSA problem, namely CasNSA. Instead of treating the opinion target extraction and target sentiment identification as discrete procedures in previous works, our new framework takes the contextualized target semantic encoding into consideration to yield better sentimental polarity judgment. Additionally, extensive empirical results show that the proposed approach effectively achieves a 68.13% F1-score on SemEval-2014, 62.34% F1-Score on SemEval-2015, 56.40% F1-Score on SemEval-2016, and 50.05% F1-score on a Twitter dataset, which is higher than the existing approaches. Ablated experiments demonstrate that the CasNSA model substantially outperforms state-of-the-art methods, even when using fixed words embedding rather than pre-trained BERT fine tuning. Moreover, in-depth performance analysis on the social comment datasets further validates that our work gains superior performance and reliability effectively and efficiently in realistic scenarios.

**Keywords:** machine learning; natural language processing; fine-grained sentiment analysis; fine-tuning transformer BERT; multi-level cascade tagging scheme



**Citation:** Ding, H.; Huang, S.; Jin, W.; Shan, Y.; Yu, H. A Novel Cascade Model for End-to-End Aspect-Based Social Comment Sentiment Analysis. *Electronics* **2022**, *11*, 1810. <https://doi.org/10.3390/electronics11121810>

Academic Editor: George A. Papakostas

Received: 30 March 2022

Accepted: 31 May 2022

Published: 7 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid advent of the internet social media era, aspect-level sentiment analysis has received a lot of attention from both industry and academic communities. No matter the e-commerce market, such as eBay or social forums such as Twitter and YouTube, to improve users' experiences and ensure its market competitiveness, it has placed increasing attentions toward its users' expressed sentiment polarity identification. Jalil et al. [1] gathered a large quantity of heterogeneous "COVID-Senti" data from the social networking media platform, Twitter. After the sophisticated data pre-processing procedure, they analyzed the COVID-19 influence and the vaccinations' effectiveness by focusing on these Twitter user sentiments classification. To better embrace the industrial internet of things (IIoT) or Industry 4.0 era, Khan et al. [2] focused on generalized aspect-based category detection (ACD) and proposed a novel convolutional attention-based bidirectional LSTM

for the detection of customer opinion and emotions. In this way, IIoT can provide better services by analyzing users' feedback.

Fine-grained sentiment polarity analysis from unstructured text is an essential task in automatic public opinions detection and consumer reviews attitude recognition. In general, aspect-based sentiment analysis (ABSA) has been a focus of research in recent years. Aspect-based sentiment analysis [3], which identifies people's opinions on specific topics, is an extraction and classification fused problem in natural language processing (NLP). Various ABSA works focus on the sentiment polarity (positive, neutral, and negative) of a target word in given comments or reviews [4,5]. For example, in the restaurant service scenario, "Nice *ambience*, but the highly overrated *fast-food restaurant* .", the consumer mentions two opinion targets, namely "*ambience*" and "*fast-food restaurant*", and expresses a positive attitude toward the first one, and negative sentiment toward the second one.

As the dominant line of research in fine-grained opinion mining, aspect-based sentiment analysis (ABSA) aims to identify sentiment of target entities and their aspects. Specifically, given a target entity of interest, ABSA methods can extract its properties and identify the sentiment expressed about those properties [6]. From a technological point of view, the methodologies can be divided into two sub-tasks, namely opinion target extraction and target sentiment identification [4,5], which corresponds to the above-mentioned interested target entity properties extraction and expresses sentiment identification.

Despite the previous success made in the sentiment analysis research area, most existing methods ignore the interactive relations among these sub-tasks. As a result, existing methods model these sub-tasks independently of each other, which hinders their practical use performance, i.e., the goal of some methods [7–10] is only to detect the opinion target mentioned in the text. In addition, other methods [11–15] identifying target sentiment polarity assume that the target mention is given. To perform the ABSA task in more practical settings, i.e., extracting the targets and the corresponding sentiments simultaneously, one typical way is to pipeline these two sub-tasks end-to-end. Essentially, these existing pipeline methods [7,16–19] still treat these sub-tasks as separate two steps and are not sufficient to yield satisfactory results for the complete ABSA task. These pipeline methods complete the ABSA task through target boundary tags (e.g., B, I, E, S, and O) and sentiment tags (e.g., POS, NEG, and NEU) prediction. In recent years, some ideal joint approaches [17,19–24] for ABSA have been proposed, which regard it as complex integrated boundaries and types tagging the prediction task, and make the two sub-tasks jointly trained. These joint methods differ from the above pipelines, and they utilize a set of specially designed tags integrated from discrete target boundary and sentiment tagging tasks, namely B-{POS, NEG, NEU}, I-{POS, NEG, NEU}, E-{POS, NEG, NEU}, S-{POS, NEG, NEU}, O-{POS, NEG, NEU}, denoting the beginning of, inside of, end of, and single-word opinion target with the positive, negative and neutral sentiment, and O denoting the positions of none of the sentiment words, respectively. An example of introducing the differences between these tagging schemes is shown in Table 1, and we can intuitively discover that the "Integrated" tagging solution, which is presented in Row 2, is relatively more complex than the "Discrete" tagging scheme. The "Integrated" scheme greatly expands the corresponding tagging species and huge search space, which is prone to increasing the complexity of tagging predictions and decrease the performance of the overall ABSA tagger.

**Table 1.** The discrete tagging approach and integrated tagging approach. "Discrete" and "Integrated" refers to these two tagging schemes, respectively.

Input	Nice	Ambience	,	But	Highly	Overrated	Fast-Food	Restaurant	.
Discrete	O	S	O	O	O	O	B	E	O
	O	POS	O	O	O	O	NEG	NEG	O
Integrated	O	S-POS	O	O	O	O	B-NEG	E-NEG	O

Through the above comprehensive and detailed analysis of those popular pipeline and joint methods, we judiciously deem that it is crucial to investigate the interactive relationships between these two sub-tasks for determining the target-oriented sentiment polarity more accurately. In the relational triple extraction research area, the work *CasREL* [25] and *tagging decomposition strategy* [26] firstly decompose the joint relational triples extraction into head-entity (**HE**) and tail-entity and relation (**TER**) extraction sub-tasks, and formulate the relations mapped from the subjects detection to objects detection, which finally efficiently solves the triple-overlapping problem. Inspired by these works, we investigate and exploit the high coupled inter-dependency of these sub-tasks and propose a novel cascade social comment sentiment analysis model, namely **CasNSA**, to guide the identification of sentiment polarity with the auxiliary task of the target boundary prediction.

Specifically, our framework CasNSA contains three principal components: the contextual semantic representation (**CSR**) module, target boundary recognizer (**TBR**) and sentiment polarity identifier (**SPI**). Firstly, the CSR component encodes the social comment sentences and then provides embedded vector sequences. To further investigate the representation power of transformers [27], we simultaneously adopt a transformer-based pre-trained language encoder (e.g., **BERT** [20,28]) and pre-trained word embedding (e.g., **Word2Vec** [29], and **GloVe** [30]) in this part. The CSR component finally provides a hidden state, which ranges from the start to end token throughout the inputted sentence. In the following, The TBR module utilizes the hidden states generated by the CSR component as the inputted sentence's semantics. At the sentiment inference time, the SPI module utilizes the hidden state on the position "[CLS]" of the inputted token sequence and takes the advantages of the target boundary information generated from the TBR module. In this manner, the novel hierarchical tagging method constrains the sentiment analysis within the specific opinion-target context and thus achieves better overall performance for the ASBA task.

Generally speaking, compared with other mainstream approaches for the ABSA task, our method is a simple yet insightful neural network architecture. Our CasNSA regards the opinion-target extraction as the sequence labeling problem and regards the target-oriented sentiment analysis as the multi-label identification problem, respectively. At the same time, we also conduct a series of contrastive ablation experiments by designing different constructions of the CSR component. Our experimental results demonstrate the BERT's absolute modeling advantage over traditional RNNs based on pre-trained fixed embeddings (e.g., Word2Vec, and GloVe). As a result, this proves BERT's powerful semantic comprehension capabilities, one of the most popular transformer-based pre-trained language models.

Additionally, experiments prove that the proposed approach provides better F1-scores of 68.13% , 62.34% , 56.40% and 50.05% for SemEval 2014 [31], 2015 [32], 2016 [33] and the Twitter dataset [34], respectively. To some extent, these experiments prove that our proposed model outperforms on real-life datasets compared to many previously widely used methods [14,17,20].

The principal contributions in this work can be summarized as follows:

- We go deep into the complete aspect-based sentiment analysis task, and formulate it as the sequence tagging problem of the opinion-target extraction sub-task (**OTE**) and the multi-label identification problem for the target-oriented sentiment analysis sub-task (**TSA**). To be specific, we introduce a novel ABSA approach named **CasNSA**, which is composed of three main sub-modules: *contextual semantic representation module*, *target boundary recognizer* and *sentiment polarity identifier*.
- While many methods attempt to model sub-tasks' correlations, including machine learning methods and deep learning methods, our method attempts to address the ABSA task by the neural network construction. To the best of our knowledge, the unique modeling mechanism is proposed to handle the ABSA task for the first time. Based on our formulated deduction for the complete ABSA task, we employ the specific opinion-target context representation provided by the OTE into the TSA

procedure, which can further constrain and guide the sentiment polarity analysis, thus achieving better performance for the complete end-to-end ABSA task.

- The further empirical comparison result confirms the effectiveness and rationality of our model **CasNSA**. We first consider the interactive relations modeling between opinion target determining and target-specific sentiment identification. Furthermore, the ablation study also proves that the transformer-based BERT is more efficient in contrast with traditional pre-trained embedding methods (e.g., Word2Vec and GloVe).

This paper is structured as follows: Section 1 introduces the *aspect-based social comment sentiment analysis* research background, and Section 2 describes the most relevant related works. Section 3 is the core of this paper since it contains the preliminaries of the **E2E-ABSA** task; the interactive relations formulation among the **OTE** sub-task and the **TSA** sub-task; and the architecture details of our proposed **CasNSA**. Section 4 discusses all validation experiments, including a comparison and ablation study on several widely used datasets. Finally, Section 5 highlights the major conclusions of this work and the potential future works.

## 2. Related Works

When reviewing the literature on the ABSA task, we summarize that the existing methods can be classified into either separate approaches or end-to-end approaches. In contrast, the end-to-end research branches can still be subdivided into pipeline approaches and unified approaches.

### 2.1. Separate Approaches

It is common for a complex NLP task to decompose the task into several separate sub-processes. The ABSA task is also treated as two-steps containing opinion-target extraction (**OTE**) and target-oriented sentiment analysis (**TSA**) by most existing studies. Some studies develop separate methods for **OTE** [7–10], while other methods develop for **TSA** [11–15,35,36].

Li et al., 2018 [7] exploited two useful clues, namely opinion summary and aspect detection history, and presented a new framework for tackling aspect term extraction. Experiments over four benchmarks prove that the method can leverage the coordinate and tagging schema constraints to help increase the aspect term extraction performance. Jaechoon et al., 2021 [8] used the statistical analysis of sentence structures to extract the sentiment-target word pairs. They proposed a model which contains a sentiment word extractor and a target word extractor by utilizing parsing and statistical methods. The proposed model shows high performance in both accuracy and F1-score, compared with others over the dataset containing 4000 movie reviews. Xu et al., 2018 [9] focused on supervised product reviews aspect extraction using deep learning. Unlike other highly sophisticated supervised neural network models, they first proposed a simple but effective CNN model employing double pre-trained embeddings, which is composed of general-purpose embeddings and domain-specific embeddings. Their model outperforms existing sophisticated SOTA methods without using any additional supervision.

To address the issue that RNNs are different to analyze long-term patterns, Song et al., 2019 [35] proposed an attentional encoder network which eschews recurrence and employs pre-trained BERT to model the context and target. Experiments show that their method achieves a new SOTA on three public datasets. Wang et al., 2018 [13] observed that the current memory network's sentiment polarity detection depends on the given target and cannot be inferred from the context. They proposed the target-sensitive memory network to tackle this problem, and its effectiveness was experimentally evaluated. Hazarika et al., 2021 [11] incorporated a variety of inter-aspect dependency patterns by all aspects' simultaneous classification along with sentence temporal dependency representations using RNNs. Their model effectively exploits the contextual information of the inter-aspect dependencies, and the benchmark SemEval 2014 [31] suggests their approach's effectiveness. To alleviate the dilemma that most previous models rely on large-scale source data while corpora la-

being is especially expensive and labor-intensive, Li et al. [14] exploited a novel approach named the coarse-to-fine task transfer, which aims to leverage knowledge learned from a rich-resource domain. In addition, they proposed a multi-granularity alignment network to resolve both the challenges of aspect inconsistency and cross domain's feature mismatch. Empirically, extensive experiments demonstrate the effectiveness of their framework.

Aiming at reducing the gap between human-level interpretability and the accuracy of the ABSA task, Yadav et al. [37] designed an interpretable learning architecture for aspect-based sentiment analysis (ABSA), employing the recently introduced Tsetlin Machines [38]. Their experiment successfully provided a human-interpretable ABSA architecture with comparable accuracy. It is worth noticing that the main selling point of their method is the transparent and interpretable learning approach rather than accuracy.

## 2.2. Pipeline Approaches

In order to make these separate works in practical use for the complete ABSA task, one typical way is to pipeline these two sub-tasks together to a relatively integrated method [6,7,9,16–19,39].

To alleviate the challenge that a sequence tagging formulation always suffers from a huge search space and sentiment inconsistency, Hu et al., 2019 [16] proposed a span-based extract-then-classify pipeline framework. In this pipeline framework, multiple opinion targets are directly extracted from the sentence under the target span boundaries' supervision, and corresponding polarities are then classified using their span representations. It is worth noticing that they further investigated the pipeline, joint, separate variant models, and their pipeline model achieved the best performance compared with the other two models. Li et al. [17] heuristically combined their previous works HAST [7] and TNet [18] as a pipeline approach, in which these two sub-modules are the current state-of-the-art models on the tasks of target boundary detection and target sentiment classification, respectively. To demonstrate the effectiveness of the work, Chen et al. [19] proposed RACL for which they constructed four pipeline baselines through several SOTA separate methods' combinations, with DECNN [9] and CMLA [39] for the sub-task of OTE, TCap [40] and TNet [18] for the sub-task of TSA.

## 2.3. Unified Approaches

Unfortunately, the above pipeline studies usually suffer from the error propagation challenge from the upstream to the downstream, and the two sub-tasks' inter-relevance is largely neglected. To alleviate these issues, many joint frameworks [17,19–24,41,42] aim to learn the OTE, and TSA sub-tasks jointly are proposed, as these collaborative features can enhance the two sub-tasks in a mutual way.

Finding that the sentiment scope within which each named entity is embedded is typically not explicitly annotated in the data, and unlike traditional methods that cast this task as simple sequence labeling, Li et al. (2017) [41] proposed a novel approach that can explicitly model the latent sentiment scopes and achieve better results compared to existing approaches [43] based on conventional conditional random fields (CRFs) [44]. Ma et al., 2018 [24] carefully designed the hierarchical stack bidirectional gated recurrent units (HSBi-GRU) to jointly learn both sub-tasks' abstract features. They proposed a HSBi-GRU based joint model that allows the target label to have an influence on their judgement for the sentiment label. Experiments demonstrate that their joint learning approach outperforms other baselines and the HSBi-GRU's effectiveness in learning abstract features. Li et al., 2019 [17] involved a two-layer recurrent neural network stacked in which the lower layer performs the auxiliary of predicting the target boundary and guides the upper layer to predict the unified tags for the primary task of target-oriented sentiment analysis. Meanwhile, they introduced a gate mechanism that modifies the interrelation between the previous and current word features to maintain the opinion target's sentiment consistency. Li et al., 2019 [20] investigated the pre-trained language models, e.g., BERT contextualized embedding modeling power on the E2E-ABSA task. They introduced four alternative

model implementations (e.g., linear layer, RNN layer, self-attention mechanism and conditional random fields) to explore BERT's potential semantic modeling power. In addition, they standardized the comparative study by utilizing a hold-out validate dataset to pick an optimal model, which was always neglected by previous related works. To fully exploit the interactive relations among aspect term extraction, opinion term extraction and aspect-level sentiment classification, Chen et al., 2020 [19] proposed a relation-aware collaborative learning (RACL) framework, which allows the three sub-tasks to work coordinately via the relation induction and multi-task learning mechanisms in a stacked multi-layer network. Experiments for the E2E-ABSA task on three real-world datasets prove that their RACL framework outperforms the unified baselines.

Motivated by the success of disentangled representation learning in the computer vision field, Silva et al. [45] investigated the effectiveness of the powerful disentangled representation learning (DRL) [46], and they utilized the decoding-enhanced BERT with disentangled attention (DeBERTa) [47] to solve the ABSA tasks. Experimental results on the ABSA's benchmarks [31] show that incorporating disentangled attention can yield a promising performance and outperforms many state-of-the-art models.

Existing unified approaches solve the pipelines' error propagation issue and are predominant on the E2E-ABSA task now, but the existing problems, such as the complex search in huge space (integrated labels selection, e.g., B,I,O,E-{POS,NEG,NEU}) and the neglect of modeling the interrelation on the OTE and TSA sub-tasks have become the bottleneck of further improving the accuracy and F1-score of the E2E-ABSA task. Following these questions, we present CasNSA's whole architecture and corresponding technological innovations.

### 3. The CasNSA Framework: Architecture Details

On the whole, our CasNSA framework can be subdivided into three principal sub-modules: contextual semantic representation block (**CSR**), target boundary recognizer (**TBR**) and sentiment polarity identifier (**SPI**). For the CSR module, we try several specific model structure implementations which derive from combinations of two variants of CSR and three variants of **SPI** to pick the optimal model settings. The overall architecture of our **CasNSA** within alternative sub-modules is shown in Figure 1, and we describe its details below.

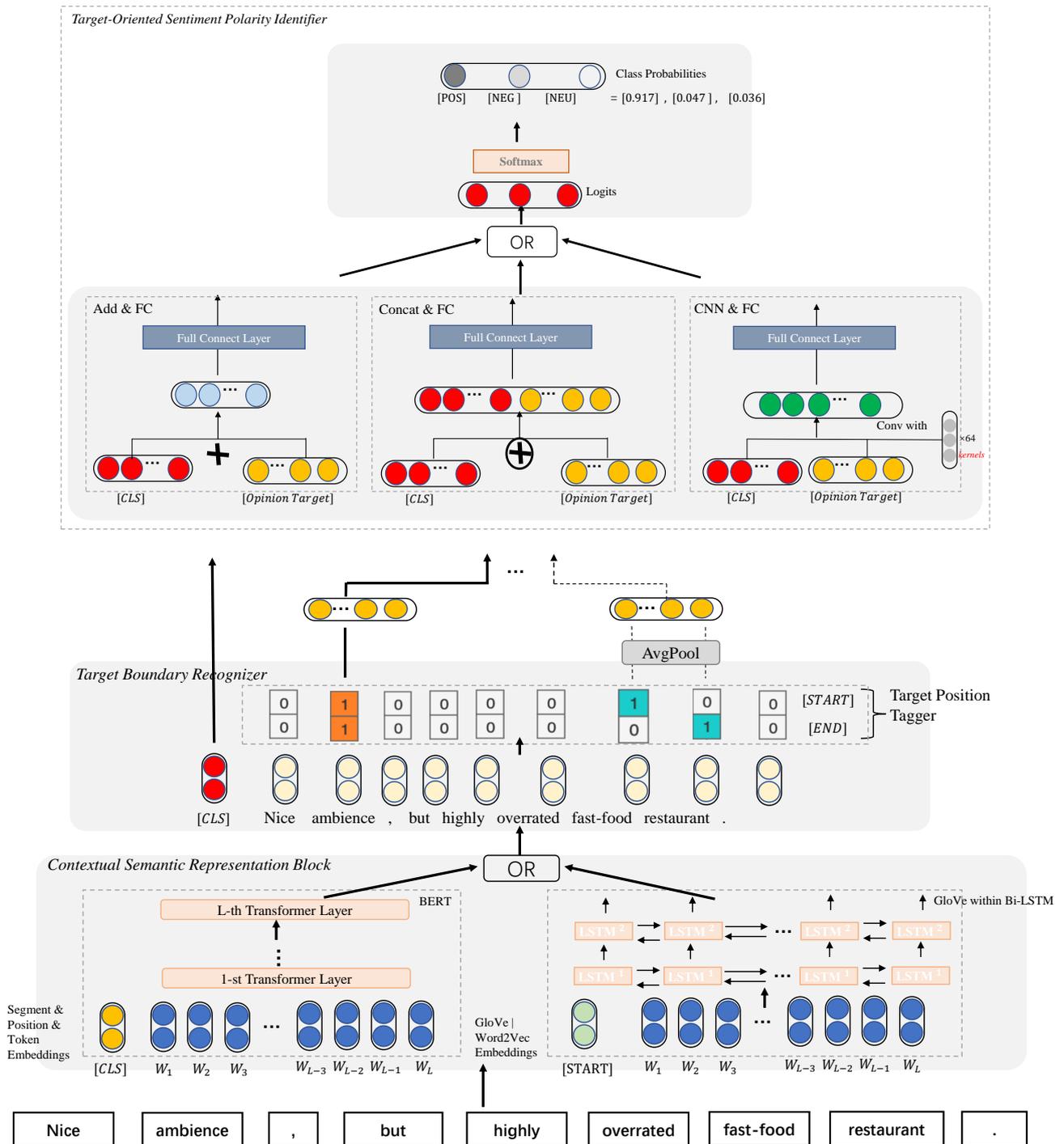
#### 3.1. Task Preliminaries

In this work, given a user comment sentence  $\{w_1, \dots, w_i, \dots, w_L\}$ , where  $L$  means the length of this sentence, end-to-end fine-grained aspect-based sentiment analysis (ABSA) is formulated as sequence tagging problems and multi-label classification problems. Specifically, it handles the sub-tasks: opinion-target extraction (OTE) and target-oriented sentiment analysis (TSA).

**Definition 1.** *Opinion-target extraction (OTE) aims to predict a sequence of target-oriented position tags  $Y^t = \{y_1^t, \dots, y_i^t, \dots, y_L^t\}$ , in which  $y_i^t \in \{B, I, S, O\}$  denoting the beginning of, inside of, single token of, and outside of an opinion target. In particular, this sequence tagging task can be instantiated in many ways, including conditional random fields (CRFs) tagger, binary tagger, etc. In this paper, we employ the simple binary tagger, which adopts two identical binary classifiers to detect the start and end positions of each opinion target, and we describe its concrete implementations in Section 3.4.*

**Definition 2.** *target-oriented sentiment analysis (TSA) aims to conduct the label classification for every specific opinion target. TSA detects its corresponding sentiment polarity for each possible target from three sentiment type, containing positive, negative and neutral. Specifically, given a sentence  $S = \{w_1, \dots, w_i, \dots, w_L\}$  and all opinion target  $T = \{t_1, \dots, t_i, \dots, t_N\}$ , the TSA-relevant component predicts the possible tag  $Y^s = \{y_1^s, \dots, y_i^s, \dots, y_N^s\}$  where  $y_i^s \in \{POS, NEG, NEU\}$  for each  $t_i$ , denoting the positive, negative and neutral sentiments, respectively. For this sub-task,*

we employ three different feature recognition procedures to determine the optimal model setting, and the concrete details are shown in Section 3.5.



**Figure 1.** The architecture overview of CasNSA. The three key components CSR, SPI and TBR is shown from bottom to top.

### 3.2. Explicit Task Modelling

Here, let us deeply explore the interactive relational mappings between the sub-tasks of OTE and TSA. The goal of E2E-ABSA is to identify a set including all possible opinion targets  $T = \{t_1 \dots t_i \dots\}$  and its corresponding sentiment polarities  $S = \{s_1 \dots s_i \dots\}$ . We directly design and formulate the E2E-ABSA overall training objective toward this goal. In

contrast with previous approaches, such as that of Li et al., 2019 [20], where the optimization objective is defined right at the integrated tagging scheme, including ‘B-POS’, ‘I-NEG’ and ‘E-NEU’, our optimization function considers the interrelation hidden in the integrated tags. Then, we can model the E2E-ABSA as the sequence tagging and multi-label classification separately.

Formally, given a user review  $d_i$  from the training dataset  $D$  and a set of potentially opinion-oriented targets and corresponding sentiments  $R_i = \{(t_i, s_i)\}$  in  $d_i$ , we aim to maximize the whole optimized target likelihood of the training dataset  $D$ .

$$Probability\ Score_{(D,R)} = \prod_{i=1}^{|D|} \left[ \prod_{(t_i,s_i) \in R_i} p((t_i, s_i) | d_i) \right] \tag{1}$$

$$= \prod_{i=1}^{|D|} \left[ \prod_{t_i \in R_i} p(t_i | d_i) \prod_{s_i \in R_i | t_i} p(s_i | t_i, d_i) \right] \tag{2}$$

$$= \prod_{i=1}^{|D|} \left[ \prod_{t_i \in R_i} \prod_{j=1}^L (p_j^t)^{\mathbf{I}\{y_j^t=1\}} (1 - p_j^t)^{\mathbf{I}\{y_j^t=0\}} \prod_{s_i \in R_i | t_i} \prod_{k=1}^N (p_k^s)^{\mathbf{I}\{y_k^s=1\}} (1 - p_k^s)^{\mathbf{I}\{y_k^s=0\}} \right] \tag{3}$$

where Equation (1) denotes the overall optimizing goal,  $(t_i, s_i) \in R_i$  denotes all existing targets and their sentiments of the  $i$ -th sentence  $d_i \in D$ , respectively. This formula means the probabilistic optimization for predicting sentimental target entities in all corpora and the corresponding sentiment categories. Equation (2) denotes the inter-scheme tag dependencies modeling between target detection and sentiment polarity identification. We decompose the last joint probabilistic optimization into two step-by-step cascade probabilistic optimizations. Firstly, we extract all sentiment target entities. Secondly, we perform the corresponding sentiment discrimination based on the context semantics of the extracted entities. Equation (3) further decomposes the target optimizing function of the complete E2E-ABSA task. More specifically, we transfer the opinion-target entity extraction task to the binary label classification and transfer the target-oriented sentiment analysis task to the standard multi-label classification task. We utilize the standard cross-entropy loss function [48] to optimize it.

This formulation provides several benefits. At first, in Equation (2), we take the mutual dependencies between the two sub-tasks into consideration, which almost all related researchers often neglect. The sub-tasks would be mutually influenced such that errors in each component can be constrained by the other, and thus it can help model the fine-grained aspect-based sentiment analysis better. Then, the E2E-ABSA task decomposition revealed in Equation (3) decreases the tagging complexity because the substitution of most unified approaches adopts the integrated tagging schemes, such as ‘B-POS’ and ‘I-NEG’. Finally, this sophisticated formulation represents the deep hierarchical neural network and thus can be instantiated in many implementations.

### 3.3. Contextual Semantic Representation Block

In the beginning, given a user comment  $S = \{w_1, w_2, \dots, w_L\}$  for sentiment analysis, we need to interpret the natural language sequence  $S$  into a sequence of semantic feature vectors  $V = \{v_1, v_2, \dots, v_L\}$ . As is shown in the bottom right of Figure 1, following most traditional methods [17,19,27,49], we employ the pre-trained word embeddings (e.g., *GloVe*, and *Word2Vec*) to transfer the sentence tokens into a sequence of the vector. To investigate the language modeling power armed with Transformer, as is shown in the bottom left of Figure 1, besides using pre-trained *GloVe*, we simultaneously introduce an alternative scheme which utilizes the multi-layer bidirectional Transformer-based language model BERT [28] as the sentence encoder.

- **Scheme 1:** As is illustrated in the bottom right of Figure 1, we first use the 300 dimension GloVe [30] to initialize word embeddings. In the model training and reference time,

we add an extra marker '[START]' before the start index of the inputs to generate the sentence-level feature vector. Then the embedding operation queries each word's corresponding embedding and conducts the transfer process  $\{[START], w_1, w_2, \dots, w_L\} \rightarrow \{v_{[START]}, v_1, v_2, \dots, v_L\}$ . To prevent the vanishing-gradient problem [50] existing in RNNs, we choose the two-layer Bi-LSTM as the basic encoder in which the Bi-LSTM hidden size is set to 200. Existing works [12] have demonstrated a better learning capability than the original LSTM. Compared with vanilla recurrent neural network LSTM, bidirectional-LSTM is the same as LSTM in the mechanical aspect, but Bi-LSTM allows the reversed information flow in which the inputs can be fed from the end index to the beginning index. Finally, the encoder layer provides a forward hidden state  $(\overrightarrow{h}_{[START]}, \overrightarrow{h}_1, \overrightarrow{h}_2, \dots, \overrightarrow{h}_L)$  and a backward hidden state  $(\overleftarrow{h}_{[START]}, \overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_L)$ . We list the LSTM feature propagation's relevant formulations as follows:

$$f_j = \sigma(\mathbf{W}_{xf}x_j + \mathbf{W}_{hf}\overrightarrow{h}_{j-1} + \mathbf{b}_f) \tag{4}$$

$$i_j = \sigma(\mathbf{W}_{xi}x_j + \mathbf{W}_{hi}\overrightarrow{h}_{j-1} + \mathbf{b}_i) \tag{5}$$

$$o_j = \sigma(\mathbf{W}_{xo}x_j + \mathbf{W}_{ho}\overrightarrow{h}_{j-1} + \mathbf{b}_o) \tag{6}$$

$$c_j = f_j \circ c_{j-1} + i_j \tanh(\mathbf{W}_{xc}x_j + \mathbf{W}_{hc}\overrightarrow{h}_{j-1} + \mathbf{b}_c) \tag{7}$$

$$\overrightarrow{h}_j = o_j \circ \tanh(c_j) \tag{8}$$

The variables  $f_j, i_j$  and  $o_j$  in the above equations are the input, forget and output gate's activation vectors, respectively. The three gated states  $f_j, i_j$  and  $o_j$  are calculated through a series of complex operations. The updated new memory of LSTM corresponds to the matrix multiplication of the input token feature  $x_j$  and the updating matrix  $W_{xf}, W_{xi}$  and  $W_{xo}$ . The remained old memory of LSTM corresponds to the matrix multiplication of the last hidden state  $\overrightarrow{h}_{j-1}$  and the forgotten matrix  $W_{hf}, W_{hi}$  and  $W_{ho}$ . Finally, LSTM converts the logical value into a prob-value between 0 and 1 through an activation function  $\sigma$ . Furthermore,  $\circ$  is the cell state vector, and  $\sigma$  and  $\tanh$  are the sigmoid and hyperbolic tangent functions.

After the information flows through LSTM, we concatenate the forward  $\overrightarrow{h}_j$  and backward  $\overleftarrow{h}_j$  and obtain the combined features  $H = \{h_{[START]}, h_1 \dots h_j \dots h_L\}$  where the  $j$ -th hidden state  $\mathbf{h}_j = [\overrightarrow{h}_j; \overleftarrow{h}_j]$ , then the obtained hidden state sequence  $H$  is used by the other two downstream tasks, OTE and TBR.

- Scheme 2:** For the sake of the disadvantage that the traditional fixed embedding layer (e.g., Word2Vec, and GloVe) only provides a single context-independent representation, as is illustrated in the bottom left of Figure 1, our CSR module further adopts pre-train Transformer BERT [28] during our experiments. Here, we briefly introduce BERT. Bidirectional encoder representations from Transformers, or BERT, is a revolutionary self-supervised pre-train technique that learns to predict intentionally hidden (masked) sections of text. Crucially, the representations learned by BERT have been shown to generalize well to downstream tasks. When BERT was first released in 2018, it achieved state-of-the-art results on many NLP benchmarks. Specifically, BERT is composed of a stack of  $N$  ( $N \in \{8, 12, 16, \dots\}$ ) identical Transformer blocks. We denote the Transformer block as  $Transformer = \{TransBlock_L\}_1^N$ , in which  $N$  represents the BERT's depth.

Firstly, we pack the sequence of vector inputs  $\{h_{[CLS]}, h_1 \dots, h_L\}$  as  $H_0$ , where the  $h_i$  is the initialized BERT embedding vector of the  $i$ -th token of the sentence. Then the  $L$  - layer transformer blocks refine the token-level semantic representation layer by

layer. Taking the  $j$ -th transformer blocks step as an example, the BERT hidden features  $H_j$  are calculated through Equation (9):

$$H^j = TransBlock_j(H^{j-1}) \tag{9}$$

where the  $H^j = \{h_{[CLS]}^j, h_1^j, \dots, h_L^j\}$  denotes the  $j$ -th BERT feature representations and the  $H^{j-1} = \{h_{[CLS]}^{j-1}, h_1^{j-1}, \dots, h_L^{j-1}\}$  denotes the  $\{j-1\}$ -th BERT feature representations. Finally, we regard  $H^L$  as the contextualized representations of the input sentence, and our CasNSA's other key components (OTE and TBR) use them for the further downstream model-reasoning step.

### 3.4. Target Boundary Recognizer

Similar to the CRF decoding layer, we employ two (start position and end position) binary classifiers with a softmax decoding layer on top of BERT as an opinion-target boundary tagging step. The two classifiers jointly mark each opinion-target's start and end position as "1" and mark the current tags which are irrelevant to target boundaries as "0". In particular, if the current boundary tag denotes the beginning of any opinion target, the "[START]" tagger which aims to detect the start position of this target is tagged with "1", and the "[END]" is tagged with "0"; if the current boundary tag denotes the end of any opinion target, the "[END]" tagger which aims to detect the end position of this target is tagged with "1", and the "[START]" is tagged with "0". The single-word opinion target position is tagged with "1" in both of the classifiers. During the multi-targets detection, we adopt the proximity principle, which regards the phrase between the "[START]" classifier's "1"-tagged position and the corresponding nearest "[END]" classifier's "1"-tagged position. We calculate the probability of whether the character is an opinion-targets boundary by the following formulations.

$$p_i^{start} = \sigma(\mathbf{W}_{start} \mathbf{h}_i + \mathbf{b}_{start}) \tag{10}$$

$$p_i^{end} = \sigma(\mathbf{W}_{end} \mathbf{h}_i + \mathbf{b}_{end}) \tag{11}$$

where the  $\mathbf{h}_i$  represents the  $i$ -th output vector of the *contextual semantic representation* module,  $\mathbf{W}_{(\cdot)}$  and  $\mathbf{b}_{(\cdot)}$  are the learnable matrix weights and bias values of the "[START]" and "[END]" classifiers,  $\sigma_{(\cdot)}$  denotes the activation function, and the whole two formulas of  $p_i^{start}$  and  $p_i^{end}$  denote the encoded representations of the  $i$ -th character in the input sentence. It considers a position to be a boundary and marks the position as "1" when the encoded representations  $p_i^{start}$  and  $p_i^{end}$  exceed a certain threshold (e.g., 0.5 in this paper), otherwise it regards it as a relevant character and marks it as "0".

As shown in Table 2, different from the traditional decoding layers (e.g., CRF), this example concisely illustrates our novel binary tagging strategy. More especially, the token "designs" is the first and also the last word of the opinion-target "designs", so tags are both "1" in the "[START]" and "[END]" classifiers when recognizing these single-word target boundaries.

**Table 2.** This is an example that vividly shows our proposed binary opinion-target boundary tagging strategy.

Sentence	The	Reason	I	Choose	Apple	MacBook	Is	Their	Designs	And	User	Experiences
[START]	0	0	0	0	1	0	0	0	1	0	1	0
[END]	0	0	0	0	0	1	0	0	1	0	0	1

After the *target boundary recognizer* generating the tag sequences  $Y^{[START]} = \{p_i^{start}\}_{i=1}^L$ ,  $Y^{[END]} = \{p_i^{end}\}_{i=1}^L$ , we compute the objective by utilizing the binary cross-entropy loss function [48]:

$$J_{spi}(\Theta) = - \sum_{s \in \mathcal{D}} \sum_{i=1}^L (p_i^{start} \cdot \log(p_i^{start}) + p_i^{end} \cdot \log(p_i^{end})) \tag{12}$$

where  $\mathcal{D}$  and  $L$  denote the training set and the length of one sentence in  $\mathcal{D}$ ,  $p_i^{start}$  and  $\hat{p}_i^{start}$  are the  $i$ -th token's gold tag and the predicted value by the "[START]" binary classifier,  $p_i^{end}$  and  $\hat{p}_i^{end}$  are the  $i$ -th token's gold tag and the predicted value by the "[END]" binary classifier.

### 3.5. Sentiment Polarity Identifier

Similar to the *target boundary recognizer*, the *sentiment polarity identifier* also uses the *contextual semantic representation* output features as its inputs. However, the key features that the target-oriented sentiment identification requires include (1) the depended opinion target; (2) the context that indicates the sentence-level sentiments; (3) the mutual relationships between the opinion-target features, and the contexts. Under these considerations, we propose the target and context joint-aware representation  $\bar{\mathbf{t}}_i$ , given a extracted  $i$ -th opinion target in which its start and end indices of the sentence are  $j$  and  $k$ , and we define  $\bar{\mathbf{t}}_i$  as follows:

$$\bar{\mathbf{t}}_i = [\mathbf{h}_{target}; \mathbf{h}_{context}] = \left[ [\mathbf{h}_j : \mathbf{h}_k]; \mathbf{h}_{[CLS]}^c \right] \tag{13}$$

Formally, we take  $[\mathbf{h}_j : \mathbf{h}_k] = \{\mathbf{h}_j, \mathbf{h}_{j+1}, \dots, \mathbf{h}_k\}$  as  $\mathbf{h}_{target}$ , in which  $\mathbf{h}_j$  and  $\mathbf{h}_k$  denote the start and end position feature representations of the  $i$ -th opinion target. We regard the output vector  $\mathbf{h}_{[CLS]}^c$  which is located on the "[CLS]" position in BERT or "[START]" position in Bi-LSTM as the sentence-level context semantic  $\mathbf{h}_{context}$ . For multiple word opinion target representation  $[\mathbf{h}_j : \mathbf{h}_k]$  where  $j \neq k$ , we employ the mean-pooling operation to incorporate these word features to  $\mathbf{h}_{target}$ .

Then, the assembled features  $\bar{\mathbf{t}}_i$  which contain all relevant signals about the  $i$ -th target for sentiment polarity identifying are sent to the *sentiment polarity identifier* (we mark it as  $f_{SPI}$ ), and then the SPI module predicts the corresponding sentiment polarity labels  $\mathbf{S}_i^{SPI} \in \{[POS], [NEG], [NEU]\}$ .

$$\mathbf{S}_i^{SPI} = f_{SPI}(\bar{\mathbf{t}}_i) \tag{14}$$

The SPI component's feature decoding function  $f_{SPI}$  can be instantiated in many ways. In our experiments, we explore several methods to conduct the feature integration procedure, including (1) simple bitwise adding; (2) simple vector concatenation; and (3) CNN-based feature extraction:

$$f_{SPI}^1([\mathbf{h}_{target}; \mathbf{h}_{context}]) = \mathbf{h}_{target} + \mathbf{h}_{context} \tag{15}$$

$$f_{SPI}^2([\mathbf{h}_{target}; \mathbf{h}_{context}]) = \mathbf{h}_{target} \oplus \mathbf{h}_{context} \tag{16}$$

$$f_{SPI}^3([\mathbf{h}_{target}; \mathbf{h}_{context}]) = CNN(\mathbf{h}_{target} \oplus \mathbf{h}_{context}) \tag{17}$$

These feature incorporation steps are illustrated in the top part of Figure 1. Among them, after the bitwise adding layer and vector concatenate layer, an extra full connection layer is employed. The CNN-based method generates the intermediate feature by running a CNN on the character sequence of  $\bar{\mathbf{t}}_i$ , and the window size of CNN's convolutional kernel for the feature-based vector is set to 3. In the last layer of the whole network, we add a linear layer whose output dimension is set to 3 in this setting, and the target-oriented sentiment polarity can be classified smoothly. For training, we perform a multi-label classification

by adopting cross entropy [48] as the loss function, and the loss of the sub-task sentiment polarity identifying is calculated by the following:

$$J_{spi}(\Theta) = - \sum_{s \in \mathcal{D}} \sum_{t_i \in T_s} \hat{S}_{t_i} \cdot \log(S_{t_i}) \quad (18)$$

where  $\mathcal{D}$  and  $s$  denote all training samples and one training instance of  $\mathcal{D}$ ,  $t_i$  is the  $i$ -th opinion target revealed in  $s$ ,  $S_{t_i}$  is the gold label of target-oriented sentiment polarity, and  $\hat{S}_{t_i}$  is the sentiment prediction result.

It is worth noting that  $\bar{t}_i$  is the gold opinion target at the training time. In contrast, at the inference time, we select the predicted opinion target one by one from TBR module to complete the joint extraction task.

### 3.6. Training Objective of Target and Sentiment Joint Extractor

Following the previous unified work [25,26,51], we combine the two sub-tasks loss function, and the objective  $J_{unify}(\Theta)$  is defined as follows:

$$J_{unify}(\Theta) = J_{tbr}(\Theta) + \lambda \cdot J_{spi}(\Theta) \quad (19)$$

where  $\lambda$  is a coefficient to moderate the mutual contribution weight of these two sub-tasks, and we set it as 1 during our experiments. We train and optimize the model parameters by minimizing  $J_{unify}(\Theta)$  through the Adam stochastic gradient descent [52] over shuffled mini-batches in which each batch contains 16 training samples.

## 4. Experimental Results

### 4.1. Datasets and Evaluation Metrics

**Datasets.** We evaluate our proposed CasNSA on four widely benchmark datasets, including SemEval challenges 2014 [31], 2015 [32], 2016 [33], and the Twitter dataset [34,43]. These benchmark statistics are summarized in Table 3. The SemEval 2014 dataset includes reviews from two domains: restaurant and laptop. We merged them as the total SemEval 2014 dataset. The SemEval 2015 and SemEval 2016 datasets contain thousands of restaurant reviews. The Twitter dataset was built by Mitchell et al. [43], which is a small English dataset that yielded about 3288 unique sentiment pairs, such as <tweet, NEU>.

**Table 3.** This is an example which shows our proposed binary opinion-target boundary tagging strategy.

Dataset	# POS	# NEG	# NEU	Train	Dev	Test	Total
SemEval2014	4226	1986	1442	5477	608	1600	7685
SemEval2015	1241	445	70	1183	685	130	1998
SemEval2016	1712	565	101	1799	200	676	2675
Twitter	698	271	2254	1903	220	234	2357

**Evaluation Metrics.** Following previous works [17,19–21,41], we adopt the standard macro-averaged precision (macro-P), macro-averaged recall (macro-R), and macro-averaged  $F_1$ -score (macro- $F_1$ ) percentages as evaluation metrics. For each token in all evaluated sentences, a token-level sentiment polarity prediction is marked correct if and only if its predicted tag equals its gold tag, otherwise we mark it false.

### 4.2. Parameter Settings

During the contextual representation procedure, we examine our CasNSA with two types of word representations: the pre-trained word embedding and the pre-trained language transformer encoder. Specifically, in the former implementation, the 300-dimension GloVe [30] vectors are employed as the pre-trained word embedding; the hidden state dimension is set to 256. In the latter implementation, we use the “bert-base-uncased” BERT which has 12 transformer blocks and a hidden dimension size of 768 as the pre-trained

encoder. We denote these two implementations as  $CasNSA_{GloVe}$  and  $CasNSA_{BERT}$ . The learning rate is set as  $1 \times 10^5$ . The training batch size is 36 for each iteration. In total, we train the model for up to 100 fixed epochs using the Adam optimizer [52]. After the fifth training epoch, we conduct a model evaluation on the development set per training epoch. We train 10 models with different random seeds and report the average results following these setting. As for the *sentiment polarity identifier* component, which integrates the opinion-target feature with sentence-level contextual representation, the CNN convolutional kernel size is set to 3; the hidden size of the extra full connection layer is set to 200. we initialize all linear and RNN layer parameters by applying the Xavier-Uniform [53] strategy to all weight matrices. All biases are initialized from the uniform distribution  $\mathcal{U}(-0.2, 0.2)$ . Intuitively [17,19], the ratio coefficient  $\lambda$  between the objective  $J_{spi}(\Theta)$  and the objective  $J_{tbr}(\Theta)$  is empirically set as 1.

#### 4.3. Compared Models

To evaluate the effectiveness of CasNSA for the complete E2E-ABSA task, we compare it with several strong state-of-the-art baselines as follows:

- **LSTM-CRF** [54] is a standard sequence tagging framework, which is constructed through the LSTM and CRF decoding layer.
- **BERT-LSTM-CRF** [55] Different from the above LSTM-CRF model, BERT-LSTM-CRF is a competitive model which employs pre-trained language model BERT rather than the pre-trained word embeddings to learn the character-level word representations.
- **E2E-TBSA** [17] E2E-TBSA is a novel framework which involves two stacked LSTMs for performing the OTE and TBR sub-tasks, respectively. Meanwhile, it utilizes a unified tagging scheme to formulate ABSA as a sequence tagging problem.
- **BERT+SAN** [20] This is one of the competitive E2E-ABSA models which stacks a designed self-attention network (SAN) [56] layer on the top of the BERT feature extractor backbone.
- **RACL<sub>GloVe</sub>** [19] RACL is a relation-aware collaborative learning framework with multi-task learning and relation propagation techniques. RACL<sub>GloVe</sub> is one of the two RACL implementations which outperforms many state-of-the-art pipeline baselines and unified baselines for the E2E-ABSA task.
- **ABSA-DeBERTa** [37] ABSA-DeBERTa is a simple downstream fine-tuning model using BERT with disentangled attention for aspect-based sentiment analysis. ABSA-DeBERTa's disentangled attention mechanism incorporates complex dependencies between aspects and sentiments words and thus obtains state-of-the-art results on benchmark datasets.

The above works are all powerful and representative of the fine-grained aspect-based sentiment analysis task. Based on their officially released code, we instantiate their model, reproduce their performance, and report the corresponding performance in the next subsection.

#### 4.4. Overall Comparison Results

Table 4 shows the performance comparisons of our models against other methods for end-to-end textual sentiment analysis. The evaluation metric calculation includes extracting all the exact opinion targets and identifying their corresponding polarities. Our proposed approach, CasNSA and its several variants, overwhelmingly outperforms all other pipelines and achieve state-of-the-art results in terms of the precision/recall/F1-value evaluation metrics on four widely used datasets. The best-performing model of the SemEval2014, SemEval2015 and Twitter benchmarks is the pre-trained BERT-based model *Bert-CasNSA-svc*. CasNSA applies the simple vector concatenate operation as a feature integration procedure, which obtains F1-scores of 68.13%, 56.40% and 50.05% on the SemEval2014, SemEval2015 and Twitter benchmarks, respectively; the best-performing model of the SemEval2015 benchmark is the pre-trained BERT-based model *Bert-CasNSA-cnn* which applies the CNN convolutional operation as feature integration. Thus, *Bert-CasNSA-cnn*

obtains an F1-score of up to 62.34%; more precisely, as presented in Rows 6, 7 and 8, which are in the second group, even without taking advantage of the pre-trained language model BERT, CasNSA<sub>GloVe</sub>, which only utilizes pre-trained word embeddings, is still competitive to the existing state-of-the-art methods. From the results of ABSA-DeBERTa, we can observe that the disentangled attention BERT indicates promising results over the four benchmarks, showing that the detachment of the position and content vectors can help to solve the ABSA tasks. Although our model overall performance may be inferior to ABSA-DeBERTa, our model still has certain advantages compared to other SOTA models due to our novel tagging decomposition. The disentangled representation learning (DRL) [46] enhanced pre-trained language model DeBERTa is indeed a popular and advantageous strategy for handling many downstream language understanding tasks. Motivated by the success of disentangled representation learning, we plan to further investigate the effect of disentangled representation learning on the improvement of model performance in the next stage of research.

**Table 4.** Comparison results of different methods on the development sets for the fine-grained sentiment analysis task. The first group includes the pipeline competitors, the second group contains our CasNSA variants. The models that have marked \* denote that the results are quoted directly from their original published literature rather than our reproduction results. Bold marks the best scores among all models, and the second best scores are underlined.

Models	SemEval2014			SemEval2015			SemEval2016			Twitter		
	Prec	Rec	F1-val									
LSTM-CRF	58.66	51.26	54.71	60.74	49.77	54.71	53.71	50.27	51.91	53.74	42.21	47.26
BERT-LSTM-CRF	53.31	59.40	56.19	59.39	52.94	55.98	57.55	50.39	53.73	43.52	52.01	47.35
E2E-TBSA*	69.83	56.76	61.62	63.60	52.27	57.38	58.96	51.41	54.92	53.08	43.56	48.01
BERT+SAN*	71.32	61.05	65.79	-	-	-	60.75	50.24	55.00	-	-	-
RACL <sub>GloVe</sub> *	<u>77.24</u>	59.06	66.94	<u>69.90</u>	53.03	60.31	59.28	52.56	55.72	52.37	46.75	49.40
ABSA-DeBERTa	<b>81.57</b>	64.74	<b>72.19</b>	<b>92.68</b>	<b>63.24</b>	<b>75.18</b>	<b>94.07</b>	<b>69.27</b>	<b>79.79</b>	<b>71.56</b>	<b>53.69</b>	<b>61.35</b>
GloVe-CasNSA-sba	66.79	61.18	63.86	63.15	55.20	58.91	57.52	50.15	53.58	52.33	45.52	48.69
GloVe-CasNSA-svc	69.21	62.55	65.71	65.28	54.74	59.55	59.56	49.76	54.22	53.17	44.96	48.72
GloVe-CasNSA-cnn	68.47	61.72	64.92	64.74	54.64	59.26	60.08	49.14	54.06	52.60	44.80	48.39
Bert-CasNSA-sba	70.66	62.43	66.29	65.37	57.35	61.10	<u>62.23</u>	51.12	56.13	54.76	45.56	49.74
Bert-CasNSA-svc	71.29	<u>65.24</u>	<u>68.13</u>	67.38	56.95	61.73	62.17	<u>51.61</u>	<u>56.40</u>	<u>56.93</u>	44.65	<u>50.05</u>
Bert-CasNSA-cnn	70.48	<b>65.60</b>	67.95	66.82	<u>58.65</u>	<u>62.34</u>	61.48	51.37	55.97	55.62	45.06	49.79

As is illustrated in Table 4, the comparisons among the second group, which include our CasNSA variants, can be regarded as the ablation analysis of our works. From the second group, we conclude that our CasNSA-based methods can handle the complexity of the fine-grained sentiment analysis task. The results that describe the analysis of our decomposition innovation of the E2E-ABSA task are competitive when compared with other pipelines. As is shown in Rows 1 and 6, our CasNSA variant *GloVe-CasNSA-sba* obtains a much better performance of 63.86% than the standard pipeline *LSTM-CRF* on the SemEval2014 benchmark. Similar phenomena have also existed on other benchmarks. Furthermore, we notice that the CasNSA variants that utilize the simple bit-wise adding operation finally obtain a poor F1 score on the development set. This may be due to the valuable feature loss during the information forward transmission. In short, this comprehensive experiment analysis shows that our proposed method achieves stable and competitive performance. It further demonstrates the effectiveness of our proposed E2E-ABSA task composition strategy in handling the sentiment polarity identification task.

#### 4.5. Case Study

To investigate the superiority of the CasNSA method, in this section we conduct the following case analysis on the results of several examples by several compared methods. We choose RACL<sub>GloVe</sub> (denoted as PIPELINE) as one typical competitor. The analysis

procedure includes the  $\text{CasNSA}_{\text{GloVe}}$  and  $\text{CasNSA}_{\text{BERT}}$  model implementation, as we wish to fully investigate the performance difference between the pre-trained language model and pre-trained word embeddings. In addition, we choose the simple vector to concatenate the operation as the sentiment polarity identifier, as it achieves better performance than other implementations in the above comparison experiments.

As observed in the  $S_1$  and  $S_3$ , the PIPELINE method fails to extract the two opinion targets, while our CasNSAs method both correctly produces the right target boundary of “sweets” and “built-in applications”. It demonstrates that our *target boundary recognizer* features’ decoding capacity is effectiveness and powerful compared to other standard approaches.

$S_2$  shows the benefits of our interactive relational dependencies modeling between the OTE sub-task and TSA sub-task. The PIPELINE method becomes lost in contexts and makes a false sentiment prediction “NEU” for “Sushi”. In contrast,  $\text{CasNSA}_{\text{GloVe}}$  and  $\text{CasNSA}_{\text{BERT}}$  both correctly recognize the sentiment polarity “POS” for “Sushi” with the help of the consideration of opinion-target contextualized semantic features.

Notice that example  $S_3$  fully demonstrates the effectiveness of our work innovations. At first, the PIPELINE method is insufficient for exploiting the contextualized semantic features and fails to recognize the correct target “built-in applications”, especially for those targets with several words inside long-distance sentences. Instead, owing to our proposed novel *target boundary recognizer*,  $\text{CasNSA}_{\text{GloVe}}$  and  $\text{CasNSA}_{\text{BERT}}$  correctly filter out the target “built-in applications”. Meanwhile, although the target information can guide the model to predict the sentiment more accurately, the error is probably inherited due to the lack of contextualized features extracted by pre-trained word embeddings (e.g., GloVe). According to the results given in the last row of Table 5, we can obviously observe that the BERT-based CasNSA implementation successfully identifies the positive sentiment polarity of “iPhoto” while PIPELINE and GloVe-based CasNSA cannot. This observation illustrates the powerful sequential analysis capacity of the pre-trained language model BERT, which helps outperform other methods in many natural language processing tasks. The above cases further validate the effectiveness and superiority of our proposed hierarchical framework CasNSA in handling the fine-grained sentiment analysis task.

**Table 5.** Case analysis. The “OTE” column contains the results from the *Target Boundary Recognizer* (TBR) module for the sub-task of opinion-target extraction (OTE) and the “TSA” column contains the results from the *Sentiment Polarity Identifier* (SPI) module for the sub-task of target-oriented sentiment analysis (TSA). Words in red and italic are annotated aspect/opinion target terms with subscripts denoting their sentiment polarities. The marker (✗) denotes the incorrect prediction and (✓) denotes the correct prediction. We choose the  $\text{RACL}_{\text{GloVe}}$  as the PIPELINE method.

$S_1$	Sentence 1: The [ <i>teas</i> ] <sub>pos</sub> are great and all the [ <i>sweets</i> ] <sub>pos</sub> are homemade.					
$S_2$	Sentence 2: [ <i>Sushi</i> ] <sub>pos</sub> so fresh that it crunches in your mouth.					
$S_3$	The [ <i>performance</i> ] <sub>pos</sub> seems quite good, and [ <i>built-in applications</i> ] <sub>pos</sub> like [ <i>iPhoto</i> ] <sub>pos</sub> work great with my phone and camera.					
Sentence	PIPELINE		$\text{CasNSA}_{\text{GloVe}}$		$\text{CasNSA}_{\text{BERT}}$	
	OTE	TBR	OTE	TBR	OTE	TBR
$S_1$	teas (✓)	POS (✓)	teas (✓)	POS (✓)	teas (✓)	POS (✓)
	None (✗)	None (✗)	sweets (✓)	POS (✓)	sweets (✓)	POS (✓)
$S_2$	Sushi (✓)	NEU (✗)	Sushi (✓)	POS (✓)	Sushi (✓)	POS (✓)
$S_3$	performance (✓)	POS (✓)	performance (✓)	POS (✓)	performance (✓)	POS (✓)
	None (✗)	None (✗)	built-in applications (✓)	POS (✓)	built-in applications (✓)	POS (✓)
	iPhoto (✓)	None (✗)	iPhoto (✓)	None (✗)	iPhoto (✓)	POS (✓)

### 5. Conclusions and Perspectives

Amidst the sentiment polarity distinguished for internet social user comments, we investigate the importance and effectiveness of interactive relations between sub-task opinion target extraction and sub-task target sentiment identification, which is always neglected by most of the related domain researchers. To some extent, many researchers are

discouraged from performing further explorations, due to these two sub-tasks being highly coupled together. Specifically, we propose a novel collaborative learning framework named cascade social sentiment analysis (**CasNSA**) to tackle this critical challenge. Our **CasNSA** model takes advantage of the opinion target contextualized semantic features provided by the opinion target extraction sub-task to guide the sentiment polarity identification for predicting the aspects-oriented sentiment polarities more accurately. Extensive empirical results show that the proposed approach effectively achieves a 68.13% F1-score on SemEval-2014, 62.34% F1-score on SemEval-2015, 56.40% F1-score on SemEval-2016, and 50.05% F1-score on the Twitter dataset, which is higher than the existing approaches.

In particular, the most important novelty of this paper can be summarized in two main respects: (1) to solve the significant problems brought from the discrete modeling between the target sentiment and the opinion target extraction task, our target sentiment distinguish module utilizes the opinion target contextualized semantic features generated by the target extraction module when predicting the sentiment polarity, and thus ensures the sentiment polarity identification quality; (2) we investigate the superiority of the BERT encoding capability by introducing the BERT encoder as the social users' comment contextualized feature representations generator. The model's outstanding performance when using BERT fine-tuning firmly proves that the multi-head self-attention based Transformer is still predominant in capturing aspect-based sentiment and robust to the insufficient sample overfitting dilemma. Meanwhile, different from other pipeline methods, the novel unified framework **CasNSA** is designed to handle the aspect-based social comment sentiment analysis task in an end-to-end fashion. As a result, the **CasNSA** jointly predicts the target boundary position associated with the target-oriented sentiment polarity, thereby effectively tackling the incident error accumulation issue that exists in most pipeline methods.

Moreover, the empirical comparison results illustrate the superiority of our proposed model and the effectiveness of our proposed model's sub-components, such as the hierarchical cascade sequence tagging unit and the BERT encoder. We believe E2E-ABSA will continue to be an attractive and promising research direction with realistic industrial and domestic scenarios, such as intelligent recommendation, smart personal assistant, big data mining services, and automatic customer services.

In the future, we plan to study the following major problems. (i) To support real-world dynamic application scenarios, the social comment sentiment analysis application is always updated quickly and inevitably needs to cover new scenarios in real time. How can we augment our framework's business coverage for handling different scenarios automatically and incrementally? (ii) This framework is built on relatively small datasets under weak supervision without prior external knowledge. How can we introduce external knowledge such as world wide web textual target-entity descriptions and other open-domain knowledge to improve our **CasNSA** framework's performance?

**Author Contributions:** conceptualization, D.H. and H.S.; methodology, formal analysis, and software, D.H.; validation, H.S.; investigation, H.Y. and H.S.; resources and data curation, W.J.; writing—original draft preparation, D.H.; writing—review and editing, D.H., H.S. and W.J.); visualization, S.Y.; supervision, and project administration, H.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by an anonymous scientific research project at Shanghai Dianji University and Shanghai Yangfan Program (22YF1413600).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets and released codes presented in this study can be acquired from the third author's email: weiqiangjin@shu.edu.cn.

**Acknowledgments:** Authors are thankful to anonymous reviewers and editors for their valuable comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
NLP	Natural Language Processing—An AI research area
E2E-ABSA	End-to-End Aspect-Based Sentiment Analysis—a specific NLP research task
BERT	Bidirectional Encoder Representations from Transformers—pre-trained language model
SOTA	State-Of-the-Art—obtain the best performance until now

## References

- Jalil, Z.; Abbasi, A.; Javed, A.R.; Badruddin, K.M.; Abul Hasanat, M.H.; Malik, K.M.; Saudagar, A.K.J. COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques. *Digital Public Health* **2022**, *9*, 812735. [[CrossRef](#)]
- Khan, M.U.; Javed, A.R.; Ihsan, M. A novel category detection of social media reviews in the restaurant industry. *Multimed. Syst.* **2020**. [[CrossRef](#)]
- Liu, J.; Zhang, Y. Attention Modeling for Targeted Sentiment. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 3–7 April 2017; pp. 572–577.
- Zhou, J.; Huang, J.X.; Chen, Q.; Hu, Q.V.; Wang, T.; He, L. Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision, and Challenges. *IEEE Access* **2019**, *7*, 78454–78483. [[CrossRef](#)]
- Liu, B. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol* **2012**, *5*, 1–167. [[CrossRef](#)]
- Jacobs, G.; Véronique, H. Fine-grained implicit sentiment in financial news: uncovering hidden bulls and bears. *Electronics* **2021**, *10*, 2554–2586. [[CrossRef](#)]
- Li, X.; Bing, L.D.; Li, P.J.; Lam, W.; Yang, Z.M. Aspect term extraction with history attention and selective transformation. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018; pp. 4194–4200.
- Jaechoon, J.; Gyeongmin, K.; Kinam, P. Sentiment-target word pair extraction model using statistical analysis of sentence structures. *Electronics* **2021**, *10*, 3187–3211. [[CrossRef](#)]
- Xu, H.; Liu, B.; Shu, L.; Yu, P.S. Double embeddings and cnn-based sequence labelling for aspect extraction. In Proceedings of the The 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 592–598.
- Fan, Z.F.; Wu, Z.; Dai, X.Y.; Huang, S.; Chen, J. Target-oriented opinion words extraction with target-fused neural sequence labelling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 29 April–19 May 2019; pp. 2509–2518.
- Hazarika, D.; Poria, S.; Vij, P.; Krishnamurthy, G.; Cambria, E.; Zimmermann, R. modelling inter-aspect dependencies for aspect-based sentiment analysis. In Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics, New Orleans, LA, USA, 1–6 June 2019; pp. 266–270.
- Ma, Y.; Peng, H.; and Cambria, E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2019; pp. 5876–5883.
- Wang, S.; Mazumder, S.; Liu, B.; Zhou, M.; Chang, Y. Target-sensitive memory networks for aspect sentiment classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 957–967.
- Li, Z.; Wei, Y.; Zhang, Y.; Xiang, Z.; Li, X. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27–31 January 2019; pp. 4253–4260.
- Ma, D.H.; Li, S.J.; Zhang, X.D.; Wang, H.F. Interactive attention networks for aspect-level sentiment classification. In Proceedings of the International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 4068–4074.
- Hu, M.H.; Peng, Y.X.; Huang, Z.; Li, D.S.; Lv, Y.W. Open-domain targeted sentiment analysis via span-based extraction and classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 537–546.
- Li, X.; Bing, L.; Li, P.; Lam, W. A unified model for opinion target extraction and target sentiment prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27–31 January 2019; pp. 6714–6721.
- Li, X.; Bing, L.; Lam, W.; Shi, B. Transformation networks for target-oriented sentiment classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 946–956.
- Chen, Z.; Qian, T. Relation-aware collaborative learning for unified aspect-based sentiment analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3685–3694.
- Li, X.; Bing, L.; Zhang, W.; Lam, W. Exploiting BERT for end-to-end aspect-based sentiment analysis. In Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019), Hong Kong, China, 4 November 2019; pp. 34–41.
- Sun, C.; Huang, L.Y.; Qiu, X.P. utilising BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 29 April–19 May 2019; pp. 380–385.

22. Huang, B.X.; Carley, K.M. Syntax-aware aspect level sentiment classification with graph attention networks. *arXiv* **2019**, arXiv:1909.02606.
23. Zhang, M.; Zhang, Y.; Vo, D.T. Neural networks for open domain targeted sentiment. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 612–621.
24. Ma, D.; Li, S.; Wang, H. Joint learning for targeted sentiment analysis. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 4737–4742.
25. Wei, Z.P.; Su, J.L.; Wang, Y.; Tian, Y.; Chang, Y. A novel cascade binary tagging framework for relational triple extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 1476–1488.
26. Yu, B.W.; Zhang, Z.Y.; Shu, X.B.; Liu, T.W.; Wang, Y.B.; Wang, B.; Li, S.J. Joint extraction of entities and relations based on a novel decomposition strategy. In Proceedings of the European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 29 August–8 September 2020; pp. 2282–2289.
27. Yu J.F.; Jiang, J. Adapting bert for target-oriented multimodal sentiment classification. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 5408–5414.
28. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 29 April–19 May 2019; pp. 4171–4186.
29. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Twenty-seventh Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
30. Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
31. Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androustopoulos, I.; Manandhar, S. SemEval-2014 task 4: aspect based sentiment analysis. *SemEval* **2014**, 27–35. [[CrossRef](#)]
32. Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Manandhar, S.; Androustopoulos, I. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. *SemEval* **2015**, pp. 486–495. Available online: <https://alt.qcri.org/semeval2015/task12/#> (accessed on 1 June 2022).
33. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androustopoulos, I.; Manandhar, S.; AL-Smadi, M.; Zhao, B.Q. SemEval-2016 task 5: Aspect based sentiment analysis. *SemEval* **2016**, 19–30. [[CrossRef](#)]
34. Andres L.F.; Coromoto, C.Y.; Nabhan, H.M. Sentiment analysis in twitter based on knowledge graph and deep learning classification. *Electronics* **2022**, *10*, 2739–2756.
35. Song, Y.W.; Wang, J.H.; Jiang, T.; Liu, Z.Y.; Rao, Y.H. Attentional encoder network for targeted sentiment classification. *arXiv* **2019**, arXiv:1902.09314.
36. Li, X.L.; Li, Z.Y.; Tian, Y.H. Sentimental Knowledge Graph Analysis of the COVID-19 Pandemic Based on the Official Account of Chinese Universities. *Electronics* **2021**, *10*, 2921–2944. [[CrossRef](#)]
37. Yadav, R.K.; Jiao, L.; Granmo, O.-C.; Goodwin, M. Human-Level Interpretable Learning for Aspect-Based Sentiment Analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, 2–9 February 2021; pp. 14203–14212.
38. Granmo, O.-C. The Tsetlin Machine—A Game Theoretic Bandit Driven Approach to Optimal Pattern Recognition with Propositional Logic. *arXiv* **2018**, arXiv:abs/1804.01508.
39. Wang, W.Y.; Pan, Sinno J.; Dahlmeier, D.; Xiao X.K. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2019; pp. 3316–3322.
40. Chen Z.; Qian T.Y. Transfer capsule network for aspect level sentiment classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 547–556.
41. Li H.; Lu W. Learning latent sentiment scopes for entity-level sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2019; pp. 3482–3489.
42. He, R.D.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. 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, Florence, Italy, 28 July–2 August 2019; pp. 504–515.
43. Mitchell, M.; Aguilar, J.; Wilson, T.; Van Durme, B. Open domain targeted sentiment. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1643–1654.
44. Lafferty, J.; McCallum, A.; Pereira, F. Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In Proceedings of the 18th International Conference on Machine Learning, Freiburg, Germany, 12 September 2001; pp. 282–289.
45. Silva, E.H.; Marcacini, R.M. Aspect-based sentiment analysis using BERT with Disentangled Attention. In Proceedings of the 8th ICML Workshop on Automated Machine Learning (AutoML 2021), Virtual Event, 23–24 July 2021.
46. Locatello, F.; Bauer, S.; Lucic, M.; Raetsch, G.; Gelly, S.; Scholkopf, B.; Bachem, O. Challenging common assumptions in the unsupervised learning of disentangled representations. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 4114–4124.
47. He, P.; Liu, X.; Gao, J.; Chen, W. DeBERTa: Decoding-enhanced BERT with Disentangled Attention. *arXiv* **2020**, arXiv:abs/2006.03654.

48. Deng, J. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
49. Catelli, R.; Serena, P.; Massimo, E. Lexicon-Based vs. Bert-Based Sentiment Analysis: A comparative study in Italian. *Electronics* **2022**, *11*, 374–393. [[CrossRef](#)]
50. Roodschild, M.; Sardiñas, J.G.; Will, A. A new approach for the vanishing gradient problem on sigmoid activation. *Prog. Artif. Intell.* **2020**, *9*, 351–360. [[CrossRef](#)]
51. Tao, Q.; Luo, X.; Wang, H.; Xu, R. Enhancing Relation Extraction Using Syntactic Indicators and Sentential Contexts. In Proceedings of the IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), Portland, OR, USA, 4–6 November 2019; pp. 1574–1580.
52. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 515–526.
53. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
54. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. In Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics, San Diego, CA, USA, 12–16 June 2016; pp. 260–270.
55. Liu, L.; Shang, J.; Xu, F.; Ren, X.; Gui, H.; Peng, J.; Han, J. Empower sequence labeling with task-aware neural language model. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018.
56. Shen, T.; Zhou, T.; Long, G.; Jiang, J.; Pan, S.; Zhang, C. Disan: Directional self-attention network for rnn/cnn-free language understanding. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018.