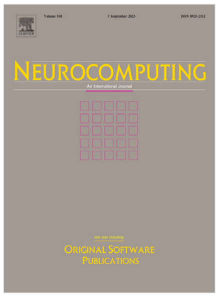
Journal Pre-proof



Syntax-enhanced aspect-based sentiment analysis with multi-layer attention

Jingli Shi, Weihua Li, Quan Bai, Yi Yang, Jianhua Jiang

PII: S0925-2312(23)00853-6
DOI: https://doi.org/10.1016/j.neucom.2023.126730
Reference: NEUCOM 126730

To appear in: *Neurocomputing*

Received date : 15 November 2022
Revised date : 19 July 2023
Accepted date : 20 August 2023

Please cite this article as: J. Shi, W. Li, Q. Bai et al., Syntax-enhanced aspect-based sentiment analysis with multi-layer attention, *Neurocomputing* (2023), doi: https://doi.org/10.1016/j.neucom.2023.126730.

Revised Manuscript (Clean version)

Syntax-Enhanced Aspect-Based Sentiment Analysis with Multi-Layer Attention

Jingli Shi[a], Weihua Li[a], Quan Bai[b], Yi Yang[c], Jianhua Jiang[d]

[a] *Engineering, Computer & Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*
[b] *School of Information & Communication Technology, University of Tasmania, Hobart, Australia*
[c] *School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China*
[d] *Jilin Province Key Laboratory of Fintech, Jilin University of Finance and Economics, Changchun, China*

Abstract

As a key task of fine-grained sentiment analysis, aspect-based sentiment analysis aims to analyse people's opinions at the aspect level from user-generated texts. Various sub-tasks have been defined according to different scenarios, extracting aspect terms, opinion terms, and the corresponding sentiment. However, most existing studies merely focus on a specific sub-task or a subset of sub-tasks, having many complicated models designed and developed. This hinders the practical applications of aspect-based sentiment analysis. Therefore, some unified frameworks are proposed to handle all the subtasks, but most of them suffer from two limitations. First, the syntactic features are neglected, but such features have been proven effective for aspect-based sentiment analysis. Second, very few efficient mechanisms are developed to leverage important syntactic features, e.g., dependency relations, dependency relation types, and part-of-speech tags. To address these challenges, in this paper, we propose a novel unified framework to handle all defined sub-tasks for aspect-based sentiment analysis. Specifically, based on the graph convolutional network, a multi-layer semantic model is designed to capture the semantic relations between aspect and opinion terms. Moreover, a multi-

Email addresses: `jingli.shi@aut.ac.nz` (Jingli Shi), `weihua.li@aut.ac.nz` (Weihua Li), `quan.bai@utas.edu.au` (Quan Bai), `yyang@hfut.edu.cn` (Yi Yang), `jjh@jlufe.edu.cn` (Jianhua Jiang)

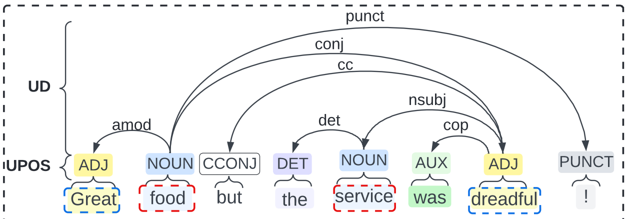
layer syntax model is proposed to learn explicit dependency relations from different layers. To facilitate the sub-tasks, the learned semantic features are propagated to the syntax model with better semantic guidance to learn the syntactic representations comprehensively. Different from the conventional syntactic model, the proposed framework introduces two attention mechanisms. One is to model dependency relation and type, and the other is to encode part-of-speech tags for detecting aspect and opinion term boundaries. Extensive experiments are conducted to evaluate the proposed novel unified framework, and the experimental results on four groups of real-world datasets explicitly demonstrate the superiority of the proposed framework over a range of baselines.

Keywords: Aspect-based Sentiment Analysis, Graph Convolutional Network, Aspect Sentiment Triplet, Syntactic Structure

1. Introduction

Sentiment Analysis (SA) aims to analyse people's attitudes, opinions, and sentiment distributions toward certain products, services, or opinions [1]. As an important problem in SA, Aspect-Based Sentiment Analysis (ABSA) focuses on finding fine-grained sentiments on different aspects of an item. ABSA involves various sub-tasks, including Aspect Term Extraction (ATE), Opinion Term Extraction (OTE), Aspect-Level Sentiment Classification (ALSC), Aspect-oriented Opinion Extraction (AOE), Aspect Extraction and Sentiment Classification (AESC), Aspect-Opinion Pair Extraction (AOPE), and Aspect Sentiment Triplet Extraction (ASTE) [2, 3, 4]. In all these sub-tasks, there are three key elements: **Aspect Term**, referring to the word or phrase in a user review of a product or service, **Opinion Term**, which is the word or phrase, expressing customer's attitudes on the opinion target, and **Sentiment Polarity**, which is identified as positive, negative, or neutral towards the target in a sentence. For example, given a customer review "*Great food but the service was dreadful !*", the aspect terms include "*food*" and "*service*", the corresponding opinion terms are "*great*" and "*dreadful*", and sentiment polarities are "*positive*" and "*negative*". Figure 1 shows the outputs of different sub-tasks of SA for this sample review.

Although ABSA has received increasing attention in academia and industry, the existing works merely focus on a single task or a few of the sub-tasks



Sub-Tasks	Output
AE	{food, service}
OE	{great, dreadful}
ALSC	{positive$_{food}$, negative$_{service}$}
AOE	{great$_{food}$, dreadful$_{service}$}
AESC	{(food, positive), (service, negative)}
AOPE	{(food, great), (service, dreadful)}
ASTE	{(food, great, positive), (service, dreadful, negative)}

Figure 1: A customer review with its dependency tree and the outputs for different sub-tasks. UD means universal dependency, and UPOS refers to the universal part-of-speech.

of SA [3]. In the early works, ATE [5, 6, 7, 8] and OTE [9, 10, 11] have been recognised as two common tasks in ABSA. Recently, many studies have been dedicated on other sub-tasks, such as AOE [12, 13, 14, 15, 16], ALSC [17, 18, 19, 20], and AESC [21, 22, 23]. Due to the limitation of annotated datasets, AOPE [24, 25, 26, 27, 28] and ASTE [26, 29, 3, 30, 31, 32, 33, 34, 35] turn out to be rather challenging. Moreover, solving all defined sub-tasks using a unified framework is more important because such a unified framework can be easily applied to practical applications in the real world [3]. A few recent works explored the unified model [29, 33, 3]. However, the important syntactic features are neglected, e.g., dependency relations, dependency relation types, and part-of-speech tags, which can improve the performance of ABSA. Therefore, the challenge of designing a unified model remains for ABSA.

Deep learning models demonstrated impressive results and achieved outstanding performances in addressing the ABSA sub-tasks. To encode the semantics of user reviews, neural sequential models, such as LSTM [36], GRU [37], Transformer [38], and ELMo [39], have been extensively applied to ABSA [40, 41, 27, 26, 42, 32, 43, 44]. Recently, pre-trained language models, e.g., BERT [45], RoBERTa [46] and BART [47], achieved superior performances in various ABSA subtasks without explicit consideration of syntactic information [48, 3, 33, 34, 49]. Unfortunately, semantic information learned by the sequential semantic models is not sufficient to solve the problems of ABSA since there are many syntactic relationships between aspect terms, opinion terms, and the corresponding sentiment polarities, as shown in Figure 1. To alleviate this problem, the hierarchical tree models, e.g., TreeLSTM [50] and Graph Convolutional Network (GCN) [51], have been introduced to capture syntactic information enriching the semantic models and facilitating the ABSA tasks. Moreover, recent studies have shown that the external syntactic structure knowledge can bring further strengths to sentiment analysis modelling [5, 35, 32].

Although these works have achieved state-of-the-art performances, several challenges still exist in addressing the problems of ABSA in a unified generative formulation. **First**, most existing works usually study a specific sub-task alone in ABSA, which causes various complicated ABSA models to be designed and hinders the practical usage of the designed models. Instead of developing multiple models, an integrated model has proved to be more effective in solving each task of ABSA in a unified manner [3]. **Second**, syntactic features are neglected in current unified frameworks of aspect-based sentiment analysis. Most existing works only focus on learning semantic representation while neglecting the syntactic features that can essentially promote the extraction of aspect and opinion terms, and improve the performance for ABSA sub-tasks [29, 33, 3]. Although these unified frameworks are based on pre-trained language models that can capture implicit syntactic information from a sentence, limitations still exist due to the absence of explicit syntactical features that can enhance the specific ABSA tasks. **Third**, in most current studies, the linguistic part-of-speech (POS) and syntactic dependency label are two overlooked syntactic features that can have a positive impact on model performance [52, 16, 32, 53, 54, 55]. For example, in Figure 1, the POS tag is *NOUN* for the aspect terms *food* and *service*. Such syntactic features can point out the boundary between neighbours of phrases and further benefit the detection of aspect and opinion terms. After identifying the aspect term *food* and opinion term *great*, the corresponding sentiment polarity can be easily predicted as *positive* through the dependency relation $\{food \curvearrowleft great|amod\}$, where *amod* refers to the adjectival modifier.

In this paper, to tackle the challenges mentioned above, we propose a novel neural network model to integrate explicit syntactic information with semantic features for all ABSA sub-tasks. Specifically, the pre-trained language model is first utilised to generate semantic embeddings, while syntactic embeddings are obtained through the syntax embedding layer. We further design a **M**ultiple **S**yntactic **S**tructure (MSS) fusion encoder with semantic and syntax graph networks, which play different roles in the proposed framework. The semantic graph network is able to learn the representation via adjacency neighbourhood of context, and syntax graph network is applied to encode syntactic structure information through structural connections. Semantic and syntax embeddings are also constructed into graph-structured data, in which the directly connected nodes are fed into a GCN to encode local information by a convolution operation. To learn global information, the state of each node in the graph passes through the multilayer semantic and syntax graph networks in MSS. Furthermore, the hidden features from the MSS multi-layers are concatenated with the POS-aware attention to provide enhanced features. Finally, the output of concatenated features is forwarded to the feed-forward neural network and Softmax layer for tasks of extraction and classification.

To sum up, our contributions are listed as follows:

- Firstly, we propose a novel neural network architecture that can handle all defined ABSA sub-tasks. Instead of developing different models for different sub-tasks, the proposed model converts the sub-tasks into question-answering tasks and tackles them using a unified framework. Different from the existing unified models, the proposed model integrates the syntactic information with semantic features to form reinforced representations for predicting aspect and opinion terms and classifying the corresponding sentiment.

- Secondly, to incorporate explicit syntax information, a **M**ultiple **S**yntactic **S**tructure (MSS) fusion encoder is proposed, leveraging syntax information to enrich the semantic features of user review.

- Thirdly, extensive experiments are conducted on four groups of real-world datasets. The empirical results show that the proposed framework can produce outstanding performances for all the ABSA sub-tasks.

The remainder of this paper is organised as follows. Related works are reviewed in Section 2. In Section 3, relevant concepts are formally defined, and the problem is formulated. After briefly describing the proposed model in Section 4, the experiments and analysis of experimental results are presented in Section 5. Finally, our contributions and future works are summarised in Section 6.

2. Related Work

In this section, the existing studies of ABSA single task and multiple sub-tasks are reviewed. Then, the early studies are introduced to describe the application of syntax-based models for ABSA. Finally, recent works based on attention mechanisms are presented with the pre-trained language models.

2.1. Sub-tasks of Aspect-based Sentiment Analysis

Aspect-based sentiment analysis (ABSA) has been widely studied in the recent decade due to the increasingly popular e-commerce [56, 57]. Liu et al. propose an unsupervised framework based on lifelong learning to improve opinion aspect extraction [5]. Xu et al. first employ a dual embedding mechanism, namely general-purpose and domain-specific embeddings, to a novel yet simple Convolutional Neural Network (CNN) model for aspect extraction [6]. Based on an Inward-Outward LSTM, a target-fused neural network model is proposed to perform target-oriented opinion words extraction [24]. A transferable network is introduced for fine-grained opinion extraction in [11]. The proposed network can exploit local and global memory interactions to capture correlations among aspect or opinion words. Li et al. propose a novel model for target-oriented sentiment classification, where a CNN layer is employed to extract salient features from the transformed word representations originating from a bi-directional Recurrent Neural Network (RNN) layer [17]. In [52], a GCN layer is designed over the dependency tree of a sentence for aspect-specific sentiment classification. To address severe domain shift between pretraining and downstream ABSA datasets, Liu et al. introduce a pretraining framework with instance-level and knowledge-level alignments [58]. To fuse sentiment knowledge and inter-aspect feature, Han et al. construct dependency tree with sentiment knowledge information of words, and attention mechanism to model inter-aspect interaction for sentiment classification [59]. Their model can maintain more domain-invariant knowledge on target dataset for aspect-based sentiment classification. Although these models can achieve remarkable performance compared to the previous works, only a specific ABSA task can be solved using each model. This makes applying such models to practical applications challenging due to the inconvenience in real-world scenarios. For example, the manufacturers prefer to keep track of customer opinions of their products by understanding more than one sentiment element from reviews, which requires conducting multiple ABSA sub-tasks.

To address the above limitations, some recent works focus on the compound ABSA tasks that aim to jointly extract multiple sentiment elements or identify the elements in pair or triplet. For example, Xu et al. propose a two-stage method, where syntactic patterns are incorporated in a sentiment graph to detect aspect or opinion words in the first stage, and the extraction results are refined by a self-learning strategy in the second stage [60]. Similarly, Li et al. develop an LSTM-based deep multi-task learning framework that handles aspect and opinion extraction jointly [61]. Hu et al. propose a span-based extract-then-classify framework that applies the supervision of aspect span boundaries to identify multiple aspect terms and then classify the corresponding sentiment by aspect span representations [23]. Aspect-opinion pair extraction is first introduced in [24], and the task is solved by leveraging an LSTM-based sequence labelling model. Chen et al. utilise BERT to learn context representations and construct a synchronous double-channel recurrent network for aspect-opinion pair extraction [27]. To identify all elements of ABSA in one shot, aspect sentiment triplet extraction is introduced, where a two-stage framework based on the LSTM and GCN networks is proposed [29]. Chen et al. adopt a multi-turn machine reading comprehension framework using BERT to extract the triplet [34]. These approaches can significantly improve the performance compared to the previous studies, but solving all the ABSA sub-tasks in a unified way still turns out to be a major challenge. A few recent studies attempt to address this issue. For example, Mao et al. handle all sub-tasks by constructing two machine reading comprehension problems with BERT as the backbone network [33]. Yan et al. propose a unified framework to solve all sub-tasks, where the pre-trained language model BART is exploited to construct a sequence-to-sequence network [3]. Nevertheless, these existing unified frameworks neglect the syntactical information and word dependencies of a sentence. Such features have been proven as essential information for the tasks of ABSA [49].

2.2. Syntax-based Aspect Level Sentiment Analysis

In customer reviews, explicit syntactic relations are usually presented between opinion words and the corresponding aspect words [5]. To enhance the performance of aspect and opinion extraction, the Graph Convolutional Network (GCN) is widely applied to exploit word dependencies and syntactical structure for ABSA sub-tasks. Zhang et al. build a GCN over the dependency tree of a sentence to explore the relevant syntactical information and long-range word dependencies for aspect-based sentiment classification [52].

Sun et al. present a convolution over the dependency tree model based on Bi-directional Long Short Term Memory (Bi-LSTM) and GCN to learn contextual and dependency features for identifying the sentiment polarity [62]. Veyseh et al. incorporate the syntactic structures of the sentences into LSTM and GCN models for targeted opinion word extraction [13]. To enhance the task of aspect-opinion pair extraction, a label-ware GCN is introduced for modelling rich syntactic knowledge in [63]. By leveraging dependency relations and types, a type-ware GCN is developed to address ABSA sub-tasks [53]. To fully utilise the relations between words for aspect sentiment triplet extraction, Chen et al. propose a multi-channel GCN model by considering linguistic features [35]. Liu et al. incorporate GCN into the gating mechanism to enhance GCN to learn aspect-related affective features [64]. Note that most existing GCN-based models only focus on the syntactic dependency edge without considering dependency labels and POS. Such features can provide distinct evidence of word relations and boundaries. Moreover, the performance improvement appears limited. This is because such models fail to distinguish the significant syntactic relations, and the user-generated reviews are often informal and complex.

2.3. Attention-based Models for Sentiment Analysis

Recently, attention mechanisms have been applied to a wide range of deep learning models for ABSA sub-tasks due to their inherent capability in the semantic alignment of aspect and context words [52]. To achieve aspect and opinion terms co-extraction, Wang et al. propose a coupled multi-layer attention model that consists of a couple of attentions in each layer, i.e., one for extracting aspect terms and the other for opinion terms extraction [65]. Zhang et al. present a novel retrieval-based attention mechanism to retrieve important semantic features for aspect extraction [52]. Wu et al. employ an attention layer to enhance the connection between aspect and opinion words for aspect-oriented opinion extraction [26]. Chen et al. propose a supervised self-attention mechanism to extract opinion entities and relations for aspect-opinion pair extraction [27]. To achieve the task of target-oriented opinion word extraction, Jiang et al. design a novel attention-based GCN to exploit syntactic information over dependency graphs [16], and Feng et al. apply the same idea to aspect-opinion pair extraction by developing a target-specified sequence labelling with multi-head self-attention model [15]. Aiming to solve the problem of unavailable dependency tree, Chen et al. design a discrete latent opinion tree by building connection between aspect-to-context attention scores and syntactic distances [66]. Additionally, the sentiment classification results can be easily interpreted via the discrete structure. Zhong et al. propose a knowledge graph augmented network to incorporate external knowledge with syntactic and contextual information [67]. The method can capture sentiment feature representations from context, syntax and knowledge-based perspectives with an attention mechanism. However, the existing models either only focus on the score calculation of semantic attention between aspect/opinion and context words or fail to consider the guidance of semantic representation on multiple syntactic features in the attention mechanism.

To alleviate the issues of the existing models, in this paper, a novel unified framework is proposed to achieve all the ABSA sub-tasks by using a complete end-to-end solution. Besides dependency relations, POS and dependency type are applied to leverage the enhanced syntactic structures for ABSA sub-tasks. Furthermore, two attention mechanisms are introduced to capture the important word relations and the target boundaries. Inspired by the research work [55], in this paper, two GCNs are designed to exploit semantic and syntactic features. The indirect word relations and key syntax relations are comprehensively learned through multiple layers in GCNs, which can benefit the extraction of aspect and opinion terms since some opinion terms are connected with the corresponding aspect terms through syntax rules.

3. Preliminaries

In this section, the formal definitions related to ABSA sub-tasks are presented, and then the problem is formally formulated based on these definitions.

3.1. Formal Definition

A review sentence is represented as a word sequence, i.e., $R = \{w_1, w_2, ..., w_n\}$, including a set of aspect terms $A = \{a_1, a_2, ..., a_l\}$, opinion terms $O = \{o_1, o_2, ..., o_m\}$, and the corresponding sentiment polarities $S = \{s_1, s_2, ..., s_p\}$. n refers to the number of words in a sentence, l and m mean the number of aspect terms and opinion terms in R, respectively. p indicates the number of sentiment polarities. Note that a_i and o_i denote a word or a span over several words in R. The ABSA sub-tasks can be defined as below.

Definition 1: ATE aims to extract all aspect terms $\{a_i|a_i \in A\}$ from a review R.

Definition 2: OTE describes the extraction of opinion terms $\{o_j|o_j \in O\}$ from a review R.

Definition 3: ALSC refers to predicting the sentiment of a given aspect target a_i as $\{POS, NEG, NEU\}$, where the three elements denote positive, negative, and neutral, respectively.

Definition 4: AOE aims to extract the corresponding opinion terms $\{o_i|o_i \in O\}$ of the given aspect terms $\{a_i|a_i \in A\}$ from a review R.

Definition 5: AESC are two sub-tasks solved in sequence, where the aspect terms $\{a_i|a_i \in A\}$ are extracted, and the sentiment $\{s_i|s_i \in S\}$ of each a_i in a review R is predicted.

Definition 6: AOP detects the aspect-opinion pairs $\{(a_i, o_j)|a_i \in A, o_i \in O\}$ in the review R.

Definition 7: ASTE denotes the extraction of all triplets $\{(a_i, o_j, s_k)|a_i \in A, o_i \in O, s_k \in S\}$ from a review R.

3.2. Problem Formulation

In this paper, we aim to address ABSA sub-tasks in a unified framework. Specifically, the proposed model can identify the three fundamental elements of ABSA, namely, aspect term, opinion term, and sentiment polarity, for either a single task or many compound tasks in a unified solution. Thus, the targeted output of the ABSA sub-tasks can be formulated below:

- Single task without oriented targets:
  $Y_{ATE/OTE} = \{w_i^{si}, w_i^{ei}\}_{i=1}^{|Y|}$
- Single task with oriented targets:
  $Y_{ALSC} = \{s_i|a_i\}_{i=1}^{|Y|}$ or
  $Y_{AOE} = \{w_i^{si}, w_i^{ei}|a_i\}_{i=1}^{|Y|}$
- Pair task:
  $Y_{AESC} = \{(w_i^{si}, w_i^{ei}, s_i)\}_{i=1}^{|Y|}$ or
  $Y_{AOP} = \{(\dot{w}_i^{si}, \dot{w}_i^{ei}, \ddot{w}_i^{si}, \ddot{w}_i^{ei})\}_{i=1}^{|Y|}$
- Triplet task:
  $Y_{AOP} = \{(\dot{w}_i^{si}, \dot{w}_i^{ei}, \ddot{w}_i^{si}, \ddot{w}_i^{ei}, s_i)\}_{i=1}^{|Y|}$,
  where the superscript si and ei indicate the start and end index of aspect or opinion terms. $\dot{w}$ and $\ddot{w}$ denote the first and second element in the pair or triplet. s_i refers to the corresponding sentiment polarity.

4. Unified Syntax-Enhanced Network

The proposed framework involves five layers: input layer, embedding layer, MSS layer, decoder layer, and output layer. Figure 2 shows the overview architecture of the proposed network.

4.1. Input Layer

Given a review sentence, the question-answer pairs for the sub-tasks of ABSA are generated by a QA generator, which is motivated by recent work [34]. For example, for the review "*Great food but the service was dreadful !*", the generated QA pairs are {*Q: what aspects are discussed? A: food, service*}, {*Q: what opinions are expressed? A: great, dreadful*}, {*Q: what sentiment polarity is for aspect food/service? A: POS/NEG*}, {*Q: what opinion is for aspect food/service? A: great/dreadful*}, {*Q: what aspect is for opinion great/dreadful? A: food/service*}, and {*Q: what the sentiment polarity is for aspect food/service and opinion great/dreadful? A: POS/NEG*}.

4.2. Embedding Layer

The embedding layer consists of two components, i.e., context embedding and syntax embedding. The context embedding converts a sequence of words into the embedding vector by the post-trained domain BERT models [68]. Specifically, the pre-trained BERT model is less task awareness and domain awareness [69, 68]. Thus, the domain BERT model is applied in the proposed framework. Given a sentence $\{w_1, w_2, ..., w_n\}$, semantic embedding $E^s = \{e_i^s \in \mathbb{R}^{d^s}\}_{i=1}^n$ can be obtained using Equation (1), where d^s denotes the dimension of the semantic embedding space. Syntax embedding encodes syntax features into vector embeddings, including dependency relation, dependency type, and POS. For words, w_i and w_j, in a sentence, the dependency graph can be constructed as $(w_i, w_j, t_{i,j})$, where $t_{i,j}$ refers to the dependency type between word w_i and w_j. The dependency relations can be converted into the corresponding adjacency matrix $J = \{adj_{i,j}\}_{n \times n}$, where $adj_{i,j} = 1$ if there is an edge between w_i and w_j, and $adj_{i,j} = 0$ otherwise. To consider the connection of a node to itself, a self-loop adjacency matrix $\hat{J} = \{\hat{adj}_{i,j}\}_{n \times n}$ is calculated using Equation (2). A dependency matrix $T = \{t_{i,j}\}_{n \times n}$ is utilised to record the dependency types, and the embedding of dependency type can be presented as $E^d = \{e_i^d \in \mathbb{R}^{d^d}\}_{i=1}^n$ with d^d denoting the dimension of the dependency type embedding space. Let $C = \{c_i\}_i^n$ denote the POS categories of a list of POS tuples (w_i, c_i) in a sentence, where c_i is the POS category of

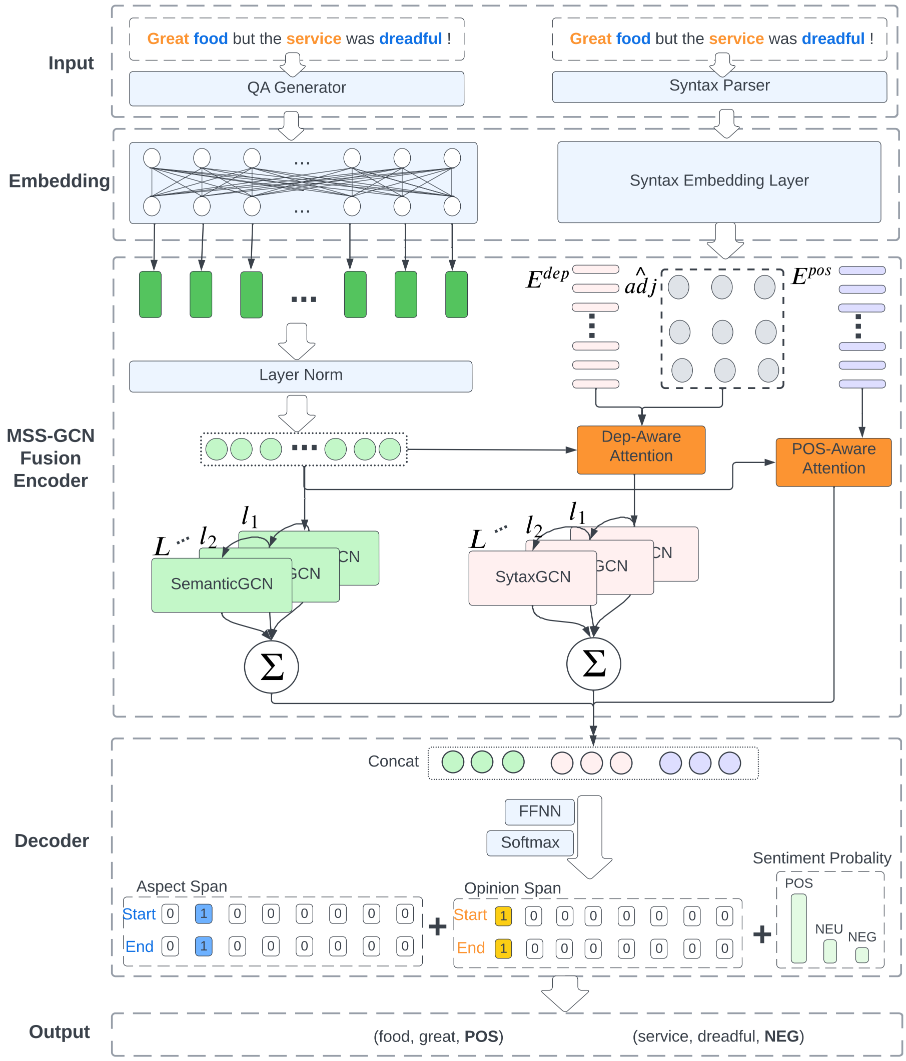


Figure 2: The overview architecture of the proposed network. This shows an example of a triplet where the aspect and opinion terms are *food*, and *great*, respectively. The corresponding sentiment polarity is *positive*.

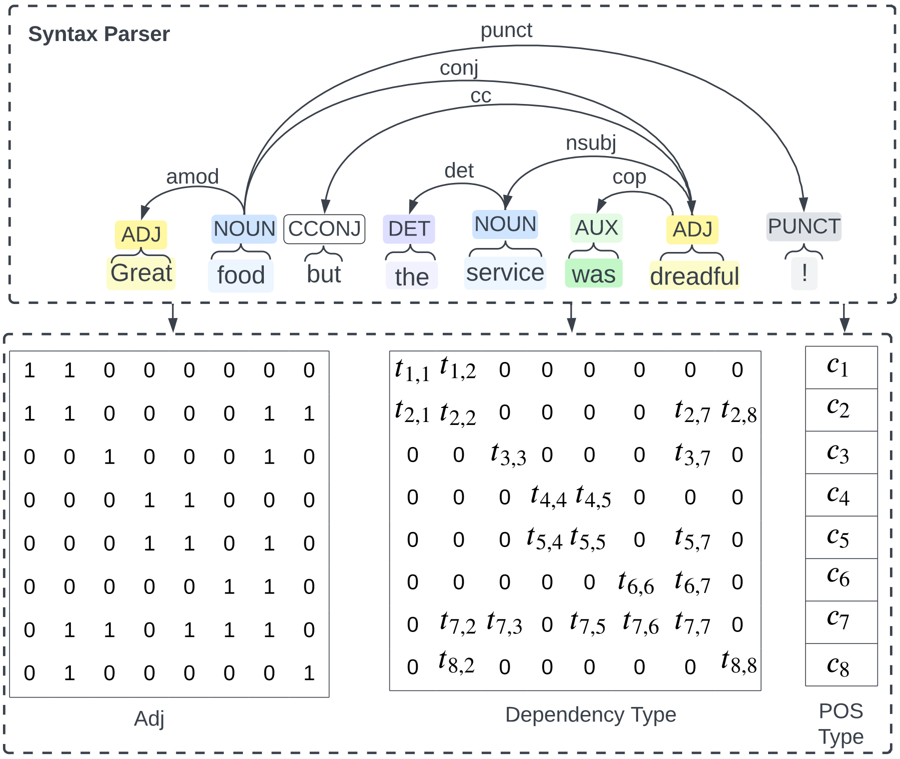


Figure 3: The syntax input generator.

word w_i. Therefore, the embedding of POS categories can be initialised as $E^p = \{e_i^p \in \mathbb{R}^{d^p}\}_{i=1}^n$ with d^p representing the dimension of the POS category embedding space. Figure 3 illustrates the example of matrix generation for dependency relation, dependency type, and POS category.

$$E^s(\{e_1^s, e_2^s, ..., e_n^s\}) = DomBERT(\{w_1, w_2, ..., w_n\}) \quad (1)$$

$$\hat{J} = J + I, \quad (2)$$

where I describes $n \,\mathrm{x}\, n$ identity matrix.

4.3. MSS Layer

To exploit the syntactic structure knowledge, a **M**ultiple **S**yntactic **S**tructure (MSS) fusion network is designed in the encoder layer to bring further strengths to semantic features. To prevent gradient vanishing and exploring, in Equation (3), layer normalisation is applied to the contextual representations from domain BERT [70].

$$\hat{E}^s(\{\hat{e}^s_1, \hat{e}^s_2, ..., \hat{e}^s_n\}) = LayerNorm(E^s)(\{e^s_1, e^s_2, ..., e^s_n\}) \tag{3}$$

where $LayerNorm$ is layer normalisation.

Then, a semantic attention mechanism is introduced to capture semantic correlations between the target and context, which significantly benefits the semantic features with more informative representations. The attention $\alpha^s_{i,j}$ is computed by Equations (4) - (6).

$$\dot{e}^s_i = \dot{W}\hat{e}^s_i + \dot{b} \tag{4}$$

$$\ddot{e}^s_j = \ddot{W}\hat{e}^s_j + \ddot{b} \tag{5}$$

$$\alpha^s_{i,j} = \frac{exp(\dot{e}^s_i \ddot{e}^s_j)}{\sum_{j=1}^{n} exp(\dot{e}^s_i \ddot{e}^s_j)}, \tag{6}$$

where $\dot{W}$ and $\ddot{W}$ denote trainable weight matrixes. $\dot{b}$ and $\ddot{b}$ refer to the bias terms.

The semantic attention matrix is used as an adjacency matrix and fed into the multiple-layer semantic GCN module, which can be formulated in Equation (7).

$$h^{s(l^s+1)}_i = \sigma(\sum_{j=1}^{n} \alpha^s_{i,j}(W^{s(l^s+1)} h^{s(l^s)}_i + b^{s(l^s+1)})), \tag{7}$$

where l^s means the number of semantic GCN layers. $W^{s(l^s+1)}$ and $b^{s(l^s+1)}$ are the trainable parameter in the (l^s+1)-th GCN layer. σ indicates the activation function, i.e., $ReLU$.

Similar to semantic GCN, dependency-aware (dep-aware) attention is designed to incorporate dependency knowledge and semantic features in the

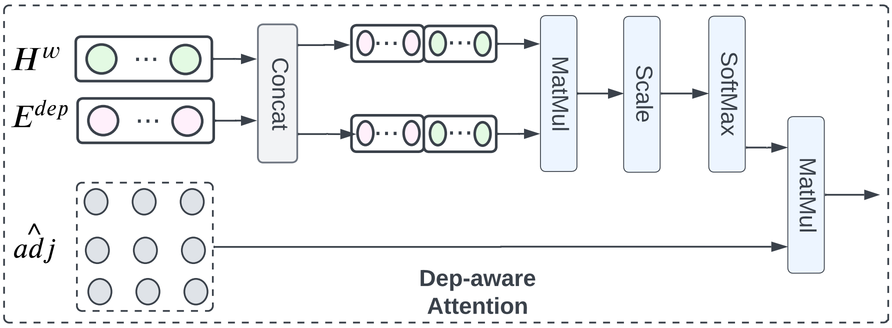


Figure 4: The illustration of Dep-aware attention.

multi-layer syntax GCN. In detail, for each l^d+1-th syntax GCN layer, the hidden representation of w_i is expressed in Equation (8).

$$h_i^{d(l^d+1)} = \sigma(\sum_{j=1}^{n} \alpha_{i,j}^{d(l^d+1)} (W_1^{d(l^d+1)} h_i^{d(l^d)} + W_2^{d(l^d+1)} e_i^{d(l^d+1)} + W_3^{d(l^d+1)} \hat{e}_i^{s(l^d+1)} + b^{d(l^d+1)})), \tag{8}$$

where $W_1^{d(l^d+1)}$, $W_2^{d(l^d+1)}$, and $W_3^{d(l^d+1)}$ are trainable parameters. $b^{d(l^d+1)}$ refers to the bias term. $alpha_{i,j}^{d(l^d+1)}$ is the dep-aware attention score between w_i and w_j, which can be obtained by Equations (9) - (11). The details of dep-aware attention are shown in Figure 4.

$$\hat{e}_i^{d(l^d+1)} = W^{d(i)}[h_i^{d(l^d)}; e_i^{d(l^d+1)}; \hat{e}_i^{s(l^d+1)}] + b^{d(i)} \tag{9}$$

$$\hat{e}_j^{d(l^d+1)} = W^{d(j)}[h_j^{d(l^d)}; e_j^{d(l^d+1)}; \hat{e}_j^{s(l^d+1)}] + b^{d(j)} \tag{10}$$

$$\alpha_{i,j}^{d(l^d+1)} = \frac{\hat{adj}_{i,j} exp(\hat{e}_i^{d(l^d+1)} \hat{e}_j^{d(l^d+1)})}{\sum_{j=1}^{n} \hat{adj}_{i,j} exp(\hat{e}_i^{d(l^d+1)} \hat{e}_j^{d(l^d+1)})} \tag{11}$$

Finally, to explicitly determine the boundary of aspect and opinion terms, the multi-layer POS-aware attention mechanism is designed to only focus on a small window of the local context surrounding the target word instead of considering all the words. This mechanism is able to avoid the expensive

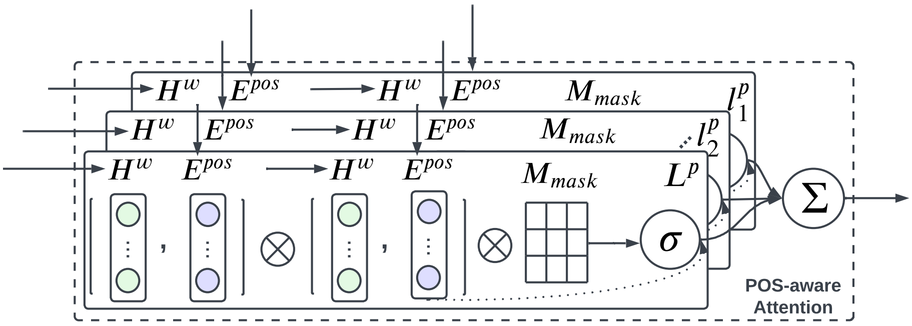


Figure 5: The illustration of POS-aware attention.

computation in attention. Specifically, in each $l^p + 1$-th layer, the POS-aware attention score is formulated in Equation (12) by integrating semantic embedding and POS embedding. Figure 5 presents the process of POS-aware attention.

$$\alpha_{i,m}^{p(l^p+1)} = \frac{exp(W^p[h_m^{p(l^p)}; \hat{e}_m^s])}{\sum_{k=i-D}^{i+D} exp(W^p[h_k^{p(l^p)}; \hat{e}_k^s])} \tag{12}$$

$$h_i^{p(l^p+1)} = \sum_{m=i-D}^{i+D} \alpha_{i,m}^{p(l^p+1)} h_i^{p(l^p)} \tag{13}$$

where D denotes the window size and W^p indicates the learnable weight matrix. l^p refers to the number of attention layers, and $h_i^{p(0)}$ is the POS embedding.

4.4. Decoder Layer

Since both semantic and syntax GCNs are multiple-layer structures and each layer has a unique capability to encode semantic or syntactic information, the GCNs can learn indirect relations between words from long distances in different layers. To utilise such information, the output of semantic and syntax GCN from each layer is ensemble through a weighted average as in Equations (14) - (15). The same operation is applied to the output of POS attention in Equation (16).

$$\hat{h}_i^s = \sum_{l^s=1}^{L^s} \theta^{s(l^s)} h_i^{s(l^s)} \tag{14}$$

$$\hat{h}_i^d = \sum_{l^d=1}^{L^d} \theta^{d(l^d)} h_i^{d(l^d)} \tag{15}$$

$$\hat{h}_i^p = \sum_{l^p=1}^{L^p} \theta^{p(l^p)} h_i^{d(l^p)}, \tag{16}$$

where L^s, L^d, and L^p present the number of layers for semantic GCN, syntax GCN, and POS-aware attention, respectively. $\theta^{(\cdot)}$ indicates the trade-off parameter.

Three categories of tasks are pre-defined as $\{c|c \in (start - index, end - index, sentiment - polarity)\}$. To determine the prediction of each category, the hidden states from GCNs and POS-aware attention are concatenated as input of the fully connected feed-forward network (FFN). Therefore, a $softmax$ function is employed to predict the output of each category by Equation (17).

$$p(y_i^c) = softmax(W^c[\hat{h}_i^s; \hat{h}_i^d; \hat{h}_i^p] + b^c), \tag{17}$$

where W^c and b^c are trainable matrix and bias, respectively.

4.5. Output Layer

After fixed epochs of training, the predicted results for each task are generated. For a single task, the model only predicts the target's start index and end index, or the sentiment polarity. For pair extraction tasks, the output is a tuple of data, e.g., (a_i, o_i) and (a_i, s_i). The output is a set of triplets, i.e., (a_i, o_j, s_k), for aspect sentiment triplet extraction. To present the prediction details, Algorithm 1 shows the process of triplet prediction.

4.6. Model Objective Function

In this subsection, the final objective function is described for model training. To train the proposed network, a joint loss is defined using the cross-entropy loss function by Equation (18).

$$\mathcal{L} = \sum_{c \in C} \theta^c \mathcal{L}^c, \tag{18}$$

where $C = \{startindex, endindex, sentimentpolarity\}$ presents the task category. θ^c means the regularisation coefficients to balance the learning process

Algorithm 1 The prediction of aspect sentiment triplet extraction.

Output: $\hat{Y} = \{(\hat{a}_1, \hat{o}_1, \hat{s}_1), ..., (\hat{a}_{|Y|}, \hat{o}_{|Y|}, \hat{s}_{|Y|})\}$
Input: $R = \{w_1, w_2, ..., w_n\}$
Initialize $\hat{Y} = \{\}$
$\hat{A}(\hat{a}_1, ..., \hat{a}_{|\hat{A}|}) = DecoderLayer(R)$
while $\hat{a}_i < \hat{A}$ **do**
    $\hat{O}(\hat{o}_1, ..., \hat{o}_{|\hat{O}|}) = DecoderLayer(\hat{A})$
    while $\hat{o}_j < \hat{O}$ **do**
        $\hat{S}(\hat{s}_1, ..., \hat{s}_{|\hat{S}|}) = DecoderLayer(\hat{A}, \hat{O})$
        while $\hat{s}_k < \hat{S}$ **do**
            $\hat{Y} = \hat{Y} \bigcup \{(\hat{a}_i, \hat{o}_j, \hat{s}_k)\}$
        end while
    end while
end while
return $\hat{Y}$

between different tasks. For each task category, the negative log-likelihood loss is formulated in Equations (19) - (21).

$$\mathcal{L}^{si} = -\sum_{i=1}^{n} p(y_i^{si}) log(\hat{p}(y_i^{si})) \tag{19}$$

$$\mathcal{L}^{ei} = -\sum_{i=1}^{n} p(y_i^{ei}) log(\hat{p}(y_i^{ei})) \tag{20}$$

$$\mathcal{L}^{sp} = -\sum_{t=1}^{T} p(y_t^{sp}) log(\hat{p}(y_t^{sp})), \tag{21}$$

where si, ei, and sp mean start index, end index, and sentiment polarity, respectively. T refers to the number of targets. $\hat{p}(\cdot)$ denotes the predicted distribution and $p(\cdot)$ indicates the annotated gold distribution.

5. Experiments

In this section, extensive experiments are conducted to evaluate the proposed network using real-world datasets. The experimental results demonstrate the comparisons between the proposed framework and the state-of-the-art models. On top of that, the performance improvement is comprehensively analysed in-depth.

5.1. Dataset

The experiments are conducted on three groups of benchmark datasets for aspect-based sentiment analysis, denoted as $\mathbb{D}_1$, $\mathbb{D}_2$, and $\mathbb{D}_3$. The detailed statistics are shown in Tables 1 - 3. All benchmark datasets originate from the Semantic Evaluation (SemEval) workshops [2, 71, 72] that consist of customer reviews on two domains, i.e., *Laptop* and *Restaurant*. However, only aspect terms and the corresponding sentiment polarities are annotated in all datasets, which hinders them from being applied to other sub-tasks of ABSA. To achieve the sub-task OTE, opinion terms are manually annotated in $\mathbb{D}_1$ [73, 65]. In $\mathbb{D}_2$, the triplet, including aspect term, opinion term, and sentiment polarity, is labelled to address sub-task ASTE [29]. As a revised variant dataset of $\mathbb{D}_2$, some missing triplets are corrected in $\mathbb{D}_3$ [30]. For each group dataset, the ratio of training, validating and testing dataset are shown in Tables 1 - 3. Besides, we further list the ratio of each sentiment polarity to all polarities.

Table 1: The statistics of datasets $\mathbb{D}_1$. The notation #S, #A, #O, #S+, #S-, #S0, #Sc, and #T denote the number of sentences, aspect terms, opinion terms, positive sentiment, negative sentiment, neutral sentiment, conflict sentiment, and (A,O,S) triplet, respectively.

Dataset	Lap14			Res14			Res15		
	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test
#S	2439 63%	609 16%	800 21%	2436 63%	608 16%	800 21%	1052 53%	263 13%	685 34%
#A	2412 63%	584 15%	824 22%	3370 62%	810 15%	1225 23%	967 56%	235 13%	542 31%
#O	2308 63%	576 15%	804 22%	3090 62%	779 15%	1130 23%	1032 55%	261 14%	581 31%
#S+	818 (61%,43%)	176 (13%,38%)	341 (26%,52%)	1744 (61%,59%)	416 (14%,58%)	726 (25%,65%)	731 (60%,76%)	171 (14%,73%)	319 (26%,59%)
#S-	690 (69%,36%)	180 (18%,39%)	128 (13%,20%)	643 (64%,22%)	161 (16%,23%)	195 (20%,17%)	193 (45%,20%)	59 (14%,25%)	179 (41%,33%)
#S0	369 (58%,19%)	94 (15%,21%)	169 (27%,26%)	520 (63%,17%)	117 (14%,16%)	195 (23%,17%)	31 (51%,3%)	3 (5%,1%)	27 (44%,5%)
#Sc	37 (61%,2%)	8 (13%,2%)	16 (26%,2%)	73 (70%,2%)	18 (17%,3%)	14 (13%,1%)	10 (36%,1%)	1 (3%,1%)	17 (61%,3%)

Table 2: The statistics of datasets $\mathbb{D}_2$. Notations #S, #A, #O, #S+, #S-, #S0, and #T denote the number of sentences, aspect terms, opinion terms, positive sentiment, negative sentiment, neutral sentiment, and (A,O,S) triplet, respectively.

Datasets	Lap14							Res14						
	#S	#A	#O	#S+	#S-	#S0	#T	#S	#A	#O	#S+	#S-	#S0	#T
Train	920	1276	1112	692	456	128	1265	1300	2059	1831	1487	408	164	2145
	62%	62%	62%	(58%,54%)	(69%,36%)	(62%,10%)	60%	61%	61%	61%	(60%,72%)	(61%,20%)	(64%,8%)	61%
Dev	228	316	279	184	110	22	337	323	523	463	366	119	38	524
	15%	15%	15%	(15%,58%)	(16%,35%)	(11%,7%)	16%	15%	15%	15%	(15%,70%)	(18%,23%)	(15%,7%)	15%
Test	339	472	416	317	98	57	490	496	828	723	635	139	54	862
	23%	23%	23%	(27%,67%)	(15%,21%)	(27%,12%)	24%	24%	24%	24%	(25%,77%)	(21%,17%)	(21%,6%)	24%
	Res15							Res16						
	#S	#A	#O	#S+	#S-	#S0	#T	#S	#A	#O	#S+	#S-	#S0	#T
Train	593	831	782	634	173	24	923	842	1181	1107	872	267	42	1289
	56%	56%	56%	(58%,76%)	(51%,21%)	(47%,3%)	57%	61%	62%	62%	(61%,74%)	(65%,23%)	(57%,3%)	62%
Dev	148	225	205	173	44	8	238	210	291	274	207	75	9	316
	14%	15%	15%	(16%,77%)	(13%,20%)	(16%,3%)	15%	15%	15%	15%	(15%,71%)	(18%,26%)	(12%,3%)	15%
Test	318	425	425	283	123	19	455	320	442	405	347	72	23	465
	30%	29%	29%	(26%,67%)	(36%,29%)	(37%,4%)	28%	24%	23%	23%	(24%,79%)	(17%,16%)	(31%,5%)	23%

Table 3: The statistics of datasets $\mathbb{D}_3$. Notations #S, #A, #O, #S+, #S-, #S0, and #T denote the number of sentences, aspect terms, opinion terms, positive sentiment, negative sentiment, neutral sentiment, and (A,O,S) triplet, respectively.

Datasets	Lap14							Res14						
	#S	#A	#O	#S+	#S-	#S0	#T	#S	#A	#O	#S+	#S-	#S0	#T
Train	906	1280	1264	817	517	126	1460	1266	2051	2071	1692	480	166	2338
	62%	63%	62%	(61%,56%)	(67%,35%)	(56%,9%)	62%	61%	60%	61%	(59%,72%)	(64%,21%)	(58%,7%)	60%
Dev	219	295	304	169	141	36	346	310	500	498	404	119	54	577
	15%	14%	15%	(12%,49%)	(18%,41%)	(16%,10%)	15%	15%	15%	14%	(14%,70%)	(16%,21%)	(19%,9%)	15%
Test	328	463	473	364	116	63	543	492	848	850	773	155	66	994
	23%	23%	23%	(27%,67%)	(15%,21%)	(28%,12%)	23%	24%	25%	25%	(27%,78%)	(20%,15%)	(23%,7%)	25%
	Res15							Res16						
	#S	#A	#O	#S+	#S-	#S0	#T	#S	#A	#O	#S+	#S-	#S0	#T
Train	605	862	941	783	205	25	1013	857	1198	1307	1015	329	50	1394
	56%	57%	58%	(61%,77%)	(51%, 20%)	(41%,3%)	58%	62%	62%	62%	(61%,73%)	(68%,24%)	(56%,3%)	62%
Dev	148	213	236	185	53	11	249	210	296	319	252	76	11	339
	14%	14%	14%	(14%,74%)	(13%,21%)	(18%,5%)	14%	15%	15%	15%	(15%,74%)	(16%,23%)	(12%,3%)	15%
Test	322	432	461	317	143	25	485	326	452	474	407	78	29	514
	30%	29%	28%	(25%,65%)	(36%,30%)	(41%,5%)	28%	23%	23%	23%	(24%,79%)	(16%,15%)	(32%,6%)	23%

5.1.1. Evaluation Metrics

In this paper, three standard evaluation metrics, i.e., precision (P), recall (R), and F1 score (F1), are adopted to evaluate our model. Specifically, it is necessary to correctly predict all elements in (a, o) and (a, o, s) for sub-task AOP and ASTE, respectively. For AOP and ASTE, the number of predicted pairs and triplets is compared to the actual number in the given dataset.

5.1.2. Implementation and Hyper-parameters

The Pytorch framework [1] is utilised to implement the proposed network. The syntax structures of all sentences are obtained by using Stanford

[1]https://pytorch.org/

NLP Toolkit [2](i.e., Stanza [74]). Two domain BERT models, i.e., BERT-PT_laptop [3] and BERT-PT_rest [4], are applied to generate semantic embeddings of customer reviews. In the ablation study, the pre-trained BERT-base [5] is further utilised to obtain word representations of datasets. All experiments are conducted on a single NVIDIA RTX A6000 GPU accelerator.

The default settings are used for domain BERTs, e.g., 12 layers of self-attention with 768 dimensional hidden vectors. The dimensionalities of both dependency type embedding d^d and POS embedding d^p are set to 200. The Adam optimiser [75] is applied with an initial learning rate of 1e-3. The epoch is set to 40, and the batch size is 10.

5.2. Baselines

The proposed approach is evaluated by comparing it against the following baselines. Most of these baseline models only focus on a single or subset of sub-tasks in a pipeline or joint manner, and very few state-of-the-art models handle all sub-tasks in a unified way. Table 4 presents the core module, selected datasets, and solved ABSA sub-tasks for each baseline.

- ***DP***[76] is a semi-supervised method based on bootstrapping, which addresses two problems, i.e., opinion lexicon expansion and opinion target extraction. The syntactic relations linking opinion words and targets are identified using a dependency parser, and then they are applied to expand the initial opinion lexicon and extract opinion targets.

- ***NCRF-AE***[77] is an end-to-end neural auto-encoder model for sequential structured prediction problems. The model consists of an encoder, a CRF [78] model enhanced by deep neural networks, and a decoder, a generative model to reconstruct the input.

- ***LSTM-RNN***[79] applies recurrent neural networks (RNNs) and word embeddings to fine-grained opinion mining tasks without any task-specific feature engineering effort. After acquiring pre-trained word embeddings, the word vectors are fine-tuned by the proposed RNN model to learn task-specific embeddings. The performance of the proposed

[2]https://stanfordnlp.github.io/stanza/
[3]https://huggingface.co/activebus/BERT-PT_laptop
[4]https://huggingface.co/activebus/BERT-PT_rest
[5]https://github.com/huggingface/pytorch-pretrained-BERT

model can be improved even further by incorporating some linguistic features, e.g., POS and phrasal information, into RNNs.

- ***RNCRF***[73] jointly identifies aspect and opinion terms through integrating recursive neural networks and CRF into a unified framework. Except for learning the high-level discriminative features, RNCRF is able to double propagate information between aspect and opinion terms, simultaneously. Moreover, the extraction performance can be further boosted by incorporating hand-crafted features into the proposed model.

- ***OTE-MTL***[31] presents a novel view of ABSA as an opinion triplet extraction task and proposes a multi-task learning framework. The proposed method can jointly detect aspects, opinions, and sentiment dependencies with two independent heads and a sentiment dependency parser head in the specific multi-head architecture.

- ***Li-Unified+***[22] aims to convert target-based sentiment analysis into a complete task and proposes a novel unified model with a unified tagging scheme. The proposed method consists of two recurrent neural networks: the upper model produces the final results of target-based sentiment analysis by predicting the unified tags; the lower model guides the upper model through performing an auxiliary target boundary prediction.

- ***RINANTE+***[10] automatically mines aspect and opinion term extraction rules based on dependency parsing outputs. Next, these mined rules are applied to annotate auxiliary data. Finally, a neural model is trained to learn from both automatically labelled and human-annotated data to extract aspect and opinion terms.

- ***TS***[29] first introduces aspect sentiment triplet extraction, which is recognised as a new sub-task in ABSA. To address this task, a two-stage framework is proposed with a complete solution in one shot. The aspect, opinion, and corresponding sentiment are predicted in the first stage, and the second stage pairs up all predicted results to form the final triplets.

- ***CMLA+***[65] provides an end-to-end solution to achieve the task of aspect and opinion terms co-extraction. The proposed multi-layer attention network consists of two attentions in each layer. One is for aspect terms extraction, while the other is for extracting opinion terms.

- ***SPAN-BERT***[23] is a span-based extract-then-classify framework that extracts multiple opinion targets under the supervision of target span boundaries and classifies the corresponding sentiment polarities using the extracted span representations from the sentence.

- ***DomBERT***[80] is a domain-oriented language model, which combines the words of general-purpose language models and domain-specific language understanding for aspect-based sentiment analysis. The DomBERT can learn from both relevant domain corpora and in-domain corpus, which benefits the domain language model learning with low-resources.

- ***CD-E2EABSA***[81] is an end-to-end framework for cross-domain aspect-based sentiment analysis. The framework can achieve domain adaptation by capturing domain-invariant features and domain-dependent features with a multi-task learning strategy. Their promising experiment results prove that their cross-domain method can perform better than most of the in-domain models.

- ***SentiPrompt***[82] is a prompt-based method for aspect-based sentiment analysis, which injects sentiment knowledge of aspect, opinion, and polarity into prompt. Moreover, the term relations is applied to the model via constructing consistency and polarity judgment templates.

- ***EMC-GCN***[35] aims to fully apply the relations between words for aspect sentiment triplet extraction. Ten types of relations combining with words are embedded as an adjacent tensor, which are treated as nodes and edges to formulate a multi-channel graph.

- ***SPAN-ASTE***[83] explicitly considers the interaction between the whole spans of aspects and opinions, predicting the corresponding sentiment relation for aspect sentiment triplet extraction. The proposed span-level model can address the extraction limitation of aspect and opinion terms with multiple words since it captures the whole span semantics of aspect and opinion terms.

- ***IMN-BERT***[84] jointly learns multiple related tasks simultaneously at the token and document level for aspect-based sentiment analysis.

The multi-task network can fully exploit joint information from aspect extraction and sentiment prediction.

- ***RACL-BERT***[85] fully exploits the interactive relations among aspect term extraction, opinion term extraction, and aspect-level sentiment classification. Moreover, it allows the three subtasks to work coordinately via multi-learning and relation propagation mechanisms for the complete ABSA task.

- ***JET-BERT***[30] is the first end-to-end model for extracting aspect sentiment triplets. In the proposed method, a position-aware tagging scheme is introduced to specify the structural information of a triplet and capture interactions among elements in the triplet. Such a scheme contributes to triplet extraction.

- ***DMRC***[33] solves all sub-tasks of ABSA in a unified end-to-end framework by joint training two BERT machine reading comprehension (MRC) models with parameters sharing.

- ***BMRC***[34] transforms ASTE into a multi-turn machine reading comprehension problem, and comprehensively identifies triplets by a bidirectional Machine Reading Comprehenhsion (MRC) structure.

- ***BART-ABSA***[3] converts all ABSA sub-tasks into a unified generative formulation, where the pre-trained model BART [47] is utilised to solve sub-tasks in an end-to-end framework.

- ***SynGen***[86] is a plug-and-play syntactic features aware module, which is easily applied to aspect-based sentiment analysis framework backbones.

5.3. Experimental Results and Model Analysis

In this subsection, we present extensive experimental results using three groups of datasets, i.e., $\mathbb{D}_1$, $\mathbb{D}_2$, and $\mathbb{D}_3$. Then, we analyse the strengths and weaknesses of the proposed method by comparing it against the state-of-the-art models mentioned previously. To ensure the fairness and rationality of the experiments, the same datasets are used by the proposed model and all the baseline models. The experimental results are shown in Tables 5 - 7.

Table 4: The summarisation of baselines in the experiments. Y means that the baseline can handle the sub-task and N indicates that the sub-task is not able to be solved by the baseline.

Baselines	Core Module	Datasets	**ATE**	**OTE**	**ALSC**	**AOE**	**AESC**	**AOP**	**ASTE**
DP	Bootstrapping	$\mathbb{D}_1$	Y	Y	N	N	N	N	N
NCRF-AE	AutoEncoder	$\mathbb{D}_1$	Y	Y	N	N	N	N	N
LSTM-RNN	LSTM	$\mathbb{D}_1$	Y	Y	N	N	N	N	N
Li-Unified+	LSTM	$\mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	N	Y	Y	Y
OTE-MTL	Multi-task	$\mathbb{D}_2$	Y	Y	N	N	N	N	Y
RNCRF	RNN+CRF	$\mathbb{D}_1$	Y	Y	N	N	N	N	N
RINANTE+	LSTM+CRF	$\mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	N	Y	Y	Y
TS	LSTM+GCN	$\mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	N	Y	Y	Y
CMLA	Attention	$\mathbb{D}_1$	Y	Y	N	N	N	N	N
CMLA+	Attention	$\mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	N	Y	Y	Y
SPAN-BERT	BERT	$\mathbb{D}_1$	Y	N	Y	N	Y	N	N
DomBERT	BERT	$\mathbb{D}_1$	Y	N	Y	N	Y	N	N
SPAN-ASTE	Bi-LSTM+BERT	$\mathbb{D}_3$	Y	Y	Y	N	N	N	Y
IMN-BERT	BERT+CNN	$\mathbb{D}_1$	Y	Y	Y	N	Y	N	N
CD-E2EABSA	BERT+Bi-LSTM	$\mathbb{D}_1$	N	N	N	N	Y	N	N
RACL-BERT	BERT	$\mathbb{D}_1$	Y	Y	Y	N	Y	N	N
EMC-GCN	BERT	$\mathbb{D}_1, \mathbb{D}_2$	N	N	N	N	N	N	Y
JET-BERT	BERT	$\mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	N	Y	Y	Y
DMRC	BERT	$\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$	Y	N	Y	Y	Y	Y	Y
BMRC	BERT	$\mathbb{D}_2$	Y	Y	Y	Y	Y	Y	Y
SentiPrompt	BART	$\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$	N	N	N	N	Y	Y	Y
BART-ABSA	BART	$\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	Y	Y	Y	Y
Ours	BERT	$\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	Y	Y	Y	Y

Table 5: The comparison results in F1 on dataset $\mathbb{D}_1$. ATE, OTE, ALSC, and AESC refer to aspect term extraction, opinion term extraction, aspect-level sentiment classification, and aspect extraction and sentiment classification, respectively.

Model	Res14				Lap14				Res15			
	ATE	OTE	ALSC	AESC	ATE	OTE	ALSC	AESC	ATE	OTE	ALSC	AESC
DP	38.72	65.94	-	-	19.19	55.29	-	-	27.32	46.31	-	-
NCRF-AE	83.28	85.23	-	-	74.32	75.44	-	-	65.33	70.16	-	-
LSTM-RNN	81.15	80.22	-	-	72.73	74.98	-	-	64.30	66.43	-	-
RNCRF	84.05	80.93	-	-	76.83	76.76	-	-	67.06	66.90	-	-
CMLA	85.29	83.18	-	-	77.80	80.17	-	-	70.73	73.68	-	-
CD-E2EABSA	-	-	-	60.18	-	-	-	62.26	-	-	-	60.18
DomBERT	77.21	-	-	-	**83.89**	-	-	-	-	-	-	-
SentiPrompt	-	-	-	**81.09**	-	-	-	**70.79**	-	-	-	64.50
SPAN-BERT	86.71	-	71.75	73.68	82.34	-	62.5	61.25	74.63	-	50.28	62.29
IMN-BERT	84.06	85.10	75.67	70.72	77.55	81.0	75.56	61.73	69.90	73.29	70.10	60.22
RACL-BERT	86.38	87.18	81.61	75.42	81.79	79.72	73.91	63.40	73.99	76.00	74.91	66.05
DMRC	86.60	-	82.04	75.95	82.51	-	75.97	65.94	75.08	-	73.59	65.08
BART-ABSA	**87.07**	**87.29**	75.56	73.56	83.52	77.86	76.76	67.37	75.48	76.49	73.91	66.61
Ours	86.65	87.01	**82.42**	76.41	82.63	**83.21**	**77.67**	68.42	**76.34**	**78.93**	**78.51**	**67.28**

Firstly, Table 5 presents the comparison results of several models for ABSA sub-tasks of ATE, OTE, ALSC, and AESC on dataset $\mathbb{D}_1$. Only parts of ABSA sub-tasks are performed because the aspect-opinion pair and sentiment triplet are not annotated. Similar to BERT-based models, domain BERT is adopted in our model. Thus, the performance is close to the models based on BERT, e.g., DomBERT, SPAN-BERT, RACL-BERT, and DMRC. Among these BERT-based models, DomBERT obtains the best performance for ATE sub-task on dataset *Lap14*, which supports the idea domain knowledge benefits the ABSA tasks. As another unified framework for ABSA, BART-ABSA obtains competitive results on *Res14* and *Lap14* for ATE and OTE sub-tasks, which highlights the potential of using a pre-trained language model for ABSA tasks. SentiPrompt, which is a prompt-tuning model, can achieve the best results on the AESC sub-task on the two datasets *Res14* and *Lap14*. It proves that the prompt-tuning model can outperform fine-tuning model by leveraging the sentiment knowledge in prompts. However, unlike models that can only tackle one sub-task of ABSA, the proposed model has the capability to address all ABSA sub-tasks in a comprehensive and unified manner. Overall, the proposed model can achieve competitive performance across all sub-tasks and datasets due to the assistance of domain knowledge and syntactic information.

Secondly, we tackle the sub-tasks of ABSA, namely OTE, AESC, AOP, and ASTE on dataset $\mathbb{D}_2$, and the results are demonstrated in Table 6. It is evident that some baselines can achieve better results for sub-tasks of ABSA than our model. Specifically, SynGen obtains better results for AESC and AOP tasks by incorporating syntax features into BART to learn the syntactic knowledge. However, the proposed model can outperform all baselines on datasets *Res14*, *Res15*, and *Res16* for the sub-task AOP. Because OTE is a single target extraction task, the dependency relation does not bring many contributions to such a task. For triplet extraction tasks, our model performs the best of all on dataset *Lap14*, explicitly demonstrating the effectiveness of the proposed network in capturing direct and indirect interactions among targets. Although our model's performance for AESC appears slightly lower than SynGen, the proposed model significantly outperforms all BERT-based models in other sub-tasks. It indicates that BERT can learn contextualised representations from sentences, while it is still an unsolved challenge to leverage BERT alone for domain-specific tasks, e.g., ABSA. Since BERT is trained on Wikipedia datasets and has almost no understanding of opinion text, it is not the best language model for opinion-aware tasks [68]. Therefore, domain BERT is applied in our model to address these challenges and improve the performance of ABSA sub-tasks.

Thirdly, the comparisons of our model with the latest baselines for ASTE on dataset $\mathbb{D}_3$ are presented in Table 7. It can be observed that SPAN-ASTE shows a superior result on Res14. However, SPAN-ASTE needs to encode all possible spans in a sentence, which is not applicable to long reviews. Moreover, SPAN-ASTE is a triplet-specific method and is not able to solve other sub-tasks of ABSA. In contrast, our model demonstrates a consistent improvement in terms of F1 score on all other datasets. It is worth noting that Li-Unified+, RINANTE+, and TS use either GloVe [87] or Word2Vec [88] to obtain word embeddings, and their performances are far behind those achieved by BERT-based models. Therefore, the pre-trained BERT can capture more informative contextual features than GloVe and Word2Vec. With the assistance of domain BERT and syntactic structure, our model is able to further enhance the performance of ABSA sub-tasks.

5.4. Further Analysis

To better understand the embedding difference between BERT and domain BERT, we plot the distributions of aspect-opinion and sentiment embedding in Figures 6 and 7. As can be observed from Figure 6, there are more clear embedding clusters of aspect and opinion terms from domain BERT than BERT. This phenomenon appears more prominent for three categories, i.e., *POS*, *NEG*, *NEU*, of sentiment in Figure 7. The embedding points

Table 6: The experimental results for OTE, AESC, AOP, and ASTE on dataset $\mathbb{D}_2$.

	Model	OTE			AESC			AOP			ASTE		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
Lap14	Li-Unified+	76.62	74.90	75.70	66.28	60.71	63.38	52.29	52.94	52.56	42.25	42.78	42.47
	RINANTE+	78.20	62.70	69.60	41.20	33.20	36.70	34.40	26.20	29.70	23.10	17.60	20.00
	TS	78.22	71.84	74.84	63.15	61.55	62.34	50.00	58.47	53.85	40.40	47.24	43.50
	OTE-MTL	-	-	-	-	-	-	-	-	-	50.52	39.71	44.31
	RACL-BERT	77.58	81.22	79.36	59.75	68.90	64.00	54.22	66.94	59.90	41.99	51.84	46.39
	EMC-GCN	-	-	-	-	-	-	-	-	-	61.70	56.26	58.81
	BART-ABSA	-	-	-	-	-	68.17	-	-	66.11	-	-	57.59
	SentiPrompt	-	-	-	-	-	-	-	-	-	63.40	58.60	60.90
	DMRC	-	-	-	67.45	61.96	64.59	65.43	61.43	63.37	57.39	53.88	55.58
	BMRC	84.67	67.18	74.90	72.73	62.59	67.27	74.11	61.92	67.45	65.12	54.41	59.27
	SynGen	-	-	-	-	-	**70.06**	-	-	**68.53**	-	-	60.71
	Ours	81.78	74.89	**80.26**	68.09	68.89	68.48	69.65	66.53	68.06	57.17	64.83	**60.76**
Res14	Li-Unified+	81.20	83.18	82.13	73.15	74.44	73.79	44.37	73.67	55.34	41.44	68.79	51.68
	RINANTE+	81.06	72.05	76.29	48.97	47.36	48.15	42.32	51.08	46.29	31.07	37.63	34.03
	TS	84.72	80.39	82.45	76.60	67.84	71.95	47.76	68.10	56.10	44.18	62.99	51.89
	OTE-MTL	-	-	-	-	-	-	-	-	-	66.04	56.25	60.62
	RACL-BERT	82.28	90.49	**86.19**	75.57	82.23	78.76	73.58	67.87	70.61	62.64	57.77	60.11
	EMC-GCN	-	-	-	-	-	-	-	-	-	71.21	72.39	71.78
	BART-ABSA	-	-	-	-	-	78.47	-	-	77.68	-	-	72.46
	SentiPrompt	-	-	-	-	-	-	-	-	-	72.79	72.94	72.86
	DMRC	-	-	-	76.84	76.31	76.57	76.23	73.67	74.93	71.55	69.14	70.32
	BMRC	87.22	82.90	84.99	77.74	75.10	76.39	76.91	75.59	76.23	71.32	70.09	70.69
	SynGen	-	-	-	-	-	**79.72**	-	-	77.59	-	-	**74.02**
	Ours	86.02	85.29	85.65	82.52	77.04	79.68	78.92	78.75	**78.83**	72.51	75.29	73.87
Res15	Li-Unified+	79.18	75.88	77.44	64.95	64.95	64.95	52.75	61.75	56.85	43.34	50.73	46.69
	RINANTE+	77.40	57.00	65.70	46.20	37.40	41.30	37.10	33.90	35.40	29.40	26.90	28.00
	TS	78.07	78.07	78.02	67.65	64.02	65.79	49.22	65.70	56.23	40.97	54.68	46.79
	OTE-MTL	-	-	-	-	-	-	-	-	-	57.51	43.96	49.76
	RACL-BERT	76.25	83.96	79.91	68.35	70.72	69.51	67.89	63.74	65.46	55.45	52.53	53.95
	EMC-GCN	-	-	-	-	-	-	-	-	-	61.54	62.47	61.93
	BART-ABSA	-	-	-	-	-	69.95	-	-	67.98	-	-	60.11
	SentiPrompt	-	-	-	-	-	-	-	-	-	62.97	62.06	62.51
	DMRC	-	-	-	66.84	63.52	65.14	72.43	58.90	64.97	63.78	51.87	57.21
	BMRC	82.99	73.23	77.79	72.41	62.63	67.16	71.59	65.89	68.60	63.71	58.63	61.05
	SynGen	-	-	-	-	-	**71.61**	-	-	69.35	-	-	**64.06**
	Ours	81.32	79.41	**80.10**	69.70	70.22	69.96	74.75	65.71	**69.94**	65.09	60.66	62.80
Res16	Li-Unified+	79.84	86.88	83.16	66.33	74.55	70.20	46.11	64.55	53.75	38.19	53.47	44.51
	RINANTE+	75.00	42.40	54.10	49.40	36.70	42.10	35.70	27.00	30.70	27.10	20.50	23.30
	TS	81.09	86.67	83.73	71.18	72.30	71.73	52.35	70.50	60.04	46.76	62.97	53.62
	OTE-MTL	-	-	-	-	-	-	-	-	-	64.68	54.97	59.36
	RACL-BERT	82.52	91.40	**86.73**	68.53	78.52	73.19	72.77	71.83	72.29	60.78	60.00	60.39
	EMC-GCN	-	-	-	-	-	-	-	-	-	65.62	71.30	68.33
	BART-ABSA	-	-	-	-	-	75.69	-	-	77.38	-	-	69.98
	SentiPrompt	-	-	-	-	-	-	-	-	-	70.20	73.35	**71.74**
	DMRC	-	-	-	69.18	72.59	70.84	77.06	74.41	75.71	68.60	66.24	67.40
	BMRC	85.31	83.01	84.13	73.69	72.69	73.18	76.08	76.99	76.52	67.74	68.56	68.13
	SynGen	-	-	-	-	-	**77.51**	-	-	77.34	-	-	71.26
	Ours	83.87	87.39	85.59	77.98	74.32	76.10	79.38	77.00	**78.16**	71.84	69.67	70.74

Table 7: The experimental results for ASTE on dataset $\mathbb{D}_3$.

Model	Lap14			Res14			Res15			Res16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Li-Unified+	40.56	44.28	42.34	41.04	67.35	51.00	44.72	51.39	47.82	37.33	54.51	44.31
RINANTE+	21.71	18.66	20.07	31.42	39.38	34.95	29.88	30.06	29.97	25.68	22.30	23.87
TS	37.38	50.38	42.87	43.24	63.66	51.46	48.07	57.51	52.32	46.96	64.24	54.21
JET-BERT	55.39	47.33	51.04	70.56	55.94	62.40	64.45	51.96	57.53	70.42	58.37	63.83
SentiPrompt	61.30	55.32	58.15	66.10	63.37	64.71	61.81	62.06	61.93	68.66	69.04	68.85
BART-ABSA	61.41	56.19	58.69	65.52	64.99	65.25	59.14	59.38	59.26	66.60	68.68	67.62
SPAN-ASTE	63.44	55.84	59.38	72.89	70.89	**71.85**	62.18	64.45	63.27	69.45	71.17	70.26
Ours	65.65	54.77	**59.72**	70.03	67.47	68.73	65.38	61.88	**63.58**	73.72	69.28	**71.43**

are sticking closer together with the same category for domain BERT than BERT. The observations show that domain BERT can generate high-quality embeddings and further improve the performance of ABSA sub-tasks.

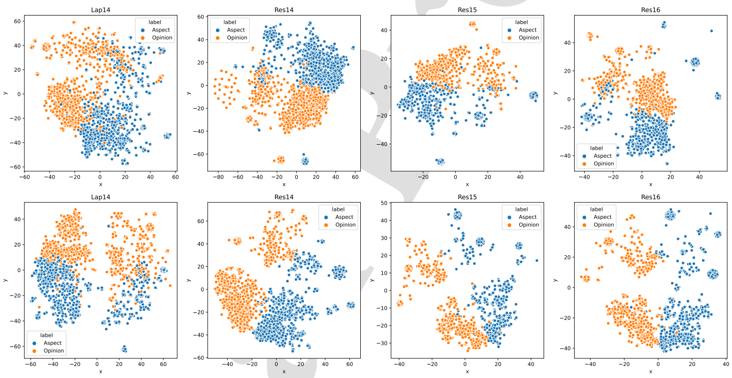


Figure 6: Aspect-Opinion embedding analysis on dataset $\mathbb{D}_2$. The first row is aspect-opinion embeddings from BERT, and the second row is from domain BERT.

5.5. Ablation Study

The ablation study in this section aims to further investigate the impact of domain BERT and MSS for tasks AESC, AOP, and ASTE on dataset $\mathbb{D}_3$. The ablation study is conducted with the following settings:

- *+BERT*, the vanilla BERT [45] is utilised for encoding the context;

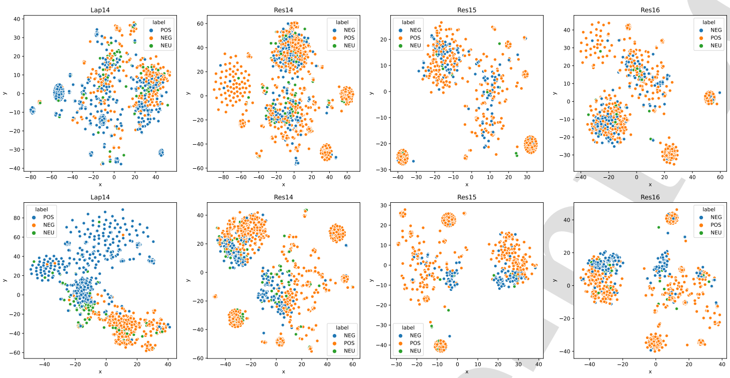


Figure 7: Sentiment embedding analysis on dataset $\mathbb{D}_2$. The first row is aspect-opinion embeddings from BERT, and the second row is from domain BERT.

- *+DomainBERT*, the context embedding is obtained by two domain BERT models;
- *+MSS*, the module MSS with dep-aware and POS-aware attention are incorporated in the proposed model;
- *Full*, all components are applied in our model.

The ablation study analysed the performance of the proposed model with different components and various embeddings on four benchmark datasets. The experimental results are shown in Table 8. By comparing models with different embeddings, we notice that domain BERT contributes to the performance improvement for ABSA sub-tasks, which proves that the learned knowledge in domain BERT is beneficial for sub-tasks of ABSA. ABSA turns out to be a very domain-specific task.

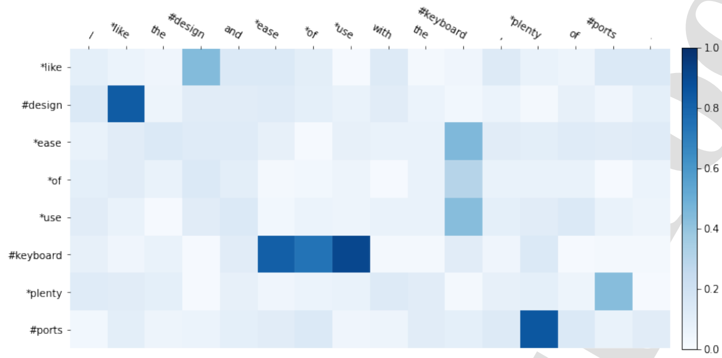
Furthermore, incorporating the MSS module enhances the performance of the proposed model on all four datasets, demonstrating the effectiveness of leveraging syntactic structure via the MSS component. It can be observed from Table 8 that the proposed model can achieve the best performance in terms of F1 score compared to all other variants. The promising findings suggest that both domain knowledge and syntactic structure information can boost semantic embeddings and capture the syntactic correlations between terms. As a result, our proposed approach leads to improved performance in all ABSA sub-tasks. Therefore, the ablation study proves that combining domain-specific knowledge and syntactic information can be an effective approach to enhance the performance of models for ABSA sub-tasks.

Table 8: Ablation study of the proposed model on Dataset $\mathbb{D}_3$.

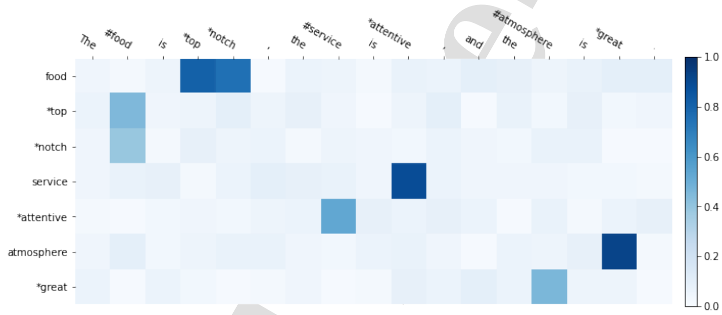
	Datasets	Lap14			Res14			Res15			Res16		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
AESC	+BERT	60.76	62.22	61.48	71.70	64.74	68.04	67.01	60.37	63.52	64.94	72.60	68.55
	+DomainBERT	74.99	56.33	64.34	75.52	66.70	70.84	68.89	66.82	67.84	70.06	73.49	71.73
	+MSS	73.03	60.91	66.42	71.61	75.92	73.70	73.14	66.35	69.58	71.61	75.27	73.39
	Full	71.36	68.56	**69.93**	75.06	75.80	**75.43**	70.81	72.62	**71.71**	80.84	73.27	**76.87**
AOP	+BERT	63.54	60.81	62.15	64.60	72.89	68.50	61.71	63.16	62.43	77.93	66.66	71.86
	+DomainBERT	70.32	58.67	63.97	66.66	75.88	70.97	68.36	63.38	65.77	68.85	78.11	73.19
	+MSS	65.27	69.98	67.54	72.37	71.57	71.96	63.61	73.02	67.99	73.52	77.51	75.46
	Full	76.86	64.13	**69.93**	74.80	73.23	**74.01**	72.40	68.52	**70.41**	81.51	73.49	**77.30**
ASTE	+BERT	49.27	52.21	51.17	58.03	65.48	61.53	69.15	45.61	54.96	62.68	61.04	61.85
	+DomainBERT	52.96	55.75	54.32	63.69	64.60	64.14	60.68	57.17	58.87	66.07	67.67	66.86
	+MSS	60.87	56.72	58.72	67.33	67.26	67.29	61.90	61.24	61.57	67.24	70.88	69.01
	Full	65.65	54.77	**59.72**	70.03	67.47	**68.73**	65.38	61.88	**63.58**	73.72	69.28	**71.43**

5.6. Case Study

In this sub-section, case studies are conducted to present representative triplet extraction examples, proving the effectiveness and validity of the proposed approach. Table 9 shows some extraction results from GT, OTE-MTL, JET-BERT, BMRC, and the proposed framework, where POS and NEG present *positive* and *negative* sentiment, respectively. For the first sentence, both aspect and opinion terms consist of a single word, and most models can predict the triplet correctly. However, there is a long distance between the second aspect term *sound* and the opinion term *tinny*, leading to triplet failure detection for models JET-BERT and OTE-MTL. These models are unlikely to consider such a long dependency distance as useful information for triplet extraction. In the second and fifth samples, multiple aspects and opinion terms are presented. The proposed model can correctly identify all targets by considering dependency relations and types. Our model accurately extracts the aspect term *heat output* and *wait staff* because of modelling the POS knowledge: $heat_{NOUN}\, output_{NOUN}$ and $wait_{NOUN}\, staff_{NOUN}$ in the third and sixth examples. In the fourth sample, the word *long* usually expresses a positive opinion of the corresponding aspect. Our model correctly predicts the sentiment polarity with the domain semantic features from the post-trained language models. For the last sample, the dependency relation $SMALL \xrightarrow{CCONJ} below\ average$ is modelled by the proposed MSS with dep-aware attention, which benefits the extraction of both opinion terms $SMALL$



(a) AOP: (design, like), (keyboard, ease of use), (ports, plenty).



(b) AOP: (food, top notch), (service, attentive), (atmosphere, great).

Figure 8: Visualisation of dep-aware attention mechanism. Marker * indicates the aspect term, and marker # refers to the opinion term.

and *below average*. Figure 8 shows two examples of attention scores from the proposed MSS with dep-aware attention mechanism. For both sentences, the aspect terms have high attention scores with opinion terms, implying that the proposed MSS with attention mechanism can capture important information and improve the performance of target extraction.

Table 9: The inference results from different models. Incorrect predictions are indicated by marker ✗.

Models	Reviews
	But with this laptop , the **bass** is very weak and the **sound** comes out sounding tinny .
GT	[(bass, weak, NEG), (sound, tinny, NEG)]
OTE-MTL	[(bass, weak, NEG), (bass, sounding tinny, NEG)✗]
JET-BERT	[(laptop, weak, NEG)✗, ()✗]
BMRC	[(bass, weak, NEG), (sound, tinny, NEG)]
Ours	[(bass, weak, NEG), (sound, tinny, NEG)]
	I like the **design** and ease of use with the **keyboard** , plenty of **ports** .
GT	[(design, like, POS), (keyboard, ease of use, POS), (ports, plenty, POS)]
OTE-MTL	[(design, like, POS), (design, ease, POS)✗, (keyboard, ease, POS)✗, (ports, plenty, POS)]
JET-BERT	[(use, ease, POS) ✗, (ports, plenty of, POS) ✗]
BMRC	[(design, like, POS), (keyboard, ease, POS)✗, (ports, plenty, POS)]
Ours	[(design, like, POS), (keyboard, ease of use, POS), (ports, plenty, POS)]
	The nicest part is the low **heat output** and ultra quiet **operation** .
GT	[(heat output, low, POS), (operation, quiet, POS)]
OTE-MTL	[()✗, (operation, quiet, POS)]
JET-BERT	[(heat output, low, POS), (operation, quiet, POS)]
BMRC	[(heat output, nicest, POS)✗, (operation, quiet, POS)]
Ours	[(heat output, low, POS), (operation, quiet, POS)]
	Startup times are incredibly long : over two minutes .
GT	[(Startup times, long, NEG)]
OTE-MTL	[(Startup, long, POS)✗]
JET-BERT	[(Startup times, long, POS)✗]
BMRC	[(Startup times, long, NEG)]
Ours	[(Startup times, long, NEG)]
	The **food** is great (big **selection** , reasonable **prices**) and the **drinks** are really good .
GT	[(food, great, POS), (selection, big, POS), (prices, reasonable, POS), (drinks, good, POS)]
OTE-MTL	[(food, great, POS), (selection, big, POS), (selection, reasonable, POS)✗, (prices, reasonable, POS), (drinks, good, NEG)✗]
JET-BERT	[(food, great, POS), ()✗, (prices, reasonable, POS), ()✗]
BMRC	[(food, great, POS), (selection, big, POS), (prices, reasonable, POS), ()✗]
Ours	[(food, great, POS), (selection, big, POS), (prices, reasonable, POS), (drinks, good, POS)]
	The **wait staff** was loud and inconsiderate .
GT	[(wait staff, loud, POS)✗,, (wait staff, inconsiderate, NEG)]
OTE-MTL	[(staff, loud, NEG)✗, (wait staff, inconsiderate, NEG)]
JET-BERT	[(staff, loud, NEG)✗, ()✗]
BMRC	[(wait staff, loud, POS)✗, (wait staff, inconsiderate, POS)✗]
Ours	[(wait staff, loud, NEG), (wait staff, inconsiderate, NEG)]
	Food portion was SMALL and below average .
GT	[(Food portion, SMALL, NEG), (Food portion, below, NEG)✗]
OTE-MTL	[(Food portion, SMALL, POS)✗, (Food, below average, POS)✗]
JET-BERT	[()✗, ()✗]
BMRC	[(Food portion, SMALL, NEG), (Food portion, below average, NEG)]
Ours	[(Food portion, SMALL, NEG), (Food portion, below average, NEG)]

6. Conclusion and Future Work

In this paper, a novel unified network is proposed to solve all sub-tasks of aspect-based sentiment analysis. Two in-domain post-trained BERTs are utilised to obtain semantic embeddings, revealing that the domain-specific information can enhance the performance of ABSA. Most previous studies merely consider the semantic features or parts of syntactic information. In contrast, the proposed model incorporates a MSS module with more informative syntactic structure, further enhancing the semantic representations. Moreover, the MSS with a couple of multi-layer attention mechanisms are applied to exploit indirect relations between terms for precise target extraction. Extensive experiments are conducted by using three groups of real-world datasets. The experimental results demonstrate the effectiveness and superiority of the proposed network. Furthermore, the ablation study is conducted to investigate the impacts of the proposed components of our model. Finally, a case study is presented to exhibit the performance of the proposed network.

In the future, we plan to improve the proposed model from two aspects: (1) the message passing mechanism can be exploited in the MSS module. (2) the learned knowledge from one domain can be transferred to other domain datasets using transfer learning.

7. Acknowledgements

This work was supported by the Callaghan Innovation [CSITR1902, 2020], New Zealand's Innovation Agency.

References

[1] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5 (2012) 1–167.

[2] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, Semeval-2014 task 4: Aspect based sentiment analysis, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014, pp. 27–35.

[3] H. Yan, J. Dai, T. Ji, X. Qiu, Z. Zhang, A unified generative framework for 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, 2021, pp. 2416–2429.

[4] X. Yang, S. Feng, D. Wang, Q. Sun, W. Wu, Y. Zhang, P. Hong, S. Poria, Few-shot joint multimodal aspect-sentiment analysis based on generative multimodal prompt, in: Findings of the Association for Computational Linguistics: ACL 2023, 2023, pp. 11575–11589.

[5] Q. Liu, B. Liu, Y. Zhang, D. S. Kim, Z. Gao, Improving opinion aspect extraction using semantic similarity and aspect associations, in: Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016.

[6] H. Xu, B. Liu, L. Shu, S. Y. Philip, Double embeddings and cnn-based sequence labeling for aspect extraction, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 592–598.

[7] H. Luo, T. Li, B. Liu, B. Wang, H. Unger, Improving aspect term extraction with bidirectional dependency tree representation, IEEE/ACM Transactions on Audio, Speech, and Language Processing 27 (2019) 1201–1212.

[8] X. Wang, H. Xu, X. Sun, G. Tao, Combining fine-tuning with a feature-based approach for aspect extraction on reviews (student abstract), in: Proceedings of the 2020 AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 13951–13952.

[9] C. Wu, F. Wu, S. Wu, Z. Yuan, Y. Huang, A hybrid unsupervised method for aspect term and opinion target extraction, Knowledge-Based Systems 148 (2018) 66–73.

[10] H. Dai, Y. Song, 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, 2019, pp. 5268–5277.

[11] W. Wang, S. J. Pan, Transferable interactive memory network for domain adaptation in fine-grained opinion extraction, in: Proceedings of the 2019 AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 7192–7199.

[12] M. Wu, W. Wang, S. J. Pan, Deep weighted maxsat for aspect-based opinion extraction, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 5618–5628.

[13] A. P. B. Veyseh, N. Nouri, F. Dernoncourt, D. Dou, T. H. Nguyen, Introducing syntactic structures into target opinion word extraction with deep learning, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 8947–8956.

[14] Z. Wu, F. Zhao, X.-Y. Dai, S. Huang, J. Chen, Latent opinions transfer network for target-oriented opinion words extraction, in: Proceedings of the 2020 AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 9298–9305.

[15] Y. Feng, Y. Rao, Y. Tang, N. Wang, H. Liu, Target-specified sequence labeling with multi-head self-attention for target-oriented opinion words extraction, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 1805–1815.

[16] J. Jiang, A. Wang, A. Aizawa, Attention-based relational graph convolutional network for target-oriented opinion words extraction, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, 2021, pp. 1986–1997.

[17] X. Li, L. Bing, W. Lam, B. Shi, Transformation networks for target-oriented sentiment classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 946–956.

[18] R. He, W. S. Lee, H. T. Ng, D. Dahlmeier, Exploiting document knowledge for aspect-level sentiment classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 579–585.

[19] L. Xu, L. Bing, W. Lu, F. Huang, Aspect sentiment classification with aspect-specific opinion spans, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 3561–3567.

[20] A. Sungheetha, R. Sharma, Transcapsule model for sentiment classification, Journal of Artificial Intelligence 2 (2020) 163–169.

[21] D. Ma, S. Li, H. Wang, Joint learning for targeted sentiment analysis, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 4737–4742.

[22] X. Li, L. Bing, P. Li, W. Lam, A unified model for opinion target extraction and target sentiment prediction, in: Proceedings of the 2019 AAAI conference on artificial intelligence, volume 33, 2019, pp. 6714–6721.

[23] M. Hu, Y. Peng, Z. Huang, D. Li, Y. Lv, Open-domain targeted sentiment analysis via span-based extraction and classification, in: Proceed-

ings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 537–546.

[24] Z. Fan, Z. Wu, X. Dai, S. Huang, J. Chen, Target-oriented opinion words extraction with target-fused neural sequence labeling, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 2509–2518.

[25] H. Zhao, L. Huang, R. Zhang, Q. Lu, H. Xue, Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3239–3248.

[26] Z. Wu, C. Ying, F. Zhao, Z. Fan, X. Dai, R. Xia, Grid tagging scheme for aspect-oriented fine-grained opinion extraction, in: Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 2576–2585.

[27] S. Chen, J. Liu, Y. Wang, W. Zhang, Z. Chi, Synchronous double-channel recurrent network for aspect-opinion pair extraction, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 6515–6524.

[28] L. Gao, Y. Wang, T. Liu, J. Wang, L. Zhang, J. Liao, Question-driven span labeling model for aspect-opinion pair extraction, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 12875–12883.

[29] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, L. Si, Knowing what, how and why: A near complete solution for aspect-based sentiment analysis, in: Proceedings of the 2020 AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 8600–8607.

[30] L. Xu, H. Li, W. Lu, L. Bing, Position-aware tagging for aspect sentiment triplet extraction, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 2339–2349.

[31] C. Zhang, Q. Li, D. Song, B. Wang, A multi-task learning framework for opinion triplet extraction, in: Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 819–828.

[32] Z. Chen, H. Huang, B. Liu, X. Shi, H. Jin, Semantic and syntactic enhanced aspect sentiment triplet extraction, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 1474–1483.

[33] Y. Mao, Y. Shen, C. Yu, L. Cai, A joint training dual-mrc framework for aspect based sentiment analysis, in: Proceedings of the 2021 AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 13543–13551.

[34] S. Chen, Y. Wang, J. Liu, Y. Wang, Bidirectional machine reading comprehension for aspect sentiment triplet extraction, in: Proceedings of the 2021 AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 12666–12674.

[35] H. Chen, Z. Zhai, F. Feng, R. Li, X. Wang, Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022, pp. 2974–2985.

[36] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[37] K. Cho, B. van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: Encoder–decoder approaches, in: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014, pp. 103–111.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st Conference on Neural Information Processing Systems, 2017, pp. 5998–6008.

[39] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 2227–2237.

[40] M. Yang, Q. Jiang, Y. Shen, Q. Wu, Z. Zhao, W. Zhou, Hierarchical human-like strategy for aspect-level sentiment classification with sentiment linguistic knowledge and reinforcement learning, Neural Networks 117 (2019) 240–248.

[41] H. Luo, T. Li, B. Liu, J. Zhang, Doer: Dual cross-shared rnn for aspect term-polarity co-extraction, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 591–601.

[42] Y. Zhang, P. Tiwari, D. Song, X. Mao, P. Wang, X. Li, H. M. Pandey, Learning interaction dynamics with an interactive lstm for conversational sentiment analysis, Neural Networks 133 (2021) 40–56.

[43] W. Li, W. Shao, S. Ji, E. Cambria, Bieru: Bidirectional emotional recurrent unit for conversational sentiment analysis, Neurocomputing 467 (2022) 73–82.

[44] Z. Li, L. Li, A. Zhou, H. Lu, Jtsg: a joint term-sentiment generator for aspect-based sentiment analysis, Neurocomputing 459 (2021) 1–9.

[45] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, 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: Human Language Technologies, 2019, pp. 4171–4186.

[46] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).

[47] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7871–7880.

[48] C. Du, H. Sun, J. Wang, Q. Qi, J. Liao, Adversarial and domain-aware bert for cross-domain sentiment analysis, in: Proceedings of the 58th annual meeting of the Association for Computational Linguistics, 2020, pp. 4019–4028.

[49] J. Dai, H. Yan, T. Sun, P. Liu, X. Qiu, Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 1816–1829.

[50] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013, pp. 1631–1642.

[51] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the 2017 Conference on Learning Representations, 2017.

[52] C. Zhang, Q. Li, D. Song, Aspect-based sentiment classification with aspect-specific graph convolutional networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019, pp. 4568–4578.

[53] Y. Tian, G. Chen, Y. Song, Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 2910–2922.

[54] Y. Liang, F. Meng, J. Zhang, Y. Chen, J. Xu, J. Zhou, A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis, Neurocomputing 454 (2021) 291–302.

[55] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, E. Hovy, Dual graph convolutional networks for 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, 2021, pp. 6319–6329.

[56] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177.

[57] E. M. Mercha, H. Benbrahim, Machine learning and deep learning for sentiment analysis across languages: A survey, Neurocomputing 531 (2023) 195–216.

[58] J. Liu, Q. Zhong, L. Ding, H. Jin, B. Du, D. Tao, Unified instance and knowledge alignment pretraining for aspect-based sentiment analysis, arXiv preprint arXiv:2110.13398 (2021).

[59] Y. Han, X. Zhou, G. Wang, Y. Feng, H. Zhao, J. Wang, Fusing sentiment knowledge and inter-aspect dependency based on gated mechanism for aspect-level sentiment classification, Neurocomputing (2023) 126462.

[60] L. Xu, K. Liu, S. Lai, Y. Chen, J. Zhao, Mining opinion words and opinion targets in a two-stage framework, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013, pp. 1764–1773.

[61] X. Li, W. Lam, Deep multi-task learning for aspect term extraction with memory interaction, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2886–2892.

[62] K. Sun, R. Zhang, S. Mensah, Y. Mao, X. Liu, Aspect-level sentiment analysis via convolution over dependency tree, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019, pp. 5679–5688.

[63] S. Wu, H. Fei, Y. Ren, D. Ji, J. Li, Learn from syntax: Improving pair-wise aspect and opinion terms extractionwith rich syntactic knowledge, in: Proceedings of the 30th International Joint Conference on Artificial Intelligence, 2021, p. 3957–3963.

[64] H. Liu, Y. Wu, Q. Liu, W. Lu, X. Li, J. Wei, X. Liu, J. Feng, Enhancing aspect-based sentiment analysis using a dual-gated graph convolutional network via contextual affective knowledge, Neurocomputing (2023) 126526.

[65] W. Wang, S. J. Pan, D. Dahlmeier, X. Xiao, Coupled multi-layer attentions for co-extraction of aspect and opinion terms, in: Proceedings of the 2017 AAAI Conference on Artificial Intelligence, volume 31, 2017.

[66] C. Chen, Z. Teng, Z. Wang, Y. Zhang, Discrete opinion tree induction for aspect-based sentiment analysis, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022, pp. 2051–2064.

[67] Q. Zhong, L. Ding, J. Liu, B. Du, H. Jin, D. Tao, Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis, IEEE Transactions on Knowledge and Data Engineering (2023).

[68] H. Xu, B. Liu, L. Shu, S. Y. Philip, Bert post-training for review reading comprehension and aspect-based sentiment analysis, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 2324–2335.

[69] Y. Lin, Y. C. Tan, R. Frank, Open sesame: Getting inside bert's linguistic knowledge, in: Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2019, pp. 241–253.

[70] J. L. Ba, J. R. Kiros, G. E. Hinton, Layer normalization, arXiv preprint arXiv:1607.06450 (2016).

[71] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, in: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 486–495.

[72] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al., Semeval-2016 task 5: Aspect based sentiment analysis, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), 2016, pp. 19–30.

[73] W. Wang, S. J. Pan, D. Dahlmeier, X. Xiao, Recursive neural conditional random fields for aspect-based sentiment analysis, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 616–626.

[74] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A python natural language processing toolkit for many human languages, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2020, pp. 101–108.

[75] P. K. Diederik, B. Jimmy, Adam: A method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations, 2015.

[76] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Computational Linguistics 37 (2011) 9–27.

[77] X. Zhang, Y. Jiang, H. Peng, K. Tu, D. Goldwasser, Semi-supervised structured prediction with neural crf autoencoder, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 1701–1711.

[78] J. D. Lafferty, A. McCallum, F. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in: Proceedings of the 18th International Conference on Machine Learning, 2001, pp. 282–289.

[79] P. Liu, S. Joty, H. Meng, Fine-grained opinion mining with recurrent neural networks and word embeddings, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1433–1443.

[80] H. Xu, B. Liu, L. Shu, S. Y. Philip, Dombert: Domain-oriented language model for aspect-based sentiment analysis, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 1725–1731.

[81] Y. Tian, L. Yang, Y. Sun, D. Liu, Cross-domain end-to-end aspect-based sentiment analysis with domain-dependent embeddings, Complexity 2021 (2021) 1–11.

[82] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang, et al., Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis, arXiv preprint arXiv:2109.08306 (2021).

[83] L. Xu, Y. K. Chia, L. Bing, Learning span-level interactions for aspect sentiment triplet extraction, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021, pp. 4755–4766.

[84] R. He, W. S. Lee, H. T. Ng, D. Dahlmeier, 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, 2019, pp. 504–515.

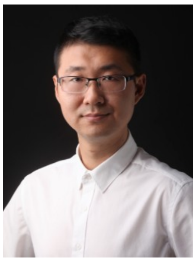
[85] Z. Chen, T. Qian, Relation-aware collaborative learning for unified aspect-based sentiment analysis, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3685–3694.

[86] C. Yu, T. Wu, J. Li, X. Bai, Y. Yang, Syngen: A syntactic plug-and-play module for generative aspect-based sentiment analysis, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.

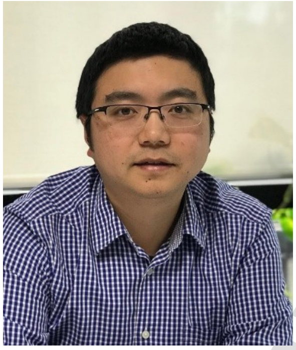
[87] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543.

[88] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013, p. 3111–3119.

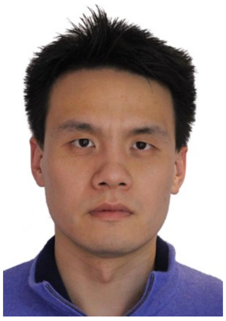
Jingli Shi received his M.Tech degree from the National University of Singapore in 2019. He is currently pursuing his Ph.D. from Auckland University of Technology, New Zealand. His research areas are natural language processing, deep learning, and sentiment analysis.



Dr **Weihua Li** received his PhD from Auckland University of Technology, New Zealand in 2018 and his M.Tech from the National University of Singapore in 2014. He is currently a senior lecturer in the School of Engineering, Computer & Mathematical Sciences. Weihua is an active researcher, and his research interest mainly focuses on Artificial Intelligence, including social intelligence, agent-based modelling, natural language processing, and problem-solving using artificial intelligence techniques. Weihua has over 40 publications in the field of Artificial Intelligence and is involved in the organisation of a number of international conferences.



Quan Bai is an associate professor in the School of Information & Communication Technology, University of Tasmania, Australia. His research has focused on machine learning, intelligent systems, and AI-enabled human-computer interactions. He is the leader of the AI Research Group at the University of Tasmania. He completed his Ph.D. degree in Computer Science at the University of Wollongong in 2007. Since then, he has held positions as a postdoctoral research fellow at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), and a lecturer/senior lecturer at Auckland University of Technology. He has over 140 research publications and has taken leadership roles in the organisation of several international AI conferences, including AAMAS, PRICAI, PRIMA, and AUS AI.



Yi Yang received his Ph.D. degree from Auckland University of Technology, Auckland, New Zealand, in 2020, and he is currently working as a lecturer at the School of Computer Science and Information Engineering, Hefei University of Technology, China. His research interests include crowdsourcing, computer vision, and adversarial machine learning.

Jianhua Jiang was born in Hangzhou, China, in 1979. He received the B.S., M.S., and Ph.D. degrees in computer science from Jilin University, Changchun, China, in 2003, 2006, and 2010, respectively. Since 2019, he has been a full professor at Jilin University of Finance and Economics, China. He has authored three books and more than 50 research papers. His research interests include artificial intelligence, finance technology, and E-Commerce. He is a member of the China Computer Federation (CCF), Society of Management Science and Engineering of China, and China Information Economics Society.
Prof. Dr. Jianhua Jiang is the dean of the School of Management Science and Information Engineering, the Director of the Jilin Province Key Laboratory of Fintech, and the Director of the Center for Artificial Intelligence at Jilin University of Finance and Economics, and the Chief Leader in the discipline of Management Science and Engineering in Jilin University of Finance and Economics.

Jingli Shi: Writing – original draft, Conceptualization, Data curation, Methodology, Software.

Weihua Li: Conceptualization, Project administration, Writing – review & editing, Supervision.

Quan Bai: Formal analysis, Project administration, Writing – review, Supervision.

Yi Yang: Writing – review, Formal analysis, Resources, Validation.

Jianhua Jiang: Writing – review, Formal analysis, Resources, Validation.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Jingli Sih reports financial support was provided by Callaghan Innovation, New Zealand.