DOMAIN-ADAPTIVE SENTIMENT ANALYSIS ACROSS ONLINE SOCIAL NETWORKS

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Supervisors
Dr Weihua Li
A/Prof. Sira Yongchareon
A/Prof. Jian Yu
A/Prof. Quan Bai

March 2023

By
Jingli Shi
School of Engineering, Computer and Mathematical Sciences

Copyright

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signature of candidate

Acknowledgements

I would like to express my sincerest appreciation to my supervisors, Dr. Weihua Li, A/Prof. Sira Yongchareon, A/Prof. Jian Yu, and A/Prof. Quan Bai, for their guidance, support, and encouragement throughout my Ph.D. journey. Their expertise, patience, and dedication have been instrumental in helping me complete my research. I am truly grateful for the time and effort they have invested in me and for the knowledge and skills they have imparted. Their insightful comments and constructive feedback have helped me to improve my research and writing skills significantly. I would also like to thank them for their unwavering support and encouragement, even during the most challenging times. Their guidance and mentorship have been invaluable and have helped me to achieve my academic goals.

I would also like to extend my heartfelt thanks to CBA Strategic IT, especially to the managing director Christine Bear, for providing me with the opportunity to conduct my research in a dynamic and supportive environment. Your unwavering support and encouragement have been invaluable in helping me achieve this milestone. Moreover, I want to express my gratitude to Callaghan Innovation for their financial support. Their contribution has allowed me to focus solely on my research, and has made a significant impact on the outcome of my work.

Research is a collaborative effort and I am deeply grateful to my colleagues Mengyan Wang and Guan Wang, for their support and friendship throughout my Ph.D. journey. They have been a constant source of motivation and inspiration, and their contributions to my research have been invaluable.

Finally, I would like to express my deepest gratitude to my family, for their unwavering support and encouragement throughout the years. Their love and understanding have been the foundation on which I have built my academic career. I would not have been able to complete this journey without their love and support.

Abstract

Aspect-based sentiment analysis is an important task in natural language processing and has a wide range of applications in fields such as e-commerce, marketing, and customer service. The goal of this task is to identify aspect and opinion terms and classify the sentiment expressed towards a particular aspect in a given text. Despite its significance, aspect-based sentiment analysis remains a challenging task due to limitations in existing models. These limitations include an inadequate consideration of crucial implicit linguistic features for aspect term extraction, declining performance on unstructured and small datasets for aspect and relation extraction, a complex and varied model landscape for different sub-tasks, and the time-consuming construction of prompts for cross-domain aspect term extraction. In this thesis, these challenges are tackled by employing several innovative deep neural network models.

First, a novel and efficient framework is introduced for extracting aspect terms by combining contextual and linguistic features using the Artificial Bee Colony-based feature selection method. To address the high sparsity and dimensionality of raw data, an improved version of Artificial Bee Colony is employed to determine the most relevant linguistic features. These selected features and context embeddings are then integrated to enhance the accuracy of aspect extraction.

Second, a novel and deep learning-based model is proposed for recognising individuals' concerns and the associated relationships through the integration of Graph Convolutional Networks, Bi-directional Long Short-Term Memory, and Concern Graphs.

The proposed model leverages sequential features from BERT embeddings and regional features of tweets extracted through the Concern Graph module, leading to improved concern detection and high resistance to noise. This approach overcomes the limitations of limited manually labelled data.

Third, a novel integrated framework is designed to tackle all defined sub-tasks of aspect-based sentiment analysis. The framework consists of a multi-layer semantic model based on graph convolutional networks, which is designed to capture the semantic connections between aspects and opinion terms. Additionally, a multi-layer syntax model is proposed to learn explicit dependency relations at different levels. To support the sub-tasks, the semantic features learned by the semantic model are passed to the syntax model, providing enhanced semantic guidance and allowing the syntax model to learn more comprehensive syntactic representations. The framework incorporates two attention mechanisms, one for modelling dependency relations and types and another for encoding part-of-speech tags to detect aspect and opinion term boundaries. This differs from conventional syntactic models, making the proposed framework unique.

Finally, a novel soft prompt-based joint learning method is proposed for aspect term extraction across domains. The method utilises external linguistic features to learn domain-invariant representations between source and target domains through multiple objectives, effectively bridging the gap between domains with varying distributions of aspect terms. Furthermore, the method incorporates a set of transferable soft prompts, consisting of multiple learnable vectors, to enhance the detection of aspect terms in the target domain. The proposed soft prompt-based joint learning method represents a novel approach to cross-domain aspect term extraction.

Publications

Shi, J., Li, W., Bai, Q., Yang, Y., & Jiang, J. Soft Prompt-Based Joint Learning for Cross-Domain Aspect Term Extraction. Intelligent Systems with Applications (submitted), 2023

Shi, J., Li, W., Bai, Q., Yang, Y., & Jiang, J. Syntax-Enhanced Aspect-Based Sentiment Analysis with Multi-Layer Attention. Neurocomputing (submitted), 2022

Shi, J., Li, W., Bai, Q., & Ito, T. (2022). BeeAE: Effective Aspect Term Extraction with Artificial Bee Colony. The Journal of Supercomputing, 78(16), pp. 17969-17991.

Shi, J., Li, W., Yongchareon, S., Yang, Y., & Bai, Q. (2022). Graph-based Joint Pandemic Concern and Relation Extraction on Twitter. Expert Systems with Applications, 195, pp. 116538.

Shi, J., Hellesoe, L., Wang, G., Li, W., & Bai, Q. (2022). Focus-Based Text Summarisation with Hybrid Embeddings. In AI 2021: Advances in Artificial Intelligence. Springer, pp. 705-715.

Shi, J., Li, W., Yang, Y., Yao, N., Bai, Q., Yongchareon, S., & Yu, J. (2021). Automated Concern Exploration in Pandemic Situations-Covid-19 as a Use Case. In Knowledge Management and Acquisition for Intelligent Systems: 17th Pacific Rim Knowledge Acquisition Workshop. Springer, Proceedings 17, pp. 178-185.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Recently, social media platforms have become a popular outlet for individuals to share their opinions and emotions regarding products, services, or events (Zhuang, Jing & Zhu, 2006; Kumar & Harish, 2018). The analysis of public opinions has drawn significant attention from both academic researchers and professionals (Jin & Ho, 2009; F. Li et al., 2010). In the e-commerce industry, consumers frequently turn to online reviews and feedback before making a purchase, while companies rely on these reviews to assist their business decisions and improve the quality of their offerings. However, manually reading through a vast number of reviews is not feasible due to the rapid pace at which they are generated. Therefore, it is crucial to use opinion mining or sentiment analysis to help users extract valuable information from a large volume of reviews. A comprehensive analysis requires analysing the sentiment of reviews at the aspect level, which is best accomplished through the use of deep learning networks. These networks have gained popularity in aspect-based sentiment analysis due to their ability to capture intricate relationships between words and phrases in the text.

In this section, aspect-based sentiment analysis and its role in deciphering entity-specific sentiments is first introduced. This is followed by a sub-section about the significance of deep learning networks in aspect-based sentiment analysis. The research motivations sub-section outlines identified gaps existed in previous study. Research questions are provide direction to the investigation. The research methodology is then detailed. Finally, the chapter presents contributions and advancements in the field of aspect-based sentiment analysis.

1.1 Aspect-based Sentiment Analysis

As an important area of Sentiment Analysis (SA), Aspect-Based Sentiment Analysis (ABSA) concentrates on discovering fine-grained sentiments related to different aspects of a given item (Pontiki et al., 2014; Yan, Dai, Ji, Qiu & Zhang, 2021). ABSA encompasses several 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). In all of these sub-tasks, there are three crucial elements: **Aspect Term**, which refers to the word or phrase in a customer review describing a product or service, **Opinion Term**, representing the word or phrase that expresses the customer's attitude towards the aspect, and **Sentiment Polarity**, identified as positive, negative, or neutral towards the aspect in a sentence.

For instance, in Figure 1.1, consider the following two reviews, "The mouse is smooth and great", and "The book is boring." In both reviews, the objective of aspect-based sentiment analysis is to extract aspect terms "mouse" and "book" and opinion terms "smooth, great" and "boring", and classify the sentiment of two reviews as "Positive" or "Negative" in two domains Electronics and Book. This task can detect exactly what people like or dislike at the aspect level.

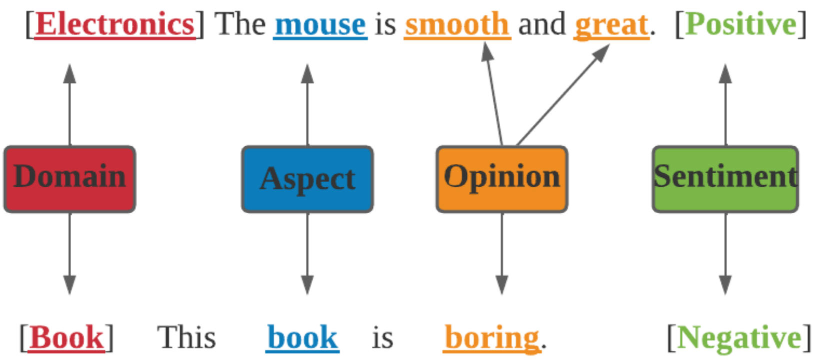


Figure 1.1: An example of sentiment analysis at the aspect level, which includes domain, aspect, opinion, and sentiment.

1.2 Deep Learning Networks

In this section, I provide a brief overview of the deep learning networks employed in the proposed models and the pre-trained language models utilised to train these models.

1.2.1 Pre-Trained Language Models

In recent years, pre-trained language models, i.e., Embeddings from Language Models (ELMO) (Peters et al., 2018) and Bidirectional Encoder Representations from Transformers (BERT) (Devlin, Chang, Lee & Toutanova, 2019a) achieve remarkable performance in the Natural Language Processing (NLP) field and are applied in numerous tasks, e.g., aspect term extraction, sentiment classification, etc. The pre-trained language models are trained on a large scale of unlabelled text data, and the models are much faster to train for new specific tasks without the requirement for annotated data.

Bidirectional Encoder Representations from Transformers

In this work, BERT is utilised as a word embedding module for automated aspect extraction. BERT is based on two major concepts to keep many of the recent advances in the NLP field: (1) the transformer architecture; (2) unsupervised pre-training. The

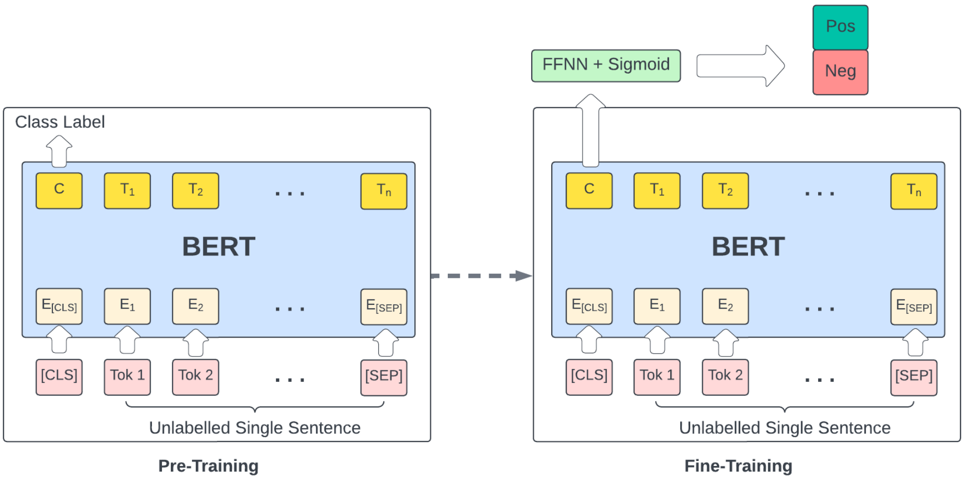


Figure 1.2: An Example of BERT Application for Sentiment Classification.

transformer is a new network architecture based solely on attention mechanisms and removes the classical recurrent and convolutional components in sequence transduction models (Vaswani et al., 2017). The BERT model includes two steps: pre-training and fine-tuning. For pre-training, BERT is trained on unannotated data for different tasks. During fine-tuning, the pre-trained parameters are first used to initialise BERT. Then the parameters are fine-tuned with annotated data for downstream tasks.

Figure 1.2 shows an example of BERT fine-tuning for sentiment classification. The input representation can represent both the sentence and the pair of sentences in a sequence of tokens to process different downstream tasks. For each input, the first token is [CLS], i.e., a special classification token. The last token in one input is [SEP], separating different sentences. The final input representation can be constructed by summing position embeddings, segment embeddings, and token embeddings, as shown in Figure 1.3.

For tasks such as ABSA, performance has been shown to improve with the help of additional training on Review text, called Post-Training (H. Xu, Liu, Shu & Philip, 2019). Therefore, in this thesis, BERT is applied to learn language semantic and basic

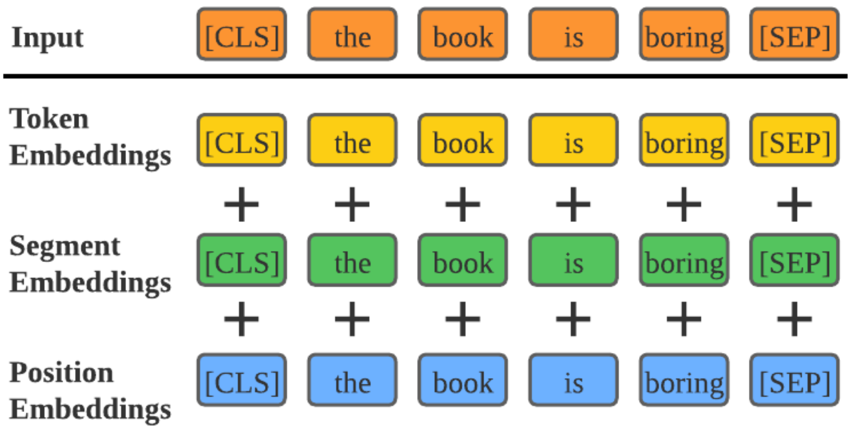


Figure 1.3: Input Representation of BERT.

syntactic knowledge from context to improve the performance of ABSA.

Text-To-Text Transfer Transformer

For the pre-trained language models, it's challenging to determine which enhancements are most impactful and assess their effectiveness when used in conjunction (Radford, Narasimhan, Salimans, Sutskever et al., 2018). By reframing NLP tasks into a unified text-to-text format, a Text-To-Text Transfer Transformer (T5) model is proposed (Raffel et al., 2020). Unlike BERT where the output can only be a class label or a span of the input, the output can be text strings, as shown in Figure 1.4 [1].

With the structure of text-to-text, it's able to use the same model, loss function, and hyper-parameters on many NLP tasks, such as machine translation (Agarwal, Kale, Ge, Shakeri & Al-Rfou, 2020), question-answering (Ngai, Park, Chen & Parsapoor, 2021), sentiment analysis (T. Gao et al., 2022). In this thesis, T5 is integrated into the proposed model to identify aspect terms in different domains.

[1] https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html

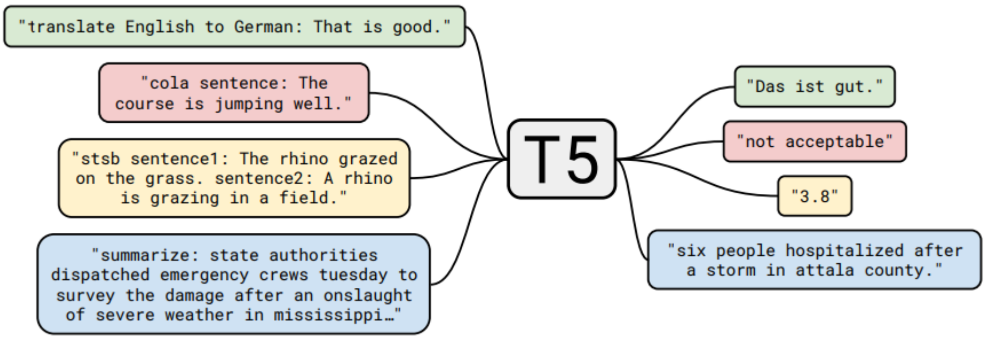


Figure 1.4: The application of T5 on NLP tasks, e.g., translation, question-answering, and classification.

1.2.2 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) have gained widespread popularity in sentiment analysis. The foundation of RNN models is a fixed-sized vector that represents a sequence, such as a sentence or a document, by feeding each token into a recurrent unit. This allows the model to capture the sequential nature of language, where the meaning of one word is influenced by the previous words (Goldberg, 2016). Moreover, RNN models offer the advantage of flexible computation steps, where the output of the model depends on previous computations. This makes RNNs capable of capturing context dependencies in language and handling texts of varying lengths (Tang, Qin & Liu, 2016).

Long Short-Term Memory

The simple RNN faces limitations due to the gradient, which can either vanish (approach zero) or explode (become excessively high) during the backpropagation process, making it challenging to train and fine-tune the parameters (Goldberg, 2016). This issue has been addressed with the development of networks such as Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho, Van Merriënboer et al., 2014), which have improved upon these limitations.

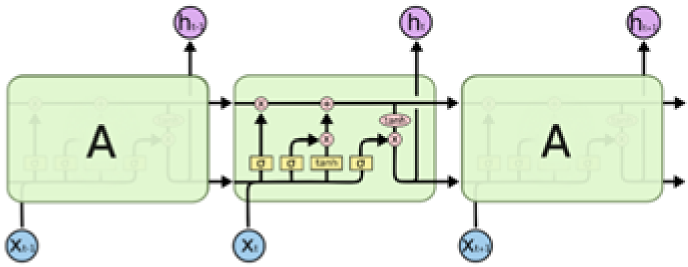


Figure 1.5: The framework of LSTM.

LSTM is one of the recurrent neural networks, as shown in Figure 1.5 [2]. Different from other RNNs, LSTM transforms the Tanh layer in RNNs into a new structure, including one memory unit and one gate mechanism. Thus, LSTM can decide how to use and update the information kept in the memory unit.

1.2.3 Graph Convolutional Network

Applying well-established neural models such as RNNs or Convolutional Neural Networks (CNNs) to arbitrarily structured graphs is a difficult task and presents a significant challenge. To address such a problem, Graph Convolutional Network (GCN) is proposed to leverage the graph structure and aggregate node information from the neighbourhoods (Kipf & Welling, 2017).

Graph Convolutional Networks possess a remarkable ability to learn representations of graph structural datasets and have demonstrated exceptional results in a broad range of NLP tasks and applications, such as Neural Machine Translation (NMT) (D. Cai & Lam, 2020), text classification (Yao, Mao & Luo, 2019), relation extraction (Hong, Liu, Yang, Zhang & Hu, 2020), Named Entity Extraction (NER) (X. Sun et al., 2022), sentiment analysis (H. Cai, Tu, Zhou, Yu & Xia, 2020).

[2]https://colah.github.io/posts/2015-08-Understanding-LSTMs

1.2.4 Model Learning Mechanisms

To adapt pre-trained language models to downstream tasks, there are two widely-used learning mechanisms, i.e., fine tuning and prompt tuning.

Fine Tuning

Fine-tuning is a popular learning strategy in deep learning that involves adapting a pre-trained language model to perform a specific task. The process involves updating the weights of a pre-trained language model using task-specific data, allowing the model to specialise in the target task while retaining the knowledge acquired during pre-training.

Fine-tuning has been widely used in various NLP tasks, such as NER (K. Xue et al., 2019), relation extraction (C. Lin, Miller, Dligach, Bethard & Savova, 2019), text summarisation (Kahla, Yang & Novák, 2021), sentiment analysis (Geetha & Renuka, 2021), etc., with great success. However, despite its popularity and success, fine-tuning also presents challenges, such as the memory consumption during training and the maintenance of multiple copies of model parameters for each task during inference.

Prompt Tuning

Prompt tuning is a relatively new technique in deep learning that involves adjusting the prompt, or the input sequence, of a pre-trained language model to improve its performance on a specific task. The idea behind prompt tuning is to leverage the powerful representations learned by large pre-trained models while fine-tuning only small portion parameters of the model to adapt it to the target task. This technique has the potential to significantly reduce the computational cost and memory requirement compared to traditional fine tuning, while still achieving comparable or even better performance.

To tackle the challenges in learning brought about by the growing size of language models, prompt tuning-based methods have been proposed as a solution. These methods utilise language prompts and task descriptions as context to make aspect-based sentiment analysis more aligned with language modelling. Early studies have focused on hard templates, which are manually crafted for single-domain ABSA tasks (C. Li et al., 2021; T. Gao et al., 2022; H. Li et al., 2022; Ben-David, Oved & Reichart, 2022).

However, the design of a prompt requires domain knowledge. To address this, soft prompts have been introduced as an alternative input. These prompts are constructed using several learnable vectors rather than human-interpretable natural language, allowing language models to effectively perform specific tasks without manual design. While much of the existing research has centered on sentiment classification (H. Wu & Shi, 2022; Asai, Salehi, Peters & Hajishirzi, 2022), this thesis explores the use of soft prompt based models for cross-domain aspect term extraction.

1.3 Research Motivations

Nowadays, a massive amount of opinionated textual data is generated daily along with the unprecedented growth of social media, online reviews websites, and e-commerce, e.g., Facebook, Twitter, Amazon, eBay, Yelp, etc. Such sentiment information can affect our behaviour and decision in daily life. Therefore, how to efficiently mine valuable information, i.e., people's opinions, sentiments, and attitudes, from online unstructured textual data has attracted great attention from researchers. To understand the rationale and significance of this study, I discuss it in three directions: aspect extraction, aspect-opinion extraction, and cross-domain sentiment analysis.

It is a common practice for customers to release reviews on the Web to express their opinions about products or services. Therefore, the owner of products or services needs to deal with the increasingly large-scale textual data to identify the strengths or weaknesses of their products or services. Especially, they want to find out the opinion targets complaints to understand the gaps between what business promises in terms of the products or services. To tackle this challenge, aspect terms should be extracted for sentiment analysis. As an important subtask of aspect-based sentiment analysis, aspect extraction aims to identify predefined opinion target words or phrases of online reviews (Pontiki et al., 2014). Researchers utilise many different approaches for aspect extraction. The existing methods can be divided into rule-based (Poria, Cambria, Ku, Gui & Gelbukh, 2014; Zainuddin, Selamat & Ibrahim, 2018), statistic-based (L. Sun, Li, Li & Lv, 2014; Schouten, Van Der Weijde, Frasincar & Dekker, 2017; Dosoula et al., 2016), corpus-based (H.-Y. Chen & Chen, 2016), pattern-based (Chatterji, Varshney & Rahul, 2017; Schouten et al., 2017) and machine learning-based (H. Xu, Zhang & Wang, 2015; J. Feng, Cai & Ma, 2019; Tulkens & van Cranenburgh, 2020). However, these approaches either dependent on manually annotated data, which require much time and financial resources, or the linguistic features (i.e., lemma, tag, dep, shape, etc.) and inherent structure (i.e., the relation between aspects) are neglected, which are critical features for aspect extraction.

Aspect and opinion extraction is a significant research direction, which has attracted increasing attention in research and is the key step in sentiment analysis. However, the aspect and opinion extraction task is largely neglected in existing methods, which leaves a critical gap to subsequent sentiment analysis tasks, such as pair-level sentiment classification, pair-level opinion clustering, etc. Some approaches have achieved significant progress on this task. For example, a bootstrapping-based method is presented to expand the initial opinion lexicon (Qiu, Liu, Bu & Chen, 2011). A partially-supervised word alignment method is proposed to extract opinion and aspect terms jointly (K. Liu, Xu, Liu & Zhao, 2013). Recently, deep learning-based models are proposed to jointly extract aspect and opinion terms (K. Liu, Xu & Zhao, 2014; H. Dai & Song, 2019; S. Chen, Liu, Wang, Zhang & Chi, 2020). However, the existing methods either ignore the internal relationship between aspect terms and opinion phrase, or consider both terms extraction as separate subtasks. Unlike current studies, this research will focus on the relation between aspect and opinion, and the automated generation of syntax features. Furthermore, this work will be able to ensure a more fine-grained sentiment analysis for product reviews and benefit several applications such as opinion summarization and product profiling. By extracting aspect-opinion pairs, it can provide sufficient information for the decision-making of buyers and sellers in e-business.

Deep learning has benefited various applications in Computer Vision (CV) and Natural Language Process (NLP). However, it is still challenged by domain shifts in NLP. For example, the sentiment is predicted as positive for “fast” expressed towards service in the reviews on the restaurant, while I cannot transfer the sentiment result to “fast” towards power consumption in the reviews on the laptop. Such domain shift becomes a challenge for most existing deep learning models because they fail to maintain comparable performance across domains.

In order to process the large scale of unlabelled and imbalanced data, it has become crucial to transfer learned knowledge to analysis unseen data. In this thesis, I also investigate cross-domain aspect term extraction, which aims to utilise the learned knowledge in the source dataset to predict the aspect terms in target datasets. There are some studies focused on the domain adaptation aspect term extraction. The external linguistic information is integrated into their language model using a self-attention mechanism for extracting cross-domain aspect terms (Pereg, Korat & Wasserblat, 2020). This method combines the inherent linguistic abilities of language models with additional syntactic features from external sources, allowing it to bridge the divide between source and target domains. To enhance the model's ability to handle target domain datasets and improve its robustness, domain-specific knowledge graphs of aspect terms is constructed automatically, and features from these graphs are incorporated into language models for cross-domain aspect term extraction (Howard et al., 2022).

However, the existing methods are only trained on the source dataset and neglect the sentiment-specific features in both source and target datasets. This research will learn domain-specific and domain-invariant features from both datasets, which sentient information lying in both data can be exploited.

1.4 Research Questions

Based on the research motivations discussed above, three research questions are proposed as follows.

Research Question 1: How effective are contextual and linguistic features in improving aspect extraction?

- **Sub-Research Question 1.1:** How to optimise linguistic features with contextual knowledge to improve aspect term extraction?
- **Sub-Research Question 1.2:** How to construct knowledge graphs to extract regional features to extract aspect terms and the corresponding relations?

Research Question 2: How to effectively handle sub-tasks using a unified framework for aspect-based sentiment analysis?

- **Sub-Research Question 2.1:** How to model dependency relation and type to learn explicit dependency relations between aspect and opinion terms?
- **Sub-Research Question 2.2:** How to encode part-of-speech tags to detect aspect and opinion term boundaries?

Research Question 3: How effective are domain-specific and domain-invariant features in cross-domain aspect-based sentiment analysis?

- **Sub-Research Question 3.1:** How to transfer domain-invariant and domain-specific knowledge among different domains?

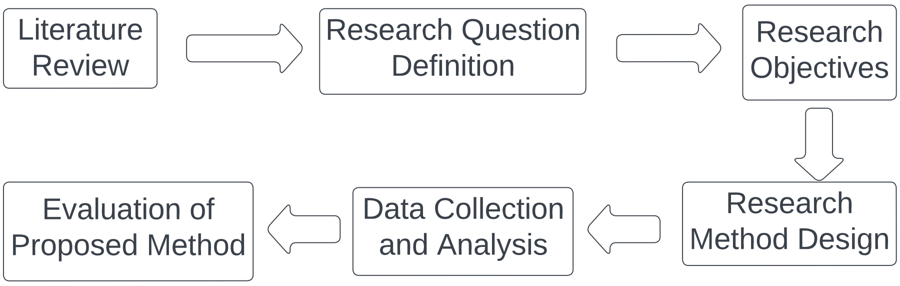


Figure 1.6: Research Methodology used in this thesis

- **Sub-Research Question 3.2:** How to interpolate transferable soft prompts into the prompt-tuning model to learn domain-invariant knowledge?

1.5 Design of Study

In this section, the research methodology employed in this thesis is presented to describe the research process. Moreover, the dataset and evaluation metrics are explained in details.

1.5.1 Research Methodology

The research methodology of this thesis presents an interactive process with six key steps, including research methodology, research method design, data analysis, and research method evaluation, as shown in Figure 1.6.

First, the literature review is conducted to explore the related research works and trace the progress of the field. Second, research questions are defined by discovering issues and finding gaps in existing studies. In order to provide answers to defined research questions, research objectives are described to specify the intended outcome of the research study. Next, research methods are designed according to the listed objectives. Then, the initial analysis of the collected dataset is carried out, and experiments are conducted on the dataset by optimising the parameters of the designed methods. Finally, experimental results are analysed, and the proposed methods are evaluated through comparison with similar models.

1.5.2 Dataset

To comprehensively evaluate the proposed methods in comparison to state-of-the-art models, the experiments within the thesis are conducted on a diverse of datasets across various domains.

- ***SemEval*** is the most commonly used dataset in sentiment analysis, sourced from SemEval challenge (Pontiki et al., 2014; Pontiki, Galanis, Papageorgiou, Manandhar & Androutsopoulos, 2015; Pontiki et al., 2016). The dataset consists of two domains, namely restaurant and laptop, distributed across three sub-datasets: SemEval-2014, SemEval-2015, and SemEval-2016. With its annotations tailored to diverse aspect-based sentiment analysis tasks, it addresses our research questions concerning sub-task analysis and domain adaptation in aspect-based sentiment analysis. The dataset is used in Chapter 3, 5, and 6.
- ***Twitter*** is tweet dataset collected during Covid-19 by (Lamsal, 2021). In the past research works, Twitter data have been used in the extraction of situational awareness information relating to any crisis. To detect latest public concerns, it is used in Chapter 3.
- ***Amazon & CNET*** are the dataset including customer reviews in domains device, service, diapers, antivirus software, and electronics. They are widely used in cross-domain sentiment analysis due to the wide range of domains. To compare the performance of domain adaptation, this dataset is used in Chapter 6.

1.5.3 Evaluation Methods

To evaluate this research work, the research questions are validated by analysing the experimental results between the proposed methods and the baselines. The collected datasets are split into training and testing datasets, and experiments are conducted on the datasets. The statistical metrics (i.e., precision, recall, and F1 score) are utilised to determine the performance of the proposed methods and compared models.

The precision (P) is the number of true positive (TP) results divided by the sum of true positive (TP) and false positive (FP) results predicted by the method. The recall (R) is the number of true positive results divided by the sum of true positive and false negative results. The F1 can be calculated using precision and recall. Mathematically, they can be expressed as Equations 1.1 - 1.3:

$$P = \frac{TP}{TP + FP} \tag{1.1}$$

$$R = \frac{TP}{TP + FN} \tag{1.2}$$

$$F1 = 2 * \frac{P * R}{P + R} \tag{1.3}$$

1.6 Contributions of the Thesis

The original research contributions of this thesis can be summarised as follows.

(1) A novel feature selection-based framework is proposed to explore the most relevant features for aspect term extraction, where both BERT embeddings and relevant linguistic features are integrated. A novel feature selection method is designed by extending the Artificial Bee Colony (Karaboga, 2005) with an adaptive threshold, which can address the high sparsity and dimensionality issue of training datasets. The framework and related results are published in (J. Shi, Li, Bai & Ito, 2022).

(2) I propose a novel end-to-end model to jointly extract concerns and relations consisting of the Concern Graph (CG) and shared state of concerns. The concern graph data structure is designed to capture the inherent structural information of concerns more efficiently. The model and results are published in (J. Shi, Li, Yongchareon, Yang & Bai, 2022).

(3) A novel neural network architecture is designed to 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. 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. The model and results are published in (J. Shi, Li, Bai, Yang & Jiang, 2022).

(4) I propose a novel joint learning method for cross-domain ATE tasks. To the best of my knowledge, it is the first attempt to solve the domain adaptation problem via soft prompts. The learnable soft prompt is designed with rich external syntactic knowledge to better leverage the domain-specific and domain-invariant knowledge across domains. The related results have been published in (J. Shi, Li, Bai, Yang & Jiang, 2023).

1.7 Thesis Structures

The rest of the thesis is arranged in the following structure:

- **Chapter 2** reviews several state-of-the-art studies of aspect term extraction, opinion extraction, sentiment analysis, and domain adaption in aspect-based sentiment analysis.

- **Chapter 3** introduces a novel and effective framework for aspect term extraction by integrating both contextual and linguistic features with the Artificial Bee Colony-based feature selection method. This chapter aims to answer Research Question 1.

- **Chapter 4** presents a novel end-to-end deep learning model to identify aspect terms and the corresponding relations based on Graph Convolutional Network and Bi-directional Long Short Term Memory integrated with Concern Graph, which tackles Research Question 1.

- **Chapter 5** develops a novel unified framework to handle all defined sub-tasks for aspect-based sentiment analysis with a multi-layer semantic model and a multi-layer syntax model. This chapter focuses on Research Question 2.

- **Chapter 6** interpolates a set of transferable soft prompts into a prompt-based joint learning method for cross-domain aspect-based sentiment analysis. This chapter addresses the challenges defined in Research Question 3.

- **Chapter 7** concludes the thesis with a summary of the benefits and drawbacks of the proposed methods, along with a roadmap for future work.

Chapter 2

Literature Review

In this chapter, contemporary studies related to cross-domain aspect-based sentiment analysis have been reviewed. First, three categories of sentiment analysis are outlined. This is followed by the literature review on aspect term extraction, which plays a crucial role in aspect-based sentiment analysis. Then, the following sections explore the existing research on the unified framework and domain adaptation for aspect-based sentiment analysis. Finally, the chapter is concluded by summarising the gaps in existing studies.

2.1 Sentiment Analysis

Sentiment analysis is widely acknowledged as a crucial task within the field of Natural Language Processing (NLP). Its primary objective is to utilise text mining and linguistic analysis techniques to identify and extract subjective information contained in text data, such as social media posts and product reviews (Y. Shi, Zhu, Li, Guo & Zheng, 2019). Sentiment analysis can be categorised into three main types in terms of text scale: document level, sentence level, and aspect level.

2.1.1 Document-Based Sentiment Analysis

The document-based sentiment analysis aims to classify the sentiment polarity or opinion of a whole document as positive, negative, or neutral. It assumes that only a single opinion target exists in each document. Currently, there are many methods proposed for document-based sentiment analysis. Sharma et al. propose a lexicon-based method to determine the semantic orientation of reviews, where WordNet is used to identify synonyms and antonyms of the opinion word lists and provide a summary of the total number of positive and negative documents (Sharma, Nigam & Jain, 2014). Tang et al. adopt Convolutional Neural Network (CNN) to learn sentence representations and Gated Recurrent Neural Network (GRNN) to encode the semantics of sentences into document representations (Tang, Qin & Liu, 2015). They conduct experiments on different datasets, and experimental results show their model outperforms baselines. Zhao et al. present a domain-independent framework using weighting rules based on the Rhetorical Structure Theory (Z. Zhao, Rao & Feng, 2017). The documents are parsed into rhetorical structure trees, and then two well-known lexicons are utilised to compute the sentiment scores of sentences. Then the scores of sentences are summed up based on weighting rules to obtain document sentiment polarity. However, the document may include some opposite sentiments, which can impact the final decision. This challenge also hinders the development of document-level sentiment analysis.

2.1.2 Sentence-Based Sentiment Analysis

Sentence-level sentiment analysis involves determining the sentiment conveyed in a sentence, particularly in product reviews. This task requires classifying sentences as either subjective (expressing opinions) or objective (representing factual information). Various methods have been proposed to address the challenges associated with this task. Nguyen et al. present a rating-based feature based on the fact that the rating scores can provide useful information to improve sentiment classification (Nguyen, Vu, Pham et al., 2014). Then, they combine extracted rating features with unigram, bigram, and trigram for sentiment-level sentiment classification. Chen et al. adopt a sentence-type method to improve the performance of sentence-level sentiment analysis (T. Chen, Xu, He & Wang, 2017). A neural network-based model is utilised to classify sentences into three types: non-target, one-target, and multi-target. Then, each type of sentence is fed into CNN to predict the sentiment of sentences. The main challenge of sentence-based sentiment analysis is identifying the expressed opinions in objective sentences. Both sentence-level and document-level sentiment analysis are important and useful in real-world applications. But necessary details of opinions on all entity targets are not provided from both methods, so aspect-level sentiment analysis attracts more attention from researchers.

2.1.3 Aspect-Based Sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA) can offer valuable information to both consumers and businesses by extracting opinions from the text regarding specific entities and the corresponding aspects (B. Liu, 2012). This method enables the analysis of vast amounts of unstructured text, yielding coarse-grained or fine-grained information that may not be captured by the traditional user ratings on some review sites.

Mowlaei et al. apply statistical methods and genetic algorithms to generate two dynamic lexicons for aspect-based sentiment analysis (Mowlaei, Abadeh & Keshavarz, 2020). Without human supervision, the dynamic lexicons can be updated constantly and fused with the static lexicons to improve performance. Li et al. propose a Bidirectional Long Short-Term Memory (BiLSTM) network for sentiment analysis (W. Li, Qi, Tang & Yu, 2020). The model exploits the relationship between aspect and sentiment polarity words in a sentence without relying on the sentiment lexicons. The experimental results show the method can outperform the lexicon-based methods. Tan et al. propose a learning method to train aspect embeddings according to the relation between aspect-category and aspect terms (Tan et al., 2020). The learned embeddings are fed into deep-learning models for aspect-category sentiment analysis tasks. Wu et al. introduce two variants of context-guided BERT to distribute attention under different contexts (Z. Wu & Ong, 2021). Both models are trained with pre-trained BERT on aspect-based sentiment analysis datasets, and the models achieve new state-of-the-art results.

2.2 Aspect Term Extraction

Aspect extraction is a fundamental sub-task in the field of aspect-based sentiment analysis. The aspect refers to the phrase in a sentence where people express their opinion. The current methods can be divided into rule-based, topic-modelling-based, machine learning-based, and deep learning-based.

2.2.1 Rule-Based Model

Rule-based methods concentrate on lexical items and relationships between words, relying on expert-designed rules to uncover patterns and extract aspects that demand domain-specific expertise or human understanding (Qiu et al., 2011; Q. Liu, Liu, Zhang, Kim & Gao, 2016). A double-propagation algorithm was proposed that relied on rules generated by a dependency parser to extract aspects in (Qiu, Liu, Bu & Chen, 2009). This algorithm employed a modest set of opinion words to uncover related aspects, and then used those aspects to identify additional opinion words. The newly discovered opinion words were added to the original set of seed opinions, and the process was repeated until no further aspects or opinions were found. Zhang et al. improve the DP model by incorporating new syntactic rules. Although DP performs well on smaller datasets, its performance begins to decline as the size of the dataset increases (L. Zhang, Liu, Lim & O'Brien-Strain, 2010). To improve the performance of aspect extraction, Liu et al. designed a method by incorporating similarity-based and association-based recommendations, with the DP approach serving as the baseline (Q. Liu et al., 2016). Their model was trained using Word2Vec embeddings on a large number of reviews. The association-based recommendation, on the other hand, is applied to extract aspects by examining the relationships between aspects and opinions.

Qiu et al. utilise the linguistic rules known as Dependency Parse (DP) to extract sentiment words and opinion targets (Qiu et al., 2011). Rule-based methods do not necessitate labelled data, reducing the need for human annotations and saving effort in the process. Yu et al. leverage a list of aspects derived from pros and cons reviews to identify aspects in free-text reviews rated highly by most users (J. Yu, Zha, Wang, Wang & Chua, 2011). Rana et al. apply a lexicon-based method to extract domain-specific explicit aspect terms (Rana & Cheah, 2016). Despite the benefits of not requiring labelled data, rule-based methods necessitate the creation of rules by humans. The accuracy of these rules depends on the correctness of the sentences being analysed, which may lead to inconsistencies and inaccuracies in the analysis.

2.2.2 Topic Modelling-Based Model

Recently, topic modelling-based methods have received significant attention in academic research (Rana, Cheah & Letchmunan, 2016). These methods primarily utilise Latent Dirichlet Allocation (LDA) and scan documents to uncover relevant topics. Brody et al. treat each sentence as an individual document and utilise the conventional LDA method. The resulting topics from each sentence were viewed as aspects (Brody & Elhadad, 2010). A hybrid approach was proposed that combined maximum entropy and topic modelling to identify both aspects and opinions (X. Zhao, Jiang, Yan & Li, 2010). Two models were proposed to identify aspects and group them, assuming that all aspects within a single sentence pertain to the same topic (Jo & Oh, 2011). Bagheri et al. propose an aspect extraction model by considering each word in the sentence as a state of a Markov chain (Bagheri, Saraee & De Jong, 2014). Wang et al. propose two semi-supervised LDA to extract aspects from the online reviews (T. Wang et al., 2014). A seed list from an E-commerce website is generated, and the prior knowledge about the product is used to extract aspects.

Topic modelling-based techniques may experience reduced accuracy due to the limitations of LDA distribution in capturing correlations between topics. For example, "time" and "day" may be topical terms related to "battery life" but have no connection to the topic of "phone". This can lead to a decrease in accuracy as individual words may not be aspects, despite their relation to a topic.

2.2.3 Machine Learning-Based Model

Machine learning-based methods like Support Vector Machine (SVM), Conditional Random Field (CRF), and Maximum Entropy (ME) are used for sequential labelling tasks for aspect extraction, achieving promising results. Wang et al. propose a semi-supervised method based on the seeding aspects (B. Wang & Wang, 2008). The seed words are used to identify related aspects. Jin et al. propose a lexicalised Hidden Markov Model approach for extracting aspects and opinions from customer reviews (Jin, Ho & Srihari, 2009). Hai et al. apply bootstrapping to learn both aspect and opinion terms from Chinese customer reviews (Hai, Chang & Cong, 2012). Li et al. integrate several heuristic rules into a binary classifier to map aspect terms from the parse tree of the review (S. Li, Wang & Zhou, 2012). Liu et al. design a partially-supervised alignment model that considers identifying opinion relations as an alignment process (K. Liu et al., 2014). Merged neural networks with linguistic patterns for aspect extraction, Poria et al. propose a 7-layer deep convolutional neural network to apply linguistic patterns to categorise each word in a sentence as an aspect (Poria, Cambria & Gelbukh, 2016).

Although machine learning-based methods are widely used for aspect extraction, they have some drawbacks. For example, these methods may ignore hidden semantic features, which can negatively impact the accuracy of the extraction. Additionally, extensive human effort is required to annotate training data for these methods, which can be time-consuming and expensive.

2.2.4 Deep Learning-Based Model

Deep learning has demonstrated great success in solving complex learning problems. Nowadays, it has been widely applied to the NLP field, which yields remarkable performance (Jelodar, Wang, Orji & Huang, 2020). However, traditional deep learning models cannot deal with long texts. Recently, efforts are made to address the issue of long text input, and some recurrent neural networks are proposed. Among these models, the Long Short-Term Memory network (LSTM) is one of the most prevalent approaches to processing texts of different lengths (Hochreiter & Schmidhuber, 1997).

Recently, most researchers have focused on deep learning-based methods to capture contextual features automatically. Poria et al. combine a deep neural network with the linguistic rule-based method to tag every word as an aspect word in the sentence (Poria et al., 2016). The method has significantly improved the accuracy of aspect extraction. Wu et al. propose a hybrid method by filtering the domain-specific aspect to further improve the performance of aspect extraction (C. Wu, Wu, Wu, Yuan & Huang, 2018). Wan et al. propose a novel method for target-aspect-sentiment detection (Wan et al., 2020). The method applies the pre-trained language model to capture the dependence on both targets and aspects for sentiment prediction. However, approaches that rely on manual data annotation for aspect extraction are time-consuming and require significant financial resources. On the other hand, methods that utilise linguistic features, such as lemma, tag, dep, shape, and inherent structure, may neglect the crucial features related to the relation between aspects, leading to decreased accuracy in aspect extraction.

2.3 Unified Framework

This section reviews current studies on ABSA, including single-task and multiple sub-task approaches. It also provides an overview of the latest studies on unified models for ABSA.

According to different combinations, ABSA can be divided into seven subtasks, i.e., aspect term extraction, opinion term extraction, aspect-level sentiment classification, aspect-oriented opinion extraction, aspect term extraction and sentiment classification, aspect-opinion pair extraction, and triplet extraction (Yan et al., 2021). The existing studies of the rest sub-tasks are reviewed in this section.

2.3.1 Opinion Term Extraction

Opinion term extraction is a critical sub-task in aspect-based sentiment analysis, and the goal is to identify all the opinion words from a customer review.

Bethard et al. present an extension of semantic parsing techniques combined with extra lexical and syntactic features that can generate labels for propositional opinions rather than other syntactic components (Bethard, Yu, Thornton, Hatzivassiloglou & Jurafsky, 2004). Ku et al. design algorithms for opinion extraction at the word, sentence, and document levels (Ku, Liang & Chen, 2006). In their algorithm, opinion summarization retrieves all relevant sentences related to the main topic from a set of documents, determines the polarity of each relevant sentence's opinion, and summarises both positive and negative sentences. Jin et al propose a novel machine learning approach built under the framework of lexicalised Hidden Markov Models (HMMs). The approach naturally integrates multiple important linguistic features into automatic learning (Jin et al., 2009). Yang et al. introduce a joint inference model that leverages information from predictors that optimise subtasks in opinion extraction to achieve a globally optimal solution (B. Yang & Cardie, 2013). However, some of these frameworks have neglected the syntactical information and word dependencies in a sentence, which are crucial for ABSA tasks.

2.3.2 Aspect-Level Sentiment Classification

Aspect-level sentiment classification aims to determine the sentiment polarity towards a specific aspect in a sentence. The central research inquiry focuses on finding the optimal way to utilise the relationship between the aspect term and sentence context to classify the sentiment.

Jiang et al. aim to enhance target-dependent Twitter sentiment classification through the implementation of (1) target-specific features and (2) consideration of related tweets (L. Jiang, Yu, Zhou, Liu & Zhao, 2011). Tang et al. develop two target-dependent long short-term memory models that automatically incorporate target information to significantly improve classification accuracy (Tang, Qin, Feng & Liu, 2016). Li et al. present a new approach for target-focused sentiment analysis using a combination of a bi-directional Recurrent Neural Network (RNN) and a CNN layer. The RNN layer transforms word representations, while the CNN layer extracts important features from these representations. (X. Li, Bing, Lam & Shi, 2018). In (C. Zhang, Li & Song, 2019), an aspect-specific sentiment classification model uses a Graph Convolutional Network (GCN) layer applied to the dependency tree of a sentence. Nevertheless, further research is needed to explore more effective and efficient approaches for aspect-level sentiment classification.

2.3.3 Aspect-Oriented Opinion Extraction

Aspect-oriented opinion extraction has gained significant attention in recent years as a key task in ABSA. The goal of aspect-oriented opinion term extraction is to identify expressions of opinion related to a specific aspect of ABSA.

Fan et al. propose a target-oriented opinion words extraction model, based on a target-fused neural network using an Inward-Outward LSTM architecture (Fan, Wu, Dai, Huang & Chen, 2019). A transferable network is proposed for fine-grained opinion term extraction (W. Wang & Pan, 2019b). Their network is capable of utilising both local and global memory interactions to uncover the correlations between aspect or opinion words. Wu et al. use an attention layer to strengthen the link between aspect and opinion words in the process of aspect-oriented opinion extraction (Z. Wu, Ying et al., 2020). Veyseh et al. utilise syntactic structures, such as the dependency tree-based distance to the aspect, to aid in the identification of opinion terms (Veyseh, Nouri, Dernoncourt, Dou & Nguyen, 2020). Mensah et al. conduct an experimental study to evaluate the significance of positional embeddings using various text encoders (Mensah, Sun & Aletras, 2021). Their findings suggest that BiLSTM-based methods possess a suitable inductive bias for opinion extraction tasks. Moreover, incorporating a GCN to consider the structural information only led to minor performance improvements.

While these methods have demonstrated promising results, there is still room for improvement, particularly in handling complex and implicit aspect expressions, and enhancing the overall accuracy of sentiment classification. Further research is required to address these challenges and improve the performance of aspect-oriented opinion extraction in ABSA.

2.3.4 Aspect Term Extraction and Sentiment Classification

Aspect term extraction and sentiment classification are two fundamental tasks in ABSA. The goal is to identify and classify the sentiment expressed toward specific aspects of a given text.

Mitchell et al. explore the simultaneous labelling of target and sentiment classes by utilising the Conditional Random Fields (CRF) technique along with conventional discrete features. The researchers introduced three models: pipeline, joint, and collapsed-based on various methods of labelling the two tasks. Their findings indicate that the pipeline model outperforms the joint model on the tweet dataset (Mitchell, Aguilar, Wilson & Van Durme, 2013). Additionally, Zhang et al. introduce word embedding representations into the CRF framework and discover that incorporating word embeddings into handcrafted features improved the performance, regardless of the pipeline, joint, or collapsed methods used (M. Zhang, Zhang & Vo, 2015). Ma et al. develop a hierarchical stack bidirectional gated recurrent units (HSBi-GRU) model to capture abstract features for both target and sentiment labelling tasks. They propose a joint model based on HSBi-GRU that considers the impact of the target label on the sentiment label (D. Ma, Li & Wang, 2018). Hu et al. propose a span-based framework of extract-then-classify that leverages the supervision of aspect span boundaries to identify multiple aspect terms, followed by the classification of sentiments based on the aspect span representations (M. Hu, Peng, Huang, Li & Lv, 2019).

2.3.5 Aspect-Opinion Pair Extraction

Extracting aspect terms and opinion words from social media is a crucial sub-task in aspect-based sentiment analysis in NLP, aiming to identify aspect terms and the corresponding opinion words in review sentences. The current research in aspect-opinion pair extraction can be divided into three categories: (1) rule-based, (2) feature engineering-based, and (3) deep learning-based.

For rule-based methods, aspect and opinion words are extracted using syntactic rules based on the sentence's dependency structure (M. Hu & Liu, 2004a, 2004b; Qiu et al., 2011). Defining rules for identifying various aspect and opinion expressions can be challenging. In response to the limitations of rule-based methods, feature engineering-based approaches have become increasingly popular (Jin & Ho, 2009; F. Li et al., 2010). However, the feature engineering-based methods is heavy human effort required to create features based on the syntax of a sentence. To minimise this effort, deep learning-based models have been introduced (P. Liu, Joty & Meng, 2015; Y. Yin et al., 2016; W. Wang, Pan, Dahlmeier & Xiao, 2016; S. Chen, Wang, Liu & Wang, 2021; K. Liu et al., 2014; W. Wang, Pan, Dahlmeier & Xiao, 2017; H. Dai & Song, 2019). Treating the extraction of aspect-opinion pair from a perspective of joint term, Zhao et al. propose a multi-task learning framework to extract the pairs by identifying the relations between aspect and opinion using span representations (H. Zhao, Huang, Zhang, Lu & Xue, 2020). Chen et al. propose a synchronous double-channel recurrent network for the extraction of aspect-opinion pair by exploring the relations between aspect and opinion terms (S. Chen et al., 2020).

2.3.6 Triplet Extraction

Aspect sentiment triplet extraction is a crucial component of ABSA that involves identifying the target aspect, its associated sentiment, and the opinion words. This information is represented as a triplet and provides a comprehensive understanding of the sentiment expressed in a given review.

In order to identify all elements of aspect-based sentiment analysis in a single step, aspect sentiment triplet extraction is introduced with a two-stage framework based on LSTM and GCN networks (Peng et al., 2020). Chen et al. formalise the

Aspect Sentiment Triplet Extraction (ASTE) task as a Multi-Turn Machine Reading Comprehension (MTMRC) task (S. Chen et al., 2021). This formaliSation enables the identification of aspect sentiment triplets in a unified framework. Xu et al. introduce the initial end-to-end model with a novel position-sensitive tagging approach that can jointly extract triplets (L. Xu, Li, Lu & Bing, 2020). The joint model can extract triplets using a sequence tagging approach that can capture the rich interactions among the elements. Existing research mainly approaches this issue through a multi-stage pipeline, disregarding the interdependence between the three elements and suffering from error propagation. To address these problems, Chen et al. present a semantic and syntactic enhanced aspect sentiment triplet extraction model to fully leverage the semantic and syntactic relationships between the triplet elements and extract them jointly.

2.3.7 Unified Models

Most research has only targeted subsets of ABSA subtasks, resulting in complex ABSA models. The unified models can address all the ABSA tasks without relying on the sub-models or changing the model structure to fit all ABSA subtasks.

Mao et al. address all sub-tasks by setting up two machine reading comprehension problems, with BERT serving as the backbone network (Mao, Shen, Yu & Cai, 2021). Yan et al. redefine each subtask target as a sequence that combines pointer indexes and sentiment class indexes, thereby converting all ABSA subtasks into a unified generative formulation (Yan et al., 2021). Using this unified formulation, they leveraged the pre-trained sequence-to-sequence model BART to solve all ABSA subtasks within a single end-to-end framework. Despite these existing unified frameworks, they ignore the syntactical information and word dependencies in a sentence, which have been proven to be crucial information for aspect-based sentiment analysis tasks (J. Dai, Yan, Sun, Liu & Qiu, 2021).

Despite being based on pre-trained language models that can capture implicit syntactic information from sentences, these unified frameworks are still limited by the absence of explicit syntactical features to improve the performance of ABSA tasks.

2.4 Domain Adaptation

Cross-domain aspect-based sentiment analysis has gained popularity in NLP in recent years. This approach involves applying learned knowledge from one domain to another domain for sentiment analysis. Traditional methods for aspect-based sentiment analysis rely heavily on a large-scale human-annotated dataset, which may not be available in real-world applications. Cross-domain aspect-based sentiment analysis has been proposed as a solution to address this issue.

Aspect Term Extraction (ATE) is a detailed sentiment analysis task that has gained significant attention. Despite this, only a few studies have attempted to focus on domain adaptation for ATE. The objective of Cross Domain ATE is to apply the knowledge gained from a source domain to a target domain where ATE label data is limited. This task is challenging due to its complexity and the shortage of labelled data in target domains. Current methods for solving this task can be classified into three categories: neural network-based models, fine-tuning Pre-trained Language Models (PLMs), and prompt-tuning PLMs.

2.4.1 Neural Network-based Model

Initial studies on Cross Domain ATE primarily focused on using hand-made domain-agnostic features and neural network models (Y. Ding, Yu & Jiang, 2017). Jakob et al. formulate the ATE problem as an information extraction task and present a Conditional Random Field (CRF) method for single-domain and cross-domain ATE (Jakob & Gurevych, 2010). Chernyshevich and Belarus develop a CRF-based system trained on mixed annotated data to detect aspect terms on all domain-specific test sets (Chernyshevich & Belarus, 2014). However, CRF-based methods are not effective when the training and test datasets are from different domains. To address this, Ding et al. propose a LSTM network method that leverages domain-independent syntactic rules (Y. Ding et al., 2017). To bridge the gap between different domains in aspect-based sentiment analysis, Wang et al. employ domain-invariant dependency relations as pivot information and propose a novel Recursive Neural Network (W. Wang & Pan, 2018). In their subsequent research, Wang et al. further integrated word representations and syntactic head relations into a Conditional Domain Adversarial Network (W. Wang & Pan, 2019a). Moreover, to capture the intra-correlations between aspect and opinion terms and between terms themselves, Wang et al. introduced an Interactive Memory Network, which incorporated auxiliary tasks and domain adversarial networks to align source and target spaces (W. Wang & Pan, 2019b). Marcacini et al. present a transductive learning method combining features of labelled aspect terms, unlabelled aspect terms, and linguistic information from both source and target domains (Marcacini, Rossi, Matsuno & Rezende, 2018). Li et al. propose an adversarial learning method to align word correlations by learning an alignment weight for each word (Z. Li et al., 2019). Despite the good performance, neural network-based methods often lack the quality of domain-invariant features and fail to fully exploit supervision signals in the target domains, leading to low precision results.

2.4.2 Fine-Tuning Pre-trained Language Models

Research has recently shown that fine-tuning language models with task-specific layers can improve performance for cross-domain aspect term extraction (ATE). Hewitt et al. find that these models can obtain word sense and geometrical dependency parse relations, which benefit the ATE task (Hewitt & Manning, 2019). Pereg et al. leverage the intrinsic knowledge of language models with external syntactic features by incorporating external linguistic information into the model through a self-attention mechanism for cross-domain ATE (Pereg et al., 2020). Gong et al. propose an end-to-end framework based on BERT that integrates feature-based adaptation and instance-based adaptation, resulting in improved performance of the language model for ATE (Gong, Yu & Xia, 2020). Anand et al. apply an evolutionary approach to learn linguistic patterns of aspect words automatically, avoiding the need for manual pattern engineering (Anand & Mampilli, 2021). Mampilli et al. combine language models with attention mechanisms for ATE and achieved promising results on both in-domain and unseen-domain datasets (Mampilli & Anand, 2022). Li et al. propose a generative cross-domain data augmentation framework that uses annotated data from the source domain to generate data for the target domain for ATE model training (J. Li, Yu & Xia, 2022). Howard et al. introduce a method to automatically construct domain-specific knowledge graphs of aspect terms and inject these features into language models for ATE in target domains (Howard et al., 2022). Klein et al. transfer learned knowledge from language models by utilising syntactic relations connecting opinions and related aspect words (Klein et al., 2022). Dong et al. propose a syntax-based BERT to capture domain-invariant features for aspect term and sentiment transfer (Dong et al., 2022). However, these language model-based methods rely heavily on annotated resources, and fine-tuning may be unstable with small-scale data. Additionally, most methods only integrate linguistic features directly into the language models, lacking word-level adaptation for aspect extraction.

2.4.3 Prompt-Tuning Pre-trained Language Models

Prompt-based methods have been suggested to overcome the challenges posed by Larger Language Models (LLMs) in learning. These methods use language prompts and task descriptions as context to make ABSA similar to language modelling. Early studies focused on hard templates manually defined for ABSA tasks in a single domain. Li et al. are the first to introduce a prompt-based model for ABSA subtasks, where sentiment knowledge prompts were constructed by combining features from aspects, opinions, and polarities (C. Li et al., 2021). Gao et al. develop a unified generative framework for different ABSA tasks, which uses task prompt to control the type of task (T. Gao et al., 2022). Their method can transfer learned knowledge to difficult tasks by assembling prompts from simple tasks. Li et al. propose a teacher-student network with a prompt-based approach to address over-fitting in basic prompt-based models (H. Li et al., 2022). Ben et al. present an example-based prompt learning method that can be applied to multiple tasks in unseen domains, including rumour detection, multi-genre natural language inference, and aspect prediction (Ben-David et al., 2022). However, designing a prompt requires domain knowledge, so soft prompts have been proposed as a solution, allowing LMs to perform specific tasks using learnable vectors instead of human-interpretable natural language.

Wu et al. use soft prompts instead of fixed, predefined templates to learn representations for different domains, and employ a novel domain adversarial training mechanism to learn domain-invariant features for sentiment classification tasks (H. Wu & Shi, 2022). Asai et al. present a multi-task language model tuning method that transfers knowledge across tasks using soft prompts (Asai et al., 2022). This method is highly efficient in terms of parameters and delivers promising results using knowledge from high-resource datasets for sentiment classification and other NLP tasks. However, existing hard and soft prompt-based methods either focus on a single domain or are limited to sentiment classification rather than aspect term extraction.

2.5 Summary

This chapter provides a comprehensive overview of related works in the field of aspect-based sentiment analysis, aspect term extraction, a unified framework for aspect-based sentiment analysis, and domain adaptation.

The gaps identified in the literature review are summarised below by examining the advantages and disadvantages of these studies.

- Most previous studies only use machine learning models and do not consider the contextual features from pre-trained models. Additionally, existing works either overlook the valuable information provided by the local context or ignore other linguistic features, negatively impacting the performance of aspect term extraction.

- For aspect term extraction, only the BERT embedding of tweets is utilised, which fails to incorporate regional word dependencies from tweets, limiting the performance of aspect term extraction.

- In relation extraction of aspect terms, the relationship between aspect terms within a single tweet posted by a user goes undetected, missing critical information for revealing meaningful insights.

- Most existing studies focus solely on a specific sub-task in ABSA, leading to the development of complex ABSA models that limit the practical applications of these proposed models.

- The current unified frameworks for aspect-based sentiment analysis overlook syntactic features. Most existing studies concentrate on learning semantic representations, neglecting the syntactic features that can significantly enhance aspect and opinion term extraction and improve performance for ABSA sub-tasks.

- In many current studies, the linguistic part-of-speech (POS) and syntactic dependency label, two key syntactic features, are often overlooked but can greatly enhance model performance.
- When pre-training PLMs for fine-tuning methods, using universal datasets instead of specific domains, the resulting models may exhibit task-agnostic behavior and subpar performance in domain adaptation.
- Prompt-based learning methods face the challenge of high cost in enumerating all possible spans of aspect terms. Additionally, existing models struggle to deliver robust performance on cross-domain datasets, owing to the varying distributions of aspect terms across domains and the difficulty of constructing prompts.

The literature review has highlighted the aforementioned research gaps in aspect term extraction, aspect relation extraction, a unified framework for ABSA, and domain adaptation of ABSA. This also highlights the need for further research in this field. The solutions proposed in this thesis aim to address these gaps by presenting corresponding models.

Chapter 3

Aspect-Based Sentiment Analysis with Artificial Bee Colony-based Feature Selection

Aspect terms are opinion targets for people to express and understand opinions in reviews. Aspect terms extraction is an essential subtask in aspect-level sentiment analysis. To extract aspect terms from a sentence, existing methods mainly focus on context features generated by pre-trained models. However, these models either neglect the crucial implicit linguistic features, e.g., post-of-tag, head, and head dependency, or fail to explore sufficient valuable features for aspect term extraction, which lead to the deficiency in the aspect term extraction task.

To address the challenges, in this chapter, I propose a novel and effective framework for aspect term extraction by integrating both contextual and linguistic features with the Artificial Bee Colony-based feature selection method. Firstly, a novel variant of Artificial Bee Colony is designed to identify the most valuable linguistic features to reduce the high sparsity and dimensionality of the raw dataset. Next, the selected features and context embeddings are integrated to improve the performance of aspect extraction. Finally, extensive experiments are conducted on real-world datasets, and the results exhibit that the proposed framework can outperform the competitive baselines. Compared with the latest baselines, the proposed framework achieves comparatively higher F1 scores of 80.7%, 84.7%, 72.2%, and 74.8% on the four groups of datasets. Furthermore, the ablation study shows that the proposed method with the designed feature selection module significantly outperforms the original Artificial Bee Colony, having 4.15%, 4.4%, 4.4%, and 3.2% improvements in F1 scores on all four datasets, respectively.

3.1 Overview

Nowadays, people tend to express their opinion or emotions toward a product, service, or event on social media platforms (Zhuang et al., 2006; Riquelme & González-Cantergiani, 2016; Manek, Shenoy, Mohan & Venugopal, 2017; Kumar & Harish, 2018; J. Shi et al., 2021). The analysis of public opinion targets has been attracting great attention from both researchers and practitioners (Jin & Ho, 2009; Jakob & Gurevych, 2010; F. Li et al., 2010). In the world of e-commerce, most users tend to seek opinions by viewing a large number of reviews or feedback online before purchasing any products or services. Besides, most e-commerce companies rely on the analysis of customers' reviews for refining their business decisions to improve the quality and gain insights into their products or services. However, manually reading through each review turns out to be unrealistic since a remarkable volume of reviews is generated rapidly. Therefore, it is essential to assist users in identifying the desired information from numerous reviews by applying opinion mining or sentiment analysis.

For the traditional sentiment analysis methods based on coarse level (document or sentence), they are not able to satisfy the users' needs because finer information is required from the product or service reviews in terms of aspects (Turney, 2002; S.-M. Kim & Hovy, 2004; Mukherjee & Liu, 2012). For example, most graphic designers only focus on the display screen and colour accuracy when they seek a laptop. However, different customers may express different opinions on each of these aspects. Therefore, it is necessary and important to analyse the reviews at the aspect level. Aspect term extraction is recognised as an important task of aspect-level sentiment analysis. It aims to detect opinion targets, referred to as the product's or service's attributes, from opinion reviews (M. Hu & Liu, 2004a). It is also an essential step for fine-grained opinion mining. In recent years, deep learning approaches have been adopted for aspect term extraction because of their outstanding performance, where context-dependent embedding models are utilised, e.g., Word2Vec (Mikolov, Yih & Zweig, 2013), GloVe (Pennington, Socher & Manning, 2014), and BERT (Devlin et al., 2019a). Apart from the contextual features from the pre-trained models, feature engineering plays an important role in identifying highly relevant features for aspect term extraction.

Although many existing feature selection methods like Entropy, Information Gain, Chi-Square, and Mutual Information are employed in aspect term extraction, they are only incorporated with machine learning models and overlook the contextual features from pre-trained models (Kumar & Harish, 2018). Meanwhile, the local context is capable of disambiguating word meanings by providing rich surrounding information, which has been identified as a crucial source of features for aspect term extraction. However, most existing studies either ignore the rich information delivered by the local context or disregard other linguistic features, which hinder the performance of aspect term extraction.

In this chapter, a novel and effective framework is proposed to address the aforementioned issues by incorporating contextual features and other linguistic features to detect aspect terms. The proposed approach involves four major steps. **First**, I define a set of linguistic features associated with aspect terms and employ the proposed Feature Selection Artificial Bee Colony (FS-ABC) to identify the most relevant features. Compared with other methods, ABC has a wider search scope and fewer control parameters, which is more suitable for search-based tasks, e.g., feature selection. **Second**, I construct new fused vectors by incorporating the selected features and embeddings obtained from BERT. **Third**, the fused vectors are fed into Bidirectional Long Short Term Memory (BiLSTM), and the output hidden states are used as input of the Conditional Random Field (CRF) (Lafferty, McCallum & Pereira, 2001) layer. **Fourth**, extensive experiments are conducted to evaluate the proposed framework by using real-world datasets. The experimental results reveal that the proposed framework can outperform the existing models. In addition, an ablation study is conducted to validate the effectiveness of the selected features. To the best of my knowledge, this is the first research work studying aspect term extraction by integrating contextual representations with selected linguistic features by the proposed Artificial Bee Colony. To sum up, the contributions of this chapter are listed as follows.

- A novel feature selection-based framework is proposed to explore the most relevant features for aspect term extraction, where both BERT embeddings and relevant linguistic features are integrated.

- A novel feature selection method is designed by extending the Artificial Bee Colony (Karaboga, 2005) with an adaptive threshold, which can address the high sparsity and dimensionality issue of training datasets.

- Extensive experiments are conducted on real-world datasets to demonstrate the effectiveness of the proposed framework and explicitly show the selected implicit features can improve the performance of aspect term extraction.

The remainder of this chapter is organised as follows. Section 3.2 describes related works on aspect term extraction methods using machine learning, deep learning algorithms, and feature selection techniques. The problem formulation and the definition of linguistic features are presented in Section 3.3. In Section 3.4, the proposed framework of aspect term extraction is explained, and the proposed feature selection method is also introduced. Section 3.5 demonstrates the experimental results and analysis on SemEval datasets and introduces the ablation study. Finally, in Section 3.6, I highlight the major contributions of this chapter, discuss the limitations of the proposed method, and conclude the remarks and directions for future work.

3.2 Related Works

3.2.1 Aspect Term Extraction

The conventional approaches of aspect term extraction mainly focus on rule-based methods (Popescu & Etzioni, 2005; Y. Wu, Zhang, Huang & Wu, 2009) and hand-crafted features-based methods (Z. Chen, Mukherjee & Liu, 2014). With the remarkable performance improvement, machine learning algorithms have become mainstream for aspect term extraction (Breiman, 2001; Platt et al., 1999). Yin et al. design the positional dependency-based word embedding to apply both dependency context and positional context to aspect term extraction (Y. Yin, Wang & Zhang, 2020). A new topic modelling-based method is proposed for aspect term extraction by integrating a novel adaptation of the Latent Dirichlet Allocation (LDA) algorithm (Ozyurt & Akcayol, 2021). However, these methods require great human efforts in defining rules and annotating data.

In recent years, deep learning techniques have been widely adopted in sentiment analysis tasks since they are capable of fusing text features to extract new representations through multiple hidden layers. For example, Liu et al. propose an RNN-based model to identify opinion targets by using word embeddings without hand-crafted features

(P. Liu et al., 2015), where the experimental results demonstrate that RNN-based models can outperform the feature-rich models based on CRF. A novel unified framework is proposed to jointly extract aspect and opinion terms by integrating recursive neural networks and CRF (W. Wang et al., 2016). Soujanya et al. present a Convolutional Neural Network (CNN) based model to extract aspects (Poria et al., 2016), where the pre-trained word embeddings, i.e., Word2Vec (Mikolov, Yih & Zweig, 2013), are employed along with Part-of-Speech (PoS) tag features. Hoang et al. show the potential of utilising BERT to generate contextual word representations, having additional generated text to detect aspect categories (Hoang, Bihorac & Rouces, 2019). Liao et al. propose a novel unsupervised model to capture global and local representation for aspect extraction (Liao et al., 2019). A joint model is presented to integrate the aspect term extraction and aspect categories detection tasks into a multi-task learning framework (Wei et al., 2021). In each task, multi-layer Convolutional Neural Networks (CNNs) are applied to compute high-level word representations. A task-specific and task-share vector is produced. With a guided Latent Dirichlet Allocation (LDA), an unsupervised approach is proposed for aspect term extraction (Venugopalan & Gupta, 2022). The model is enhanced by guiding inputs using linguistic rules and multiple pruning strategies with a BERT-based semantic filter.

Deep learning-based methods employ contextual representations with light human efforts and outperform machine learning-based models. Whereas, such methods disregard other linguistic features, e.g., lemma, tag, dep, and shape. Specifically, words with completely different spellings may have almost the same meanings. Meanwhile, a set of words in different orders can present completely different meanings. Therefore, it is important to employ linguistic knowledge to obtain meaningful information, rather than totally depending on the pre-trained embeddings. In this chapter, I propose an FS-ABC method to select the most relevant linguistic features. Along with the word embeddings from BERT, the proposed approach can mitigate the issues related to missing effective features.

3.2.2 Feature Selection

With the growing dimensions of datasets in the fields of data mining and deep learning, high-dimensional data analysis has become increasingly challenging. To alleviate the problem, feature selection is recognised as a practical pre-processing mechanism for pruning irrelevant and redundant features. Xue et al. propose a self-adaptive particle swarm optimisation method to solve large-scale feature selection problems (Y. Xue, Xue & Zhang, 2019). A novel hyper-learning binary dragonfly algorithm is proposed to detect an optimal subset of features for a given classification problem (Too & Mirjalili, 2021). Relying on XGBoost, a novel framework for feature selection is presented to select sets of informative features in classification problems (Alsahaf, Petkov, Shenoy & Azzopardi, 2022). Most existing research works on aspect term extraction neglect linguistic feature selection. For example, Savoy proposes a new feature selection method for sentiment analysis by combining Z-score and Information Gain (Savoy, 2012). Koncz et al. propose a computationally efficient feature selection method based on document frequency (Koncz & Paralic, 2011). Akhtar et al. develop a Particle Swarm Optimization (PSO) based feature selection technique (Akhtar, Gupta, Ekbal & Bhattacharyya, 2017), which leverages cascade machine learning algorithms for aspect term extraction and sentiment analysis. However, these studies only apply the selected linguistic features to machine learning models but ignore the contextual representations. To address the challenges mentioned above, I retain both contextual embeddings and linguistic features as input of BiLSTM to yield better performance than the baselines. I design the feature set on datasets SemEval 2014 (Pontiki et al., 2014), SemEval2015 (Pontiki et al., 2015), and SemEval2016 (Pontiki et al., 2016), i.e., three groups of public datasets for aspect-based sentiment analysis, where aspect terms are annotated by researchers. The used linguistic features consist of lexical and syntactic information, which are generic in nature and domain-independent, with the consideration of being applied to similar nature applications.

3.2.3 Artificial Bee Colony

Artificial Bee Colony (ABC) is a representative optimisation algorithm based on swarm intelligence (Karaboga, 2005). Recently, it still attracted a lot of attention due to its simple structure and strong exploration ability in scientific research. By adopting and maintaining the history of the previously abandoned and the global best solutions, an enhanced ABC is proposed to solve the trade-off and achieve a balance of exploration and exploitation (Shunmugapriya, Kanmani, Supraja, Saranya et al., 2013). Zorarpacı et al. propose a new hybrid model for feature selection by combining ABC with different evolution methods (Zorarpacı & Özel, 2016). The method aims to solve the problem of dimensionality, which affects the quality of the training process in machine learning tasks. Applying ABC for feature selection and parameter optimisation, Kuo et al. combine a C5 decision tree (DT) and Support Vector Machine (SVM) to extract comprehensible rules from SVMs (Kuo, Huang, Zulvia & Liao, 2018). The proposed algorithm can address two problems caused by DT and SVM: (1) Lack of explanatory ability; (2) Increased computational cost due to high-dimensional data. Li et al. propose a hybrid feature selection algorithm based on ABC to automatically identify early Parkinson's disease (H. Li et al., 2021). The method can eliminate most of the useless or noisy features and determine the optimal features, achieving better performance of classification. Zhang et al. hypothesise that each sense of a word can be represented by one or more specific dimensions, then propose attention-based word embeddings using ABC for aspect-level sentiment classification (M. Zhang, Palade, Wang & Ji, 2021). ABC is mainly used to perform feature selection for algorithm optimisation. XGBoost algorithm is one of the latest and highly successful machine learning algorithms in data science. It generally yields better performance compared to the traditional Random Forest or Neural Network models due to its sparsity awareness (T. Chen et al., 2015). As a fast and scalable tree-boosting model, XGBoost is a practical base learner for calculating feature scores. It is more predictive than traditional boosting models in high-dimensional datasets with the importance of scores to reflect more complex interactions (Alsahaf et al., 2022). To the best of my knowledge, the proposed framework is the first to integrate XGBoost with the improved ABC for aspect term extraction.

3.3 Preliminaries

In this section, I formulate the problem of aspect term extraction and define the relevant linguistic features used in the proposed framework. Then, the overall framework is presented in detail. Finally, the proposed feature selection method is elaborated.

3.3.1 Problem Formulation

Given a sentence, denoted by $S = \{w_1, w_2, ..., w_N\}$, where N is the number of words, I define a set of linguistic features $F = \{f_1, f_2, ..., f_M\}$ containing M features. The objective of feature selection is to choose the most relevant feature set $F_{best} = \{f_{d1}, f_{d2}, ..., f_{dL}\}$, where d indicates the d^{th} dataset and L represents the total number of selected features. Then, the aspect term extraction can be formulated as a sequence tagging task, which aims to learn the mapping $S \to Y$ by using F_{best}, where $Y = \{y_1, y_2, ..., y_N\}$ denotes the tags of sentence S. For each tag, it is encoded into the format of $\{B, I, O\}$, representing $Beginning\ of$, $Inside\ of$, and $Outside\ of$ an aspect, respectively.

3.3.2 Linguistic Features

In this section, I describe the linguistic features used for aspect term extraction. Most of the lexical and syntactic features are domain-independent and easy to transfer to other tasks.

Inspired by the work (Akhtar et al., 2017), I design 102 features in total for SemEval datasets (Section 3.5.1), and they can be divided into 14 categories:

- **Word and Local Context**: current word, local context [-5, ...,5][1] and their lower cases are used as features;
- **POS, Dependency, and Tag**: I use Part-of-Speech (POS), dependency, and tag of the current word and local context [-2, ..., 2] as features;
- **Character n-gram**: character 2-gram, 3-gram, and 4-gram of the current word are extracted as features;
- **Head Word, POS, and DEP**: the head word of the current word and its POS and dependency are used as features;
- **Prefix and Suffix**: the fixed-length prefix and suffix of the current word and local context [-3, ..., 3] are trimmed as features;
- **Frequent Aspect**: I construct a list of frequently occurring aspect terms and use a binary value as a feature to indicate if the current word is in this list;
- **Start with Digit**: a binary feature showing if the current word starts with a digit;
- **Orthographic**: a binary feature indicating if the current word starts with a capital letter;

[1]The context window size is 10, i.e. 5 words to the left and 5 words to the right. The below context index has the same meaning.

- **NER**: named entity features of current word and context [-2, ..., 2] are extracted using Spacy[2];
- **Length**: the length of the current word and context [-2, ..., 2] are used as features;
- **Pair of Pre-POS and POS, and Pair of POS and Next-POS**: the features include the pair of the previous word and current word and the pair of the current word and next word;
- **Similarity**: I apply GloVe to generate lexicon expansion based on similarity. I obtain the top 3 similar words in GloVe of current words and context [-3, ..., 3] as features;
- **Semantic Orientation Score**: semantic orientation (SO) score measures sentiment polarity expressed in a phrase (Hatzivassiloglou & McKeown, 1997). I calculate the SO score of the current word and context [-2, ..., 2] as features;
- **Lemma, Shape, Alpha, and Stop Word**: the current word's base form and shape are two features. The binary value shows if the word is an alpha character or a stop word.

3.4 Artificial Bee Colony-based Aspect Term Extraction

I integrate pre-trained embeddings with selected linguistic features and apply BiLSTM and CRF for aspect term extraction. The overall architecture of the framework is shown in Figure 3.1. The proposed framework consists of three core modules: (1) **BERT encoder** to encode the input sequence and generate the context representations. (2) **Artificial Bee Colony-based Feature Extractor** to select the most valuable linguistic

[2]https://spacy.io/

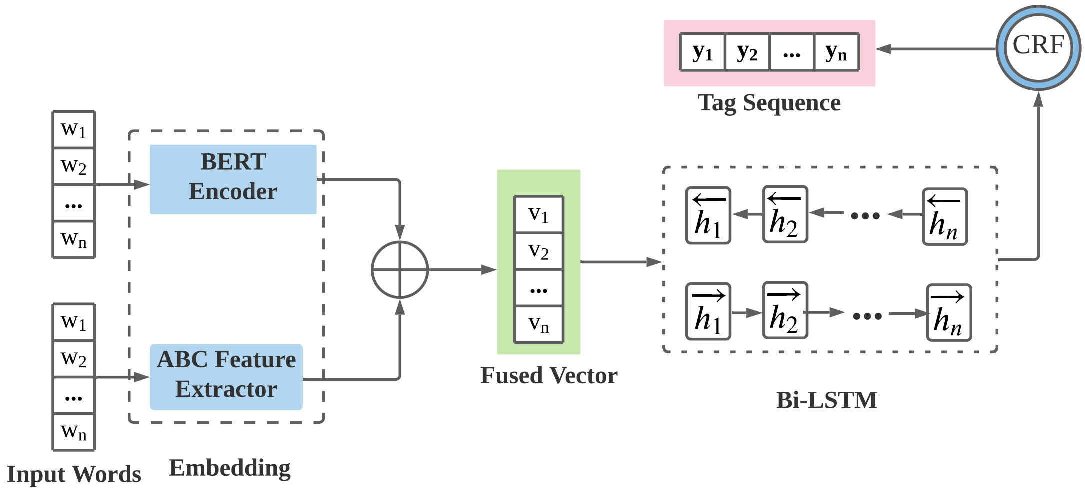


Figure 3.1: The overview of the BeeAE framework.

features by the FS-ABC. (3) **Aspect Term Extraction** to fuse the context and linguistic feature representations and predict the aspect term using a CRF layer.

3.4.1 BERT Encoder

Inspired by the success of BERT (Devlin et al., 2019a), I employ the BERT base model for pre-trained embeddings to encode the original word sequence and convert each token into context embedding. For a sentence, it is tokenised using the WordPiece vocabulary (Y. Wu et al., 2016). Two special tokens [CLS] and [SEP] are added to the beginning and the end of the tokenised sentence, respectively. Given a sentence $\{w_1, w_2, ..., w_n\}$, the input sequence $T = \{t_1, t_2, ..., t_M\}$ is encoded with M tokens after tokenisation. Next, the initial embedding e_i of each token t_i is obtained by summing its token embedding e_i^w, position embedding e_i^p, and segment embedding e_i^s. Finally, the embedding of input sequence $E = \{e_1, e_2, ..., e_M\}$ is fed into the BERT encoder, and the final output representation of each token in a sequence can be obtained using Equation 3.1. Because BERT uses WordPiece tokeniser to generate word tokens, some words may break into several tokens. To detect aspect terms in one word instead of sub-word pieces, the corresponding representations of sub-word tokens are averaged to

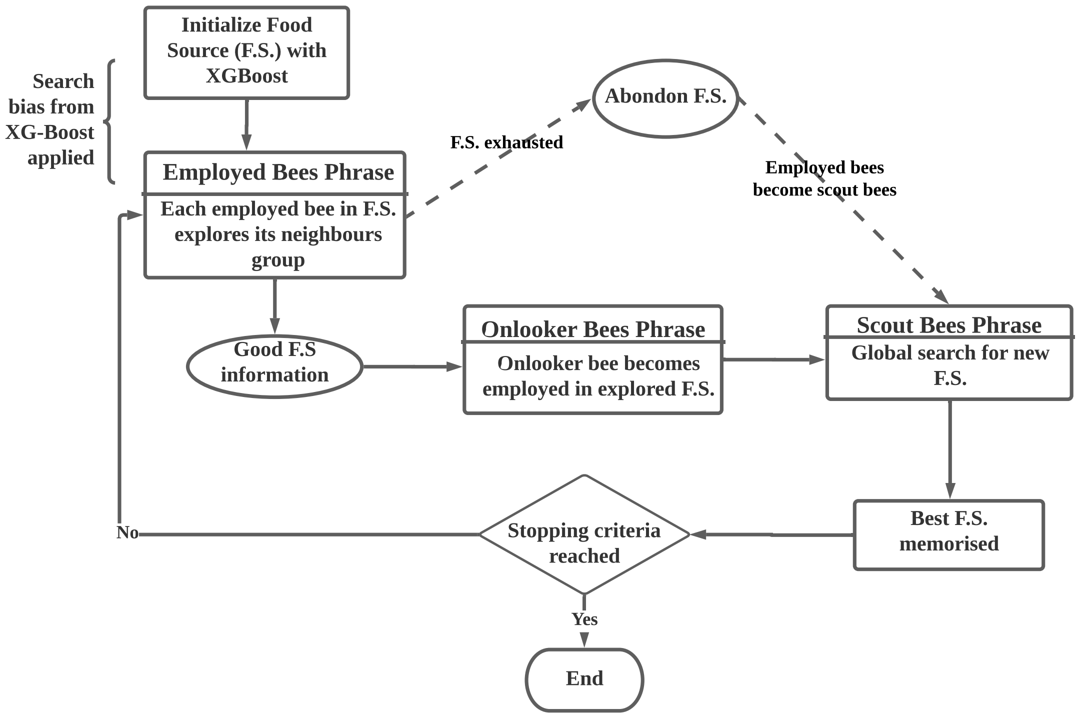


Figure 3.2: The flowchart of the FS-ABC.

get one aspect term representation. For example, the representation of the aspect term "hardware" is the average of representations of two tokens "hard" and "##ware".

$$h_i^{(bert)}(w_i) = BERT(e_i) \tag{3.1}$$

3.4.2 Artificial Bee Colony-based Feature Extractor

Inspired by the global optimisation of bees (Rao et al., 2019), I develop a novel feature selection method by extending ABC, which is a heuristic algorithm aiming at optimising numerical problems. Because of its strong ability and wide research range, the ABC algorithm is more suitable for feature selection than other biological heuristic models. The general structure of the proposed FS-ABC is shown in Figure 3.2, which involves five stages:

- **Stage I:** the initial food sources are not randomly selected as in the original ABC algorithm but guided by the highest-ranked features using the information gain values produced by the XGBoost algorithm. With the feature ranking values from XGB, I extracted a search bias for the proposed FS-ABC method, giving bias to the higher-ranked features by XG Boost. The search bias is applied when new food sources are created in this stage.

- **Stage II:** the bees explore their neighbours' groups to search for new food sources and evaluate their fitness. If the new food source is produced, the employed bees share the food source information with onlooker bees.

- **Stage III:** onlooker bees choose food sources guided by the feature score provided by XGB and the quality of the food source is calculated. The employed bees become scout bees if their solutions cannot be improved after predetermined trials, and their solutions are abandoned.

- **Stage IV:** the poor food source identified through exploration are abandoned and scout bees start to search for new solutions randomly.

- **Stage V:** the best food source is memorised which has the highest quality and the searching behaviour will be terminated if a stopping criterion is satisfied. Otherwise, repeat the five stages.

XGBoost is a scalable end-to-end tree boosting model, which consists of an ensemble of classification and regression trees (CART). It has been widely used in many Natural Language Processing (NLP) tasks (Lai, Liu & Lien, 2021; Azhar, Khodra & Sutiono, 2019) because of its advantages compared to other gradient boosting frameworks, such as addressing the over-fitting, supporting the parallelisation of tree construction, and speeding up the execution (J. Ma et al., 2020). Given data input $X = \{x_1, x_2, ..., x_i, ..., x_n \| x_i \in \mathbb{R}^{FN}\}$, where FN is the feature number, the output is predicted with the collection of decision trees in Equation 3.2.

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i), \tag{3.2}$$

where T represents the number of trees, and f_t indicates an independent tree structure with leaf scores.

Then the regularised objective introduced in XGBoost is given by Equation 3.3.

$$\zeta^t = \sum_{i=1}^{n}[l(y_i, \hat{y}^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t), \tag{3.3}$$

where g_i and h_i denote the first and second-order derivatives on the loss function, respectively. $l(*)$ means the differentiable loss function used to measure the difference between the prediction and the ground truth. Ω denotes the regularisation function.

To accelerate the feature selection, the feature score calculated from XGBoost is used to guide the selection process in ABC. The feature score is measured by the weight in XGBoost, which is the number of times a feature is used to split the data across all trees (T. Chen et al., 2015). The feature score of one feature is calculated by Equations 3.4 - 3.5.

$$fs_i = \sum_{t=1}^{T} \sum_{m}^{M-1} I(fe_t^m, fe) \tag{3.4}$$

$$I(fe_t^m, fe) = \begin{cases} 1 & if fe_t^m == fe \\ 0 & otherwise \end{cases} \tag{3.5}$$

where T indicates the number of trees. M and $M-1$ represent the number of nodes and non-leaf nodes in the tth tree, respectively. fe_t^m refers to the feature related to the node m, and $I(*)$ means the indicator function.

Algorithm 1 describes the FS-ABC method. Firstly, the initial food sources of the algorithm are calculated using Equation 3.6, where $i = 1...SN$ and $j = 1...n$. SN

refers to the number of food sources, and n means the dimension size. o^j_{min} and o^j_{max} represent the lower and upper bounds of dimension j, respectively. Next, different from the original Artificial Bee Colony method, a new food source w^j_i is generated in the employed bees phrase guided by random number $\gamma \in [-1, 1]$ and feature score fs_i produced by the XGB method. The guidance is able to save search time and eliminate some useless food sources. After finding a new food source, the quality of the new food source is evaluated by fitness $fit_i(o_i)$, which is calculated using the cost value $f_i(o_i)$ of the solution w_i. All the calculated qualities will be shared with the onlooker bees, and a food source is detected with probability p_i. If the food source O_i cannot be further improved through several trial limit TR, the food source is to be abandoned, and scout bees determine a new food source by using Equation 3.6. Finally, the best solutions are incorporated in O_{best}. The process will be repeated until criterion c reaches the limit C_{max}.

$$o^j_i = o^j_{min} + rand(0,1) * (o^j_{max} - o^j_{min}) \tag{3.6}$$

Algorithm 1 The FS-ABC algorithm

Output: O_{best}
Input: $O_i = \{o^j_i \| i = 1, 2, ..., SN, j = 1, 2, ..., n\}$
Initialize $O_{best} = None$, $O_{abandoned}$, $tr = 0$;
while $c < C_{max}$ **do**
    $w^j_i = o^j_i + \gamma * (1 - fs_i) * (o^j_i - o^j_k)$
    $fit_i(o_i) = \frac{1}{1+f_i(o_i)}$
    $p_i = \frac{fit_i(o_i)}{\sum_{j=1}^{SN} fit_j(o_j)}$
    if $max(tr) < TR$ **then**
        $O_{abandoned}$ += O_i
        Generate o^j_{new}
    end if
    $c = c + 1$
    Update O_{best}
end while

With the search bias and initial food source selection, I provide a searching space to increase the effectiveness in finding the optimal feature subset. This also turns the ABC algorithm into a semi-directed search algorithm, equipping it with the capability of conducting a global search but avoiding falling into local optima. In addition, the food source candidates are also guided by the XGB feature scores, giving the FS-ABC efficient guidance on the feature candidates' generation.

To balance the effectiveness and exploration searchability of ABC, the iterative food source requested by employed and onlooker bees is modified in the proposed ABC method. In the original ABC algorithm, the neighbouring food source explored by employed bees is updated by Equation 3.7. The random neighbour positions of the old food source are explored in order to discover the position of the new food source, which is able to enhance the exploration ability. However, it cannot be guaranteed that the performance of a randomly selected neighbour is better than that of the current food source. Then ABC converges slowly due to the uncertain search directions.

$$w_i^j = o_i^j + \gamma * (o_i^j - o_k^j) \tag{3.7}$$

In view of the disadvantage of the original ABC, the iterative search of food sources is improved by Equation 3.8. The strategy to select neighbours is proposed based on the feature score. The higher the feature score, the more chance the neighbour is selected. Therefore, the proposed ABC can make full use of the information of neighbours with a better feature score, and the bees have a greater probability of searching for food sources in the right direction with less time.

$$w_i^j = o_i^j + \gamma * (1 - fs_i) * (o_i^j - o_k^j) \tag{3.8}$$

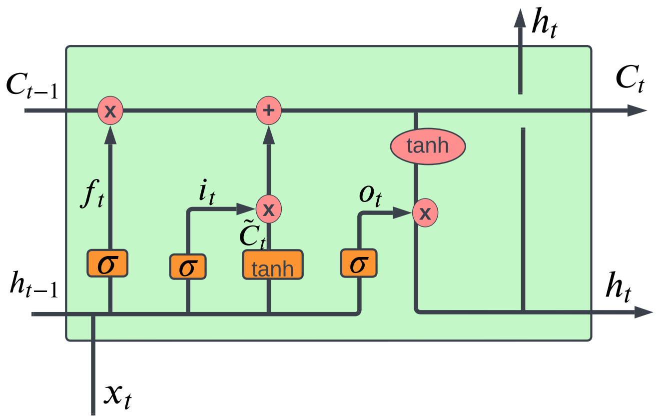


Figure 3.3: The structure of the LSTM cell.

3.4.3 Aspect Term Extraction

The proposed ABC method is exploited to select the most relevant features from the defined linguistic feature set, and the word sequence is converted into an input vector $V^{(abc)} = \{v_1^{(abc)}, v_2^{(abc)}, ..., v_N^{(abc)}\}$ through ABC Feature Extractor. The fused vector representation $V = \{v_1, v_2, ..., v_N\}$ is directly generated by $h^{(bert)}$ and $V^{(abc)}$ in Equation 3.9.

$$v_i = [h_i^{(bert)}; v_i^{(abc)}] \tag{3.9}$$

To effectively learn the fused vector representation, I further employ a BiLSTM encoder to encode each vector v_i using Equations 3.10 - 3.13.

$$i_t = \sigma(W_v^{(i)} * v_i + W_h^{(i)} * h_{t-1} + b^{(i)}) \tag{3.10}$$

$$f_t = \sigma(W_v^{(f)} * v_i + W_h^{(f)} * h_{t-1} + b^{(f)}) \tag{3.11}$$

$$o_t = \sigma(W_v^{(o)} * v_i + W_h^{(o)} * h_{t-1} + b^{(o)}) \tag{3.12}$$

$$\tilde{C}_t = tanh(W_v^{(\tilde{C})} * v_i + W_h^{(\tilde{C})} * h_{t-1} + b^{(\tilde{C})}) \tag{3.13}$$

$$C_t = i_t \odot \tilde{C}_t + f_t \odot C_{t-1} \tag{3.14}$$

$$h_t = o_t \odot tanh(C_t), \tag{3.15}$$

where σ is the sigmoid activation function, $W(\cdot)$ refers to weight parameters. In Equations 3.10 - 3.13, $b(\cdot)$ refers to the bias vector, and $\odot$ represents element-wise multiplication. C and $\tilde{C}_t$ denote cell state and cell input activation state, respectively, carrying information from the previous layer to the next layer. h is the hidden state. Because Bi-LSTM is applied in the proposed method, two representations $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are computed in forward and backward directions. Therefore, the final hidden state can be denoted as $h_t{}' = [\overrightarrow{h_t}, \overleftarrow{h_t}]$. Figure 3.3 shows the cell structure of LSTM.

Next, I feed hidden states $H = \{h'_1, h'_2, ..., h'_N\}$ from BiLSTM to predict the final structured output $\hat{Y} = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}_N\}$ by adding a CRF layer. Finally, to train the proposed framework, the cross-entropy loss is computed as the loss function formulated in Equation 3.16.

$$\mathcal{L}(\hat{Y}, Y) = -\sum_{i=1}^{N}\sum_{j=1}^{L} \hat{y}_{ij} \cdot log(y_{ij}), \tag{3.16}$$

where L indicates the number of tag categories and N means the number of tokens in the review. $\hat{y}_{ij}$ and y_{ij} denote the predicted tag and ground truth tag for word w_i, respectively.

3.5 Experiments

In this section, I first introduce the datasets used in the experiments and present the parameter settings of the proposed method. Then, all the baselines are introduced, and the experimental results are presented. Finally, I conduct an ablation study for a more comprehensive analysis of the proposed framework.

Table 3.1: The statistics of SemEval-2014, SemEval-2015 and SemEval-2016 datasets.

Datasets	Train		Test	
	Reviews	Aspects	Reviews	Aspects
SemEval-2014 (Restaurant)	3041	3686	800	1134
SemEval-2014 (Laptop)	3045	2342	800	650
SemEval-15	1315	1209	685	547
SemEval-16	2000	1757	676	622

3.5.1 Dataset

The proposed FS-ABC is evaluated on four widely used benchmark datasets, i.e., the laptop and restaurant datasets from SemEval-2014 (Pontiki et al., 2014) and two restaurant datasets from SemEval-2015 (Pontiki et al., 2015) and SemEval-2016 (Pontiki et al., 2016). Aspect terms in all datasets are manually annotated for model training and evaluation. Table 3.1 shows the statistics of SemEval datasets for training and testing.

3.5.2 Evaluation Metrics

In this chapter, three standard evaluation metrics, i.e., precision (P), recall (R), and F1 score, are adopted to evaluate the proposed model. They are formulated in Equations 3.17 - 3.19. Among these three standard metrics, the F1 score is the most widely used evaluation metric for the aspect extraction task (Liao et al., 2019; Wei et al., 2021; Venugopalan & Gupta, 2022).

$$P = \frac{TP}{TP + FP} \tag{3.17}$$

$$R = \frac{TP}{TP + FN} \tag{3.18}$$

Table 3.2: The hyperparameters used in the proposed method.

Hyperparameter		**value**
BERT	Encoder block	12
	Attention Head	12
	Max Len (input)	200
	Hidden States	768
	Dropout	0.3
BiLSTM	Hidden States	300
	Layer	3
	Dropout	0.4
	Learning Rate	1e-3

$$F1 = 2 * \frac{P * R}{P + R} \tag{3.19}$$

where TP (true positive) refers to the number of aspect terms detected correctly. FP (false positive) indicates the number of non-aspect terms predicted as aspect terms. FN (false negative) presents the number of aspect terms classified as non-aspect terms.

3.5.3 Experiment Setting

In the experiments, the contextual representations are generated by the pre-trained BERT model [3]. The "bert-base-uncased" model contains 12 encoder blocks of the transformer, and each block consists of 12 self-attention heads and 768 hidden units. The maximum length of training sentences is set to 200. A 3-layer BiLSTM is used and the dimension of the hidden states unit is 300. The dropout rate for BiLSTM is 0.4 and 0.3 for BERT embeddings. The learning rate is set to be 1e-3 for the Adam optimiser. The details of hyperparameter settings are summarised in Table 3.2.

[3]https://github.com/huggingface/transformers

3.5.4 Baselines

To evaluate the proposed FS-ABC, I compare the performance against several competitive baselines, including both machine learning-based and deep learning-based methods. On top of that, I also compare the proposed method with feature selection-based machine learning models. The baselines are listed as follows.

SVM is a traditional supervised machine learning algorithm (Platt et al., 1999). Incorporated with n-gram, analytical, and dictionary features, SVM is able to be used for aspect-based sentiment analysis tasks (e.g., aspect terms extraction, sentiment classification, etc.).

MultinomialNB assumes that the input is a bag of words and calculates the probabilities of classes assigned to words by using the joint probabilities of words and classes (Schütze, Manning & Raghavan, 2008).

RandomForest is an ensemble of decision trees for regression or classification tasks (Breiman, 2001). By constructing a multitude of decision trees at training time, it is able to output the class that is the mode of the classes output by the individual tree.

CRF is a traditional sequence model and has been widely used for subjective expression extraction (e.g., aspect extraction) (Lafferty et al., 2001). By combining parsing, syntactic, lexical, and dictionary-based features, CRF outperforms other traditional machine learning models (e.g., SVM, RandomForest, MultinomialNB, etc.).

PSO is a feature selection method developed for aspect-based sentiment analysis by using the features identified by different classifiers (i.e., Maximum Entropy, CRF, SVM) (Akhtar et al., 2017).

DLIREC is a hybrid system with two components for aspect term extraction and term polarity classification (Toh & Wang, 2014). The system implements a variety of syntactic, semantic, lexicon features, and cluster features delivered from unlabelled data.

POD is a positional dependency-based word embedding, in which the positional context is modelled, and the dependency context is enhanced by integrating more lexical information along dependency paths (Y. Yin et al., 2020).

RNN-based method is an Elman-type RNN model designed for opinion mining (Elman, 1990).

DE-CNN is a multi-layer CNN integrating GloVe and word embeddings of the specific domain (H. Xu, Liu, Shu & Philip, 2018).

LSTM, along with pre-trained word embedding, namely Word2Vec, is utilised for aspect term extraction (P. Liu et al., 2015).

HAST aims to tackle aspect term extraction by exploiting pre-trained word embedding (GloVe), aspect detection history, and opinion summary (X. Li, Bing, Li, Lam & Yang, 2018).

CGL-AE is a novel neural model for aspect extraction by coupling global and local representation (Liao et al., 2019).

3.5.5 Experimental Results and Analysis

I divide the experiments into two partitions: (1) Comparison on SemEval-2014 and (2) Comparison on SemEval-2015 and SemEval-2016. Figure 3.4 shows the comparison results on SemEval-2014 datasets. The experimental results on SemEval-2015 and SemEval-2016 are presented in Figure 3.5.

In Table 3.3, The proposed method can steadily outperform all the baselines in F1 score on both Laptop and Restaurant datasets. Although the RNN-based method achieves the best performance in Recall (R) on the SemEval-2014 Laptop dataset, the proposed method further achieves 7.9% and 2.5% absolute gains in Precision (P) and F1 score, respectively. It implies that compared with machine learning-based models, deep

Table 3.3: Experimental results on SemEval-2014

Method	Laptop			Restaurant		
	P	R	F1	P	R	F1
CGL-AE	0.312	0.605	0.412	0.280	0.502	0.361
RandomForest	0.700	0.533	0.606	0.719	0.614	0.663
MultinomialNB	0.537	0.733	0.620	0.563	0.766	0.649
SVM	0.737	0.587	0.654	0.761	0.695	0.726
CRF-based	0.782	0.673	0.723	0.812	0.781	0.796
RNN-based	0.810	**0.757**	0.782	0.828	0.804	0.816
DLIREC	0.819	0.671	0.738	0854	0.827	0.840
PSO	0.855	0.667	0.749	**0.871**	0.821	0.845
Proposed	**0.889**	0.739	**0.807**	0.856	**0.838**	**0.847**

learning-based methods can capture more important context features with complementary information to benefit aspect term extraction, yielding a better performance than the traditional machine learning algorithms, e.g., RandomForest, MultinomialNB, and SVM. The PSO-based approach performs better than other machine learning baselines in F1. The proposed method obtains slight gains of 5.8% and 0.2% on Laptop and Restaurant datasets, respectively. The results reveal (1) Conventional machine learning algorithms with feature selection can achieve a better performance than some deep learning models without linguistic features for aspect term extraction; (2) By incorporating feature selection into deep learning models, the performance of aspect term extraction can be further improved.

In Table 3.4, it can be observed that the proposed framework can achieve considerable improvements in F1 score on both SemEval-2015 and SemEval-2016 datasets. It can be implied from the results that deep learning-based models give a better performance than conventional machine learning-based methods. By exploiting context embeddings, e.g., Word2Vec and GloVe, LSTM and HAST perform better than the other deep learning-based methods. While I apply feature selection to deep learning models to obtain further gains of 0.7% and 0.4% in the F1 score on both datasets, respectively. Therefore, the comparison shows that feature selection can make significant

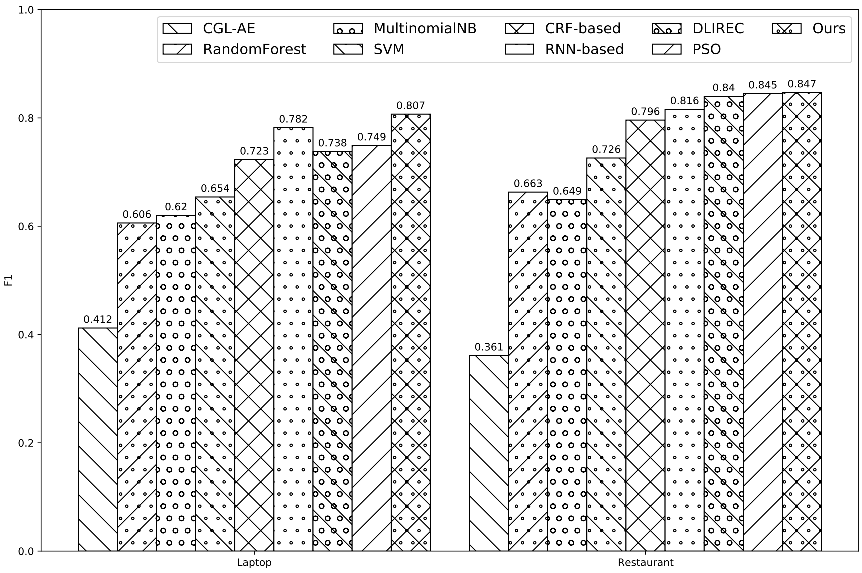


Figure 3.4: Experimental results (F1 score) on SemEval2014 laptop and restaurant datasets.

contributions to aspect term extraction.

The proposed framework is able to outperform both deep learning-based and feature selection-based methods on all four groups of datasets, which validates the effectiveness of the proposed method. To solve aspect term extraction problems, most deep learning-based methods mainly focus on developing complicated models to scale the importance of context embedding. The improvements in the proposed method show that linguistic information can capture different features of aspect terms and complement semantic representations learned by deep learning methods. Feature selection can reduce the high dimension of features. However, the performance is able to improve further if the fusion of semantics and linguistic features are properly designed.

3.5.6 Ablation Study

In this section, I conduct an ablation study to demonstrate the effectiveness of selected linguistic features and the FS-ABC in the proposed framework. To comprehensively understand the importance of the BERT embedding, linguistic features, and feature

Table 3.4: Experimental results on SemEval-2015 and SemEval-2016

Method	**SemEval-15**	**SemEval-16**
	F1	F1
CGL-AE	0.412	0.412
RandomForest	0.513	0.504
MultinomialNB	0.483	0.502
SVM	0.504	0.463
LSTM	0.683	0.704
DE-CNN	0.683	0.744
POD	0.701	0.707
HAST	0.715	0.736
Proposed	**0.722**	**0.748**

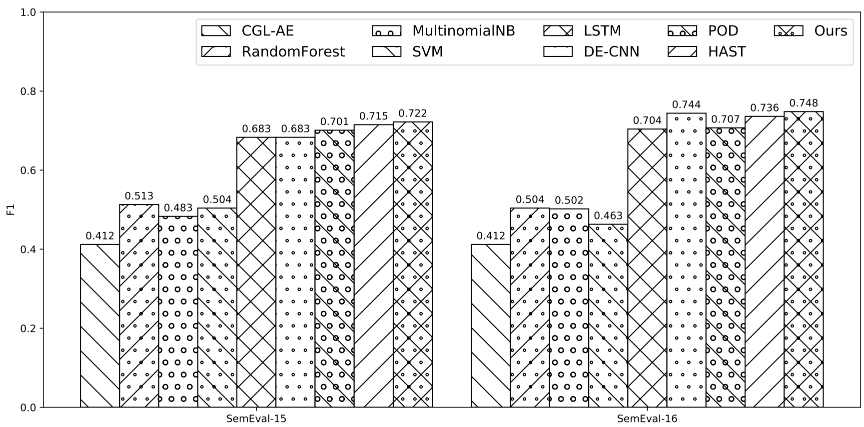


Figure 3.5: Experimental results (F1 score) on SemEval2015 and SemEval2016 datasets.

Table 3.5: Experimental results of the ablation study on different datasets for aspect extraction (F1-score), which aims to analyse the performance of the proposed method with different modules and linguistic features.

	Only BERT[3]			+ALL[4]			+ABC[5]			Ours[6]		
Datasets	P	R	F1	P	R	F1	P	R	F1	P	R	F1
SemEval-2014(L)[1]	0.794	0.704	0.746	0.838	0.747	0.790	0.813	0.725	0.766	0.867	0.755	0.807
SemEval-2014(R)[2]	0.826	0.731	0.776	0.841	0.767	0.802	0.833	0.775	0.803	0.856	0.838	0.847
SemEval-15	0.721	0.592	0.650	0.752	0.616	0.677	0.758	0.614	0.678	0.787	0.667	0.722
SemEval-16	0.762	0.623	0.686	0.788	0.642	0.708	0.794	0.652	0.716	0.815	0.693	0.748

[1]SemEval-2014 Laptop dataset.

[2]SemEval-2014 Restaurant dataset.

[3]Only word embeddings from BERT are applied.

[4]Only all linguistic features are applied.

[5]Only the proposed ABC is applied.

[6]The proposed ABC and selected linguistic features are applied.

selection, I conduct four groups of experiments: (1) **Only BERT**, in which all the selected linguistic features are removed. (2) **+ALL**, in which all the linguistic features without feature selection are included. (3) **+ABC**, in which only selected features by the original ABC are used for aspect extraction. (4) **Ours**, in which only selected features by the proposed ABC are used for aspect extraction. The experimental results of baselines and the proposed method are shown in Table 3.5. As can be seen from the table that all of the designed linguistic features contribute to the performance improvement of aspect term extraction. Meanwhile, the proposed method with selected features outperforms the method with all the linguistic features on the SemEval datasets. When only BERT embedding is applied, the proposed method is able to outperform some traditional machine learning approaches. The improvement proves the effectiveness of BERT in capturing context information. With BERT embedding and all linguistic features, the proposed method fails to surpass all baselines, which implies that some linguistic features will decrease the F1 score of aspect term extraction. To better understand the contribution of FS-ABC, the experiment is conducted to compare the F1 score with the original ABC algorithm on all four group datasets. The improvements in F1 score across all datasets demonstrate the effectiveness of the proposed FS-ABC, proving that it is able to select the most valuable linguistic features.

3.6 Discussion

In this chapter, a novel and effective framework is proposed to address the aspect term extraction task. All the possible linguistic features are explored, and the most relevant features are selected by using the proposed FS-ABC approach. Integrating with contextual representations from the pre-trained model, the selected features are fed into the proposed framework to detect aspect terms. The efficiency of the proposed method is analysed by conducting experiments on four groups of SemEval datasets, and then the experimental results are compared with the original ABC algorithm. The experimental results show that the novel and effective framework can achieve a better performance than state-of-the-art approaches.

Due to the long search time and slow convergence speed, the ABC algorithm consumes more computing resources than other swarm intelligence-based methods (e.g., particle swarm optimisation, genetics algorithm, etc.). To satisfy the needs for real applications, the applicability and effectiveness of the proposed framework are first verified by conducting experiments on four groups of public datasets. In the future, I plan to improve the proposed approach in two aspects. First, parallel computing technology can be applied to accelerate the computing speed of the ABC algorithm. Second, the proposed framework can be applied to other domain datasets, e.g., Amazon and Yelp reviews, for exploring more features.

3.7 Summary

In this chapter, I proposed a novel and effective method for aspect term extraction that integrates both contextual and linguistic features through the use of the Artificial Bee Colony-based feature selection approach. To tackle the issue of high sparsity and high dimensionality in raw data, a distinctive version of the Artificial Bee Colony algorithm is utilised to identify the most valuable linguistic features. The selected linguistic features and context embeddings are combined to improve aspect term extraction's precision. Comprehensive experiments were carried out on real-world datasets, with results demonstrating that the proposed framework outperforms competitive baselines. In comparison with the latest baselines, the proposed framework achieves significantly higher F1 scores.

This chapter mainly answers Research Question 1 mentioned in Chapter 1. The related works of this chapter have been published in (J. Shi et al., 2021) and (J. Shi, Li, Bai & Ito, 2022).

Chapter 4

Graph-based Joint Aspect and Relation Extraction

Public concern detection provides potential guidance to the authorities for crisis management before or during a pandemic outbreak. Detecting people's concerns and attention from online social media platforms has been widely acknowledged as an effective approach to relieving public panic and preventing a social crisis. However, detecting concerns in time from massive volumes of information in social media turns out to be a big challenge, especially when sufficient manually labelled data is in the absence during public health emergencies, e.g., COVID-19. In this chapter, I propose a novel end-to-end deep learning model to identify people's concerns and the corresponding relations based on Graph Convolutional Networks and Bi-directional Long Short Term Memory integrated with Concern Graphs. Except for the sequential features from BERT embeddings, the regional features of tweets can be extracted by the Concern Graph module, which not only benefits the concern detection but also enables the proposed model to be high noise-tolerant. Thus, the proposed model can address the issue of insufficient manually labelled data. I conduct extensive experiments to evaluate the proposed model by using both manually labelled tweets and automatically labelled

Tweets. The experimental results show that the proposed model can outperform the state-of-the-art models on real-world datasets.

4.1 Overview

The outbreak of coronavirus (COVID-19) in 2019 has been causing a rapid increase in both infection and death rates around the world. Especially when the pandemic moved into the second, third, or even fourth wave, it caused devastating loss of human life, impacted the global economy, transformed our daily lives, and posed a threat to our society (Killgore, Cloonen, Taylor & Dailey, 2020). According to the studies on past pandemic outbreaks, e.g., Zika (K.-W. Fu et al., 2016; Glowacki, Lazard, Wilcox, Mackert & Bernhardt, 2016), Ebola (Lazard, Scheinfeld, Bernhardt, Wilcox & Suran, 2015; Van Lent, Sungur, Kunneman, Van De Velde & Das, 2017), and H1N1 (Chew & Eysenbach, 2010; Szomszor, Kostkova & St Louis, 2011), social media platforms, e.g., Twitter, have proven to be a popular channel for spreading information, especially related to public opinions and concerns (Damiano & Catellier JR, 2020). This is because people tend to perceive more details regarding the pandemic by reading the newsfeeds and interpreting the comments from others through social networks (W. Li, Bai, Zhang & Nguyen, 2018; Y. Hu, Bai & Li, 2019). Twitter, a popular and informative social network platform, allows people to post and interact with messages known as “tweets”. They can also communicate and express opinions about the latest events (Killgore et al., 2020). User-generated tweets from Twitter turn out to be prophetic, namely, valuable indicators of what issues will likely happen in the pandemic. Therefore, it is important to make use of tweets and investigate what various people are discussing during the pandemic. The attitudes and behaviours of our society are affected directly by public concerns. Thus, how to effectively extract public concerns and analyse the corresponding relationships will assist people in understanding the anxiety and fears of society in this pandemic situation. Furthermore, the potential social crisis can also be revealed by analysing public concerns, which significantly contribute to social management control.

Motivated by this background, great effort has been dedicated to mining social media data and exploring opinions about pandemic outbreaks (da Silva, Tsigaris & Erfanmanesh, 2021). Most existing research works can be categorised into traditional survey methods, e.g., surveys and questionnaires (Nelson, Pettitt, Flannery & Allen, 2020), and machine learning model-based methods, e.g., topic modelling (Van Der Vegt & Kleinberg, 2020; Kassab et al., 2020). The existing studies are capable of extracting fundamental public concerns, e.g., “social distancing”, “hand sanitiser” and “face masks”, which require intensive human effort in labelling large datasets, turning out to be inefficient. Moreover, in any epidemic emergence situation, e.g., COVID-19, traditional approaches, such as questionnaires and clinical tests, neither collect enough data for deep learning model training nor rapidly generate a model for concern detection. Therefore, it is vital to design an end-to-end model that is capable of automatically analysing social media data and detecting public concerns without requiring a large-scale of data to be labelled manually.

Deep learning methods are increasingly applied to valuable information extraction. However, most methods rely heavily on data labelled by the annotators, requiring much time and financial resources (Kipf & Welling, 2017). Moreover, the noisy and imbalanced social media data prevent deep learning-based methods from generalisation (Rathan, Hulipalled, Venugopal & Patnaik, 2018). In many existing studies, the proposed models are not able to track real-time statistics of public concerns related to pandemics due to the required labelled dataset (L. Li et al., 2020; Jahanbin, Rahmanian et al., 2020; Hou, Du, Jiang, Zhou & Lin, 2020; Lazard et al., 2015). To mitigate this issue, preliminary research was conducted to mine public concerns by proposing an Automated Concern Exploration (ACE) framework (J. Shi et al., 2021). The proposed framework can detect concerns from tweets automatically and construct a concern knowledge graph to present the interconnections of the extracted concern entity set. However, several advent limitations are still to be addressed. (1) only BERT embedding of tweets is used, which cannot capture regional dependency word features from tweets to improve the performance of concern extraction. (2) the relation between concerns in one tweet posted by a user is not detected, which is critical to reveal meaningful information about public concerns. (3) the framework employs a rule-based method, having poor generalisability and appearing difficult to transfer to future occurring pandemics.

In this chapter, I propose and develop an end-to-end model with Concern Graph (CG) and concern states to simultaneously identify public concerns and corresponding relations. "Public concern" is formally defined with a consideration of its type and degree, and construct a concern graph to represent the regional features, improving the concern identification effectiveness. Furthermore, the proposed method can extract concern relations by integrating concern states with Graph Convolutional Network (GCN) (Kipf & Welling, 2017). Extensive experiments are conducted to evaluate the proposed method by using both manual-labelled and auto-labelled datasets. The experimental results explicitly demonstrate that the proposed method outperforms state-of-the-art models.

The novelties of the proposed research work are presented as follows: To the best of my knowledge, the proposed method is the first to apply the deep learning-based method to detect public concerns, which rapidly assists the authority to understand people's anxiety and fears about COVID-19; Furthermore, the concern relation is extracted along with concerns, helping to identify any potential social crisis; I am the first to define a concern graph which contributes to the detection of concerns and corresponding relationships, which leads to the performance improvement of the proposed method. The contributions in this research work are summarised below:

- A concern graph data structure is defined to capture the inherent structural information of concerns more efficiently.
- A novel end-to-end model is presented to jointly extract concerns and relations consisting of the Concern Graph (CG) and shared state of concerns.
- The proposed model is evaluated on manual-labelled data and auto-labelled data, and the results indicate the proposed method is effective for auto-labelled data.

4.2 Related Work

In this section, the existing studies are firstly reviewed, which are related to public concern mining and detection. Then, modern Named Entity Recognition (NER) and Relation Extraction (RE) approaches are inspected and compared since the concern detection, defined in this chapter, tends to explore the concern entities and the corresponding relations. Finally, the GCN and its variants are reviewed since GCN has been widely adopted in NER and RE based on recent studies.

4.2.1 Concern Detection

Social media has become a prevalent platform for people to communicate and express their opinions. With the outbreaks of the pandemic, i.e., Ebola, Zika, and COVID-19, how to effectively extract people's opinions and address public concern in pandemic situations has attracted great attention from researchers. Thus, great efforts have been dedicated to the analysis of public response to pandemics on social media platforms, e.g., Twitter. The current approaches are mainly categorised into two types of methods: probabilistic model-based and deep learning-based. In probabilistic-based models, Latent Dirichlet Allocation (LDA) is commonly used for public concern extractions. For example, Allison et al. apply topic modelling to detect themes of public concern from

Ebola tweets and reveal major insights to inform communication strategies (Lazard et al., 2015). Kim et al. conduct content and sentiment analysis on Ebola Twitter (E. H.-J. Kim, Jeong, Kim, Kang & Song, 2016). Five themes are identified from Zika-related Twitter content by Fu et al. through content analysis (K.-W. Fu et al., 2016). Chandrasekaran et al. conduct a temporal assessment on COVID-19-related tweets to uncover public concern trends by extracting topics and predicting sentiment scores (Chandrasekaran, Mehta, Valkunde & Moustakas, 2020). Xue et al. utilise LDA to analyse public response to COVID-19 pandemic on the social media platform, aiming to identify popular uni-grams and bi-grams topics from tweets (J. Xue et al., 2020). Wahbeh et al. adopt a qualitative analysis tool to detect recommendations, topics, and opinions related to the COVID-19 pandemic from Twitter (Wahbeh, Nasralah, Al-Ramahi & El-Gayar, 2020). Whereas, probabilistic model-based methods perform poorly on public concern identification since contextual information is ignored. By contrast, deep learning-based methods are able to retain the contextual features of sentences. Nowadays, deep learning is widely adopted as a popular approach for many Natural Language Processing (NLP) tasks, such as sentiment analysis. By employing such an approach, many studies aim to extract insightful information for assisting the authorities in making appropriate responses and reactions (T. Wang, Lu, Chow & Zhu, 2020; H. Yin, Yang & Li, 2020; L. Chen, Lyu, Yang, Wang & Luo, 2020).

However, most existing research works only identify a few pre-defined public concerns, but neglect the relations between the concerns. Without concern relations, it is difficult to identify the cause of public concerns or reveal people's thoughts behind the expressed concerns. Different from the above two types of approaches, the proposed method is able to capture regional and sequential features of a sentence and assist the extraction of public concerns with the corresponding relations.

4.2.2 Named Entity Recognition

Named Entity Recognition (NER), also referred to as Entity Extraction (ER), is one of the classic tasks of NLP, which aims to identify and classify named entities from unstructured text into pre-defined categories (Mohit, 2014). Recent studies have shown two typical NER approaches, i.e., traditional statistical models and deep learning-based methods. Zhou et al. propose an entity extraction model with a chunk tagger method based on the Hidden Markov Model (HMM), and the model outperforms the hand-crafted rules-based models (G. Zhou & Su, 2002). Lafferty et al. present Conditional Random Fields (CRF) to segment and label sequence data by building a probabilistic model (Lafferty et al., 2001). However, traditional statistical models perform poorly on complex sentences because they fail to discover hidden features in data. Compared with traditional methods, deep learning-based approaches are able to learn latent representations from raw data and achieve promising performance. Santoso et al. apply Bi-directional Long Short Term Memory (Bi-LSTM) to perform a sequence classification (NER and Part-of-Speech) by understanding the context of the input on the Indonesian language dataset (Santoso et al., 2021). Lample et al. propose a novel neural architecture by relying on character and word representations, which combines Bi-LSTM and CRF (Lample, Ballesteros, Subramanian, Kawakami & Dyer, 2016). Similarly, Ma et al. propose a novel deep learning-based model by combining Bi-LSTM, Convolutional Neural Network (CNN), and CRF (X. Ma & Hovy, 2016). Nowadays, modern state-of-the-art models adopt context-dependent embeddings, e.g., ELMo (Peters et al., 2018), Flair (Akbik, Blythe & Vollgraf, 2018), and BERT (Devlin, Chang, Lee & Toutanova, 2019b), to encode the input.

Although deep learning-based models are capable of capturing contextual features of data, interaction information between entities is neglected. Different from the above models, apart from contextual information, I also propose a designated Concern Graph

(CG) to capture specific features of entities, enabling the proposed method to perform better on Twitter data.

4.2.3 Relation Extraction

As a fundamental task in the NLP field, Relation Extraction (RE) aims to detect and classify the semantic relationship between entity mentions. Early research works mainly focus on rule-based models, in which proper rules are difficult to define without domain knowledge. To address such an issue, many efforts have been dedicated to kernel-based models with manual-labelled data (Culotta & Sorensen, 2004; G.-D. Zhou & Zhu, 2011; Seewald & Kleedorfer, 2007). The key weakness of kernel-based methods is that contextual features are not captured, leading to wrong relation extraction on data with a long sentence. Recently, deep neural networks have been applied to relation extraction due to their supremacy in terms of accuracy. Therefore, some popular deep learning models, e.g., CNN, LSTM, and GCN, are utilised to learn contextual features of data and achieve better performance than kernel-based models (D. Zeng, Liu, Lai, Zhou & Zhao, 2014; Miwa & Bansal, 2016; T.-J. Fu, Li & Ma, 2019).

Apart from extracting entity and relation separately, many other studies investigate joint methods to extract both simultaneously. For example, Arzoo et al. propose an attention-based RNN model for joint entity mentions and relations extraction (Katiyar & Cardie, 2017). Zheng et al. present a novel tagging strategy to covert sequence labelling and classification tasks to a tagging problem and extract entities and relations directly using the joint model (Zheng et al., 2017). Miwa et al. use Tree-LSTM with bidirectional sequential LSTM to extract entity and relation simultaneously (Miwa & Bansal, 2016). Without any manually extracted features, Bekoulis et al. model NER using a CRF layer and the RE as a multi-head selection problem to extract entities and relations simultaneously (Bekoulis, Deleu, Demeester & Develder, 2018). Zeng et al. propose a sequence-to-sequence model with a copy mechanism to extract entity and relation (X. Zeng, Zeng, He, Liu & Zhao, 2018). Hang et al. use one BERT-based parameter-sharing layer to capture the features of entities and relations, then extract entities and relations by applying a source-target BERT model and a three-step overlapping model, respectively (Hang, Feng, Wu, Yan & Wang, 2021).

However, the existing NER, RE, and joint entity and relation extraction models suffer from two issues. First, existing models only discover contextual features of a sentence and neglect entity features, which is vital for entity extraction. Second, relation extraction mainly relies on a sentence's contextual features, and the information of the corresponding entity relationships is ignored. This can be a severe problem for social media data, where numerous grammatical mistakes exist in sentences. To address these two issues, I combine contextual and concern features for concern identification and integrate learned concern features with the module of relation extraction.

4.2.4 Graph Convolutional Network

Graph Convolutional Network (GCN) has demonstrated advent advantages in capturing the dependency structure of sentences, and it has been widely adopted in many NLP tasks (Battaglia, Pascanu, Lai, Rezende & kavukcuoglu, 2016; Defferrard, Bresson & Vandergheynst, 2016; Hamilton, Ying & Leskovec, 2017). As an extension of GCN, Bi-directional Graph Convolutional Network (Bi-GCN) can improve the performance of graph structure data. Figure 4.1 shows the overview of Bi-GCN. In each hidden layer, the model learns the feature description for the current node and its neighbours, along with the graph structure including both directions from the current node to neighbours and the reversed direction. The output layer can obtain information from backward and forward states simultaneously.

Hong et al. present a joint model based on GCN to perform entity and relation extraction by considering the context and syntactic information of sentences (Hong et al., 2020). Zhang et al. utilise GCN over a pruned dependency tree to tackle the relation extraction (Y. Zhang, Qi & Manning, 2018). Inspired by the existing studies, I incorporate GCN into the proposed model to effectively preserve the dependency information of sentences. Furthermore, concern states are integrated with GCN to improve the accuracy of relation extraction.

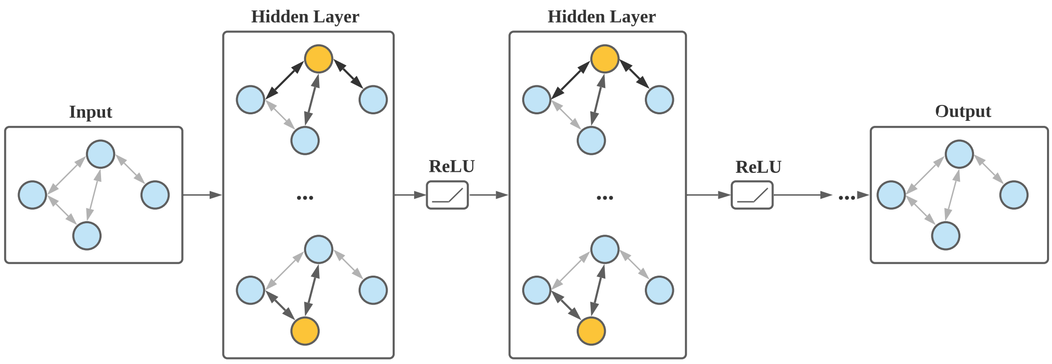


Figure 4.1: Bidirectional Graph Convolutional Network (Bi-GCN) Overview

In this chapter, I proposed an end-to-end model with a concern graph module to perform joint extraction of concerns and relations. Meanwhile, I integrate the concern states from Bi-LSTM with the input features of Bi-GCN to enhance the influences from concerns to improve relation extraction performance.

4.3 Preliminaries

In this section, the relevant definitions are presented, including public concerns, concern relations, and graphs. In addition, the concern detection problem is formally formulated.

4.3.1 Formal Definition

Definition 1: Concern refers to people's worry about a real or imagined issue. Public concern represents a word or a phrase in a tweet towards which most people express strong opinions about a particular aspect of the pandemic. Given a concern set $C = \{c_1, ..., c_n\}$, the ith potential concern detected in tweet t_j can be defined as $c_i^j = (ce_i^j, cs_i^j)$, where $ce_i^j \in CE$ is the concern entity identified in the tweet t_j and it can be words or phrases, e.g., *"China"*, *"corona emergency relief"* and *"florida medical examiner"*. $C = \{c_i^t | i \in [1, N], t \in T\}$ denotes the set of public concerns detected in the Twitter dataset T. For each concern c_i, it consists of attribute named type ct_i and concern score cs_i^j, where $ct_i \in CT$ and $CT = \{ct_1, ..., ct_n\}$ is the set of concern types, and cs_i^j of concern c_i^j is calculated by Equation 4.1:

$$cs^j = (1 - \theta) * |sp^j| + \theta * \tilde{rt}^j, \tag{4.1}$$

where the range of cs_i^j is [0,1], where the greater the value is, the more likely it becomes a concern. $\theta \in [0, 1]$ refers to the weight parameter. $sp_i^j \in [-1, 1]$ denotes the sentiment polarity of the tweet t_j, where -1 indicates an extremely negative attitude, 0 means a neutral attitude, and +1 implies an extremely positive attitude. rt_i^j represents the retweet count of tweet t_j and $\tilde{rt}^j \in [0, 1]$ describes the normalised value of rt^j.

Definition 2: Concern Relation describes the relationship between public concern pairs. I use $r_{m,n}^j \in R$ to present the relation between concern c_m^j and c_n^j in tweet t_j, where $r_{m,n}^j$ is unidirectional relation, i.e., the same as $r_{n,m}^j$, and R is the set of relations extracted from Twitter dataset T.

Definition 3: Concern Triple is the fundamental element of the public concern graph which is extracted from a tweet. To present the real meaning of concern, some short words or phrases in the tweet are very limited in context information. Whereas, the concern triple is capable of semantically representing what concern is about. A public concern triple in the tweet t_j, $ct^j_{m,n} = (s^j_m, r^j_{m,n}, o^j_n)$, has three components, i.e., s^j_m, $r^j_{m,n}$ and o^j_n, referring to as the subject, relation, and object of the concern triple, respectively. The s^j_m and o^j_n are extracted entities, and $r^j_{m,n}$ is the extracted relation based on dependency parser analysis of the tweet t_j.

Definition 4: Concern Graph aims to explore discriminative public concerns and what kind of relations exist between concerns. To present the relation of public concerns, the Concern Graph (CG) is proposed as the control signal for capturing public concerns. CG of the tweet t_j can be denoted as $G = (\nu, \varepsilon)$, where ν is the set of nodes, and ε is the edges set. As shown in Figure 4.2, nodes of CG are classified into four categories: (1) object o^j, subject node s^j; (2) relation node r^j; (3) attribute node a^j including concern type ct^j; (4) concern score cs^j.

The CG G is constructed via the following steps:

1. Detect public concern c^j_i and add it to G, where c^j_i is grounded in the tweet t_j.
2. Extract the descriptive details of concern c^j_i as the attribute $a^j_{i,l}$ including type ct^j_i and score cs^j_i, then add them to G and assign an un-directed edge from c^j_i to a^j_{il}, where $|l|$ is the number of attributes towards concern c^j_i.
3. Identify the relation r_{ik} between concerns c^j_i (subject in concern triple) and c^j_k (object in concern triple), which is a unidirectional type of relation, adding relation node r_{ik} to G and assigning edges from c^j_i to r_{ik} and from r_{ik} to c^j_k.

4.3.2 Problem Formulation

In the previous section, the related definitions are described. Based on the definitions, the proposed model aims to jointly extract typical concerns $\{c^j_i | c^j_i \in C\}$ from tweet t_j and concern relations $\{r^j_{mn} | r^j_{mn} \in R\}$, where r_{mn} is the relation between concern c_m and c_n from Twitter dataset T by constructing CG.

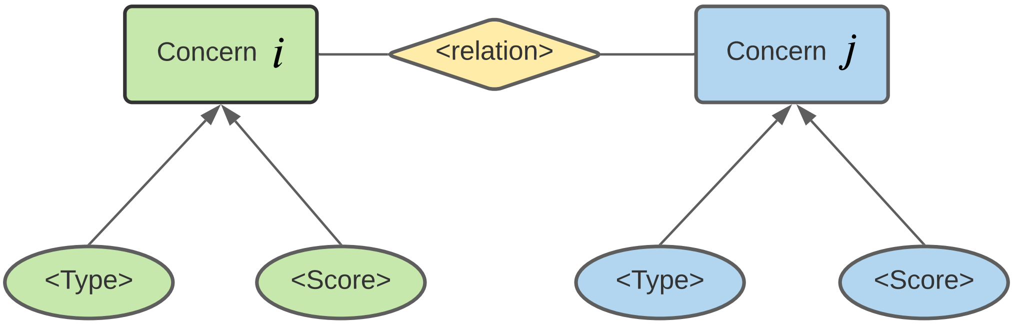


Figure 4.2: Concern Graph: each concern has two attributes, i.e. type and score, along with relation to another concern to form the concern graph.

4.4 Graph-based Concern and Relation Extraction

For a set of tweets T, the goal of the proposed method is to identify public concerns $C = \{c_1, ..., c_n\}$ and concern relations $R = \{r_1, ..., r_n\}$. In this section, the joint extraction of concerns and relations model with the concern graph is illustrated in Figure 4.3. The proposed method consists of four main components, i.e., embedding layer, encoding layer, concern decoding layer, and concern relation extraction layer. Each component is described in detail below. The embedding layer is introduced in Section 4.4.1, followed by the encoding layer in Section 4.4.2. Concern decoding and concern relation extraction layer are presented in Sections 4.4.3 and 4.4.4, respectively. The model objective function is explained in Section 4.4.5.

The proposed method is named Concern-Graph-based Concern and Relation Extraction (CG-CRE). In Algorithm 2, the training process of CG-CRE is demonstrated for improved understanding. All weight parameters are initialised in Bi-LSTM and Bi-GCN. Subsequently, the CG embedding is generated in each epoch and then computes the loss function of concern and relation. The final objective function for model training is calculated based on concern and relation loss function with a trade-off coefficient.

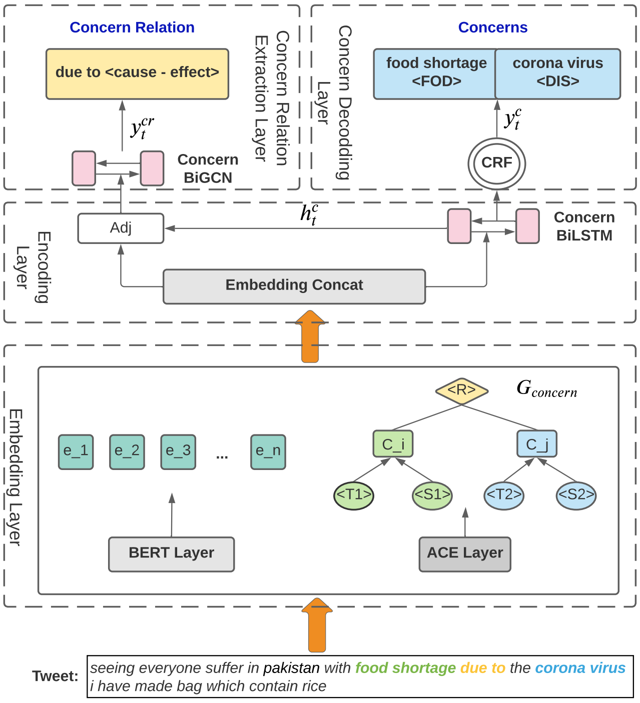


Figure 4.3: The overview of the CG-CRE model.

Algorithm 2 Training Process of CG-CRE Model

Input: T, υ_w, υ_{cg}, E, B, lr, d
T indicates labelled Twitter corpus for model training
υ_w represents BERT embedding of Twitter corpus
υ_{cg} is embedding of CG
EP is epoch number
B is batch size
lr indicates learning rate
d is embedding dimension
Output: L
L is the loss function value
Initialise weight vector W
while ep in EP **do**
  while b in B **do**
    generate embedding $\acute{x}_i$
    compute concern hidden state $h_t{}^{(b)}$ as Equation 3.15
    $L_{(c)}^{(b)} = max(\sum log(P_{(c)}^{(b)} = S_{(c)}^{(b)}))$
    $L_{(r)}^{(b)} = max(\sum log(P_{(r)}^{(b)} = S_{(r)}^{(b)}))$
    $L^{(b)} = L_{(c)}^{(b)} + \alpha * L_{(r)}^{(b)}$
  end while
end while

4.4.1 Embedding Layer

Since deep learning models are integrated into the proposed method, word tokens, and proposed CG need to be transformed into low-dimensional vectors by the embedding layer. The embedding layer includes BERT layer and Auto Concern Extraction (ACE) layer. ACE is concern extraction method proposed in the prior research work (J. Shi et al., 2021). Given a tweet $t = \{w_1, ..., w_i, ..., w_n\}$, where w_i denotes the ith word in the tweet, pre-trained BERT model is used to generate word embedding set $\grave{X} = \{\tilde{e_1}, ... \tilde{e_i}, ..., \tilde{e_n} \,|\, \tilde{e_i} \in \mathbb{R}^d\}$, where $\tilde{e_i}$ represents the embedding of word w_i and d means the embedding dimension.

To enhance model input features, I further encode proposed CG G to obtain CG node embedding $\acute{x}_i^{(0)}$ in Equation 4.2:

$$\acute{x}_i^{(0)} = \begin{cases} (v_i^{(dep)} + v_i^{(pos)}) \odot W_{cr}[0], & if\, i \in C; \\ v_i^a \odot W_{cr}[1], & if\, i \in A; \\ v_i^r \odot W_{cr}[2], & if\, i \in R; \end{cases} \tag{4.2}$$

where $v_i^{(dep)}$ and $v_i^{(pos)}$ denote the syntactic dependency relation and POS tag feature, respectively. Both $v_i^{(dep)}$ and $v_i^{(pos)}$ are used to capture the meaning of tweet and words syntactic dependency. C represents the concern set. v_i^a represents the attribute features, including concern type and score. A means attribute set. v_i^r indicates relation feature, and R is relation set. $W(\cdot) \in \mathbb{R}^{3 \times d}$ refers to parameters, where d means the feature dimension.

As illustrated in Figure 4.4, the input tweet is split into words that are used to generate inputs for the BERT layer and embedding layer. The input words are encoded into positional embeddings, segment embeddings, and token embeddings, which are essential parts of the attention mechanism of BERT. After the dependency relation vector, POS tag vector, attribute vector, and relation vector are encoded, they are fed

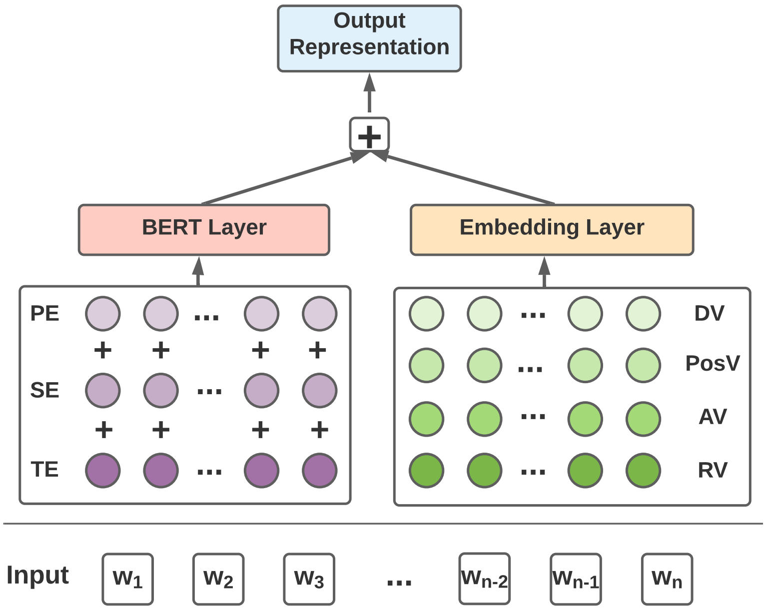


Figure 4.4: The embedding of the CG-CRE model. PE, SE, and TE mean positional embeddings, segment embeddings, and token embeddings, respectively. DV, PosV, AV, and RV refer to dependency relation vector, POS tag vector, attribute vector, and relation vector, respectively.

into the embedding layer to generate CG embeddings. The final output representation is computed by concatenating the outputs from the BERT layer and the embedding layer.

4.4.2 Encoding Layer

To capture long-distance dependencies and forward and backward features between tokens in tweets, Bi-LSTM is used in this chapter. The Bi-LSTM contains forward and backward layers, and a concatenation layer of backward and forward state information. The embeddings in Section 4.4.1 are concatenated as the input of the concern encoder layer. The Bi-LSTM encoding layer is defined by using Equations 4.3 - 4.8:

$$i_t = \sigma(W_{ex}^{(i)} * [\tilde{e}_i; \acute{x}_j^{(i)}] + W_h^{(i)} * h_{t-1} + b^{(i)}) \tag{4.3}$$

$$f_t = \sigma(W_{ex}^{(f)} * [\tilde{e}_i; \acute{x}_j^{(f)}] + W_h^{(f)} * h_{t-1} + b^{(f)}) \tag{4.4}$$

$$o_t = \sigma(W_{ex}^{(o)} * [\tilde{e_i}; \acute{x}_j^{(o)}] + W_h^{(o)} * h_{t-1} + b^{(o)}) \tag{4.5}$$

$$u_t = \sigma(W_{ex}^{(u)} * [\tilde{e_i}; \acute{x}_j^{(u)}] + W_h^{(u)} * h_{t-1} + b^{(u)}) \tag{4.6}$$

$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \tag{4.7}$$

$$h_t = o_t \odot tanh(c_t), \tag{4.8}$$

where σ is sigmoid activation function, $W(\cdot)$ refers to weight parameters, and $[;]$ is a vector concatenation operation. $\tilde{e_i}$ and $\acute{x}_j(\cdot)$ denote word embedding and embedding of CG G defined in Section 4.4.1. In Equations 4.3 - 4.6, $b(\cdot)$ refers to the bias vector, and $\odot$ represents element-wise multiplication. c and h denote cell state and hidden state, respectively, carrying information from the previous layer to the next layer. Because Bi-LSTM is applied in the proposed method, the hidden state is obtained by concatenating hidden states in both directions, namely, forward direction $\overrightarrow{h_t}'$ and backward direction $\overleftarrow{h_t}'$, therefore, the final hidden state can be denoted as $h_t' = [\overrightarrow{h_t}', \overleftarrow{h_t}']$. By passing hidden state to a fully connected neural network, the final output of Bi-LSTM can be defined in Equation 4.9:

$$O = W^o * h_t' + b^o, \tag{4.9}$$

where W^o is the output weight parameters and b^o is the bias vector.

4.4.3 Concern Decoding Layer

In the proposed model, the CRF is employed to produce a tag sequence since it can produce a higher tagging accuracy than that of the existing models (Hong et al., 2020). For one tweet $t = \{w_1, ..., w_n\}$, the goal is to predict the concern tag sequence $Y^{(c)} = \{y_1^{(c)}, y_2^{(c)}, ..., y_n^{(c)}\}$ where n denotes the number of words and superscript (c) means the notation of concern. Thus, the CRF score can be defined as in Equation 4.10:

$$S^{(c)}(t, Y^{(c)}) = \sum_{i=1}^{n} O_{i,y_i^{(c)}} + \sum_{i=1}^{n} T_{y_i^{(c)}, y_{i+1}^{(c)}}, \tag{4.10}$$

where $O \in \mathbb{R}^{n \times k}$ indicates the matrix of scores output from the previous encoding layer with k as the number of distinct tags, and $O_{i,j}$ denotes the score of the jth tag of the ith word in tweet t. T represents a matrix of transition scores as being introduced in (Huang, Xu & Yu, 2015), and $T_{i,j}$ means the score of a transition from tag i to tag j. Then, for input tweet t, the probability of a given sequence of tags over the sequence of predicted tags $Y^{(c)}$ is defined by applying the Softmax layer as in Equation 4.11:

$$P^{(c)} = \frac{e^{S^{(c)}(t, Y^{(c)})}}{\sum_{\widetilde{y}^{(c)} \in Y_X^{(c)}} e^{S^{(c)}(t, \widetilde{y}^{(c)})}}, \tag{4.11}$$

In Equation 4.11, $Y_X^{(c)}$ denotes all possible concern tag sequences for tweet t.

4.4.4 Concern Relation Extraction Layer

Given concern set $C = \{c_1, ..., c_m\}$ in tweet $t = \{t_1, ..., t_n\}$, the goal is to extract the corresponding relation $r_i \in R$. Except for sequential features, Bi-GCN is utilised to capture regional features from the tweets. Both forward and backward directions are considered and the hidden state of Bi-GCN is defined using Equations 4.12 - 4.14:

$$\overrightarrow{h_t''} = \varsigma\Big(\sum_{v \in \overrightarrow{N(w)}} (\overrightarrow{W_h} * [\overrightarrow{h_{t-1}^{v}{}''}; \overrightarrow{h_{t-1}{}'}] + \overrightarrow{b}) \Big) \tag{4.12}$$

$$\overleftarrow{h_t''} = \varsigma\Big(\sum_{v \in \overleftarrow{N(w)}} (\overleftarrow{W_h} * [\overleftarrow{h_{t-1}^{v}{}''}; \overleftarrow{h_{t-1}{}'}] + \overleftarrow{b}) \Big) \tag{4.13}$$

$$h_t'' = [\overrightarrow{h_t}''; \overleftarrow{h_t}''], \tag{4.14}$$

where ς represents ReLU activation function, h_t'' refers to the hidden state at tth layer and $\overrightarrow{h_{t-1}}'$ indicates the shared hidden state from the concern detection module. $\overrightarrow{N(w)}$ describes the neighbours of word w in the forward direction and $\overleftarrow{N(w)}$ means the neighbours of word w in the backward direction. $\overrightarrow{W_h}$ and $\overleftarrow{W_h}$ represent weight parameters in the forward and backward direction, respectively. $\overrightarrow{b}$ and $\overleftarrow{b}$ are the bias of the model. h^t refers to the final hidden state of word w, concatenating hidden states in both directions.

By using hidden states of Bi-GCN, the relation tendency score $S^{(r)}_{(r_{ij}|c_i,c_j)}$ is defined in Equation 4.15:

$$S^{(r)}_{(r_{ij}|c_i,c_j)} = W^{(r)} * \varsigma(W^{(r)}_{c_i} * h''_{c_i} + W^{(r)}_{c_j} * h''_{c_j} + b^{(r)}), \tag{4.15}$$

where superscript (r) means the notation of concern relation. $S^{(r)}_{(r_{ij}|c_i,c_j)}$ represents the tendency score of concern relation on concerns pair (c_i, c_j). $W^{(r)}$, $W^{(r)}_{c_i}$ and $W^{(r)}_{c_j}$ are weight parameters. $b^{(r)}$ denotes the bias term. The activation function (Softmax) is applied to the tendency score $S^{(r)}_{(r_{ij}|c_i,c_j)}$ to obtain the probability of relation $r_{i,j}$ in Equation 4.16:

$$P^{(r)} = \sigma(S^{(r)}_{(r_{ij}|c_i,c_j)}) \tag{4.16}$$

4.4.5 Model Objective Function

In this subsection, the final objective function for model training is described. To train the proposed model, the maximum log-likelihood is used as the loss function and maximise combined loss functions of concern and relation by using Equations 4.17, 4.18, and 4.19:

$$L_{(c)} = max(\sum_{i=1}^{|\mathbb{R}_T|} \sum_{w=1}^{|W_i|} log(P^{(c)}_w = S^{(c)}_w | t_i, \Theta)) \tag{4.17}$$

Table 4.1: Statistics of the Concern Categories

Type	Tweets	Concern Category							
		FIN	GOV	DIS	MED	PER	LOC	FOD	DAT
Manual-labelled	1761	315	457	1239	471	289	341	204	206
		9%	13%	35%	13%	8%	10%	6%	6%
Auto-labelled	40068	4341	19941	23853	6944	10977	1519	1498	11063
		5%	25%	30%	9%	13%	2%	2%	14%

$$L_{(r)} = max(\sum_{j=1}^{|\mathbb{R}_T|} log(P_w^{(r)} = S_w^{(r)} | t_j, \Theta)) \tag{4.18}$$

$$L = L_{(c)} + \alpha * L_{(r)} \tag{4.19}$$

where $|\mathbb{R}_T|$ is the size of the training dataset, t_i and t_j is the ith and jth tweet in the training dataset, respectively. $|W_i|$ is the sentence length. $\alpha \in [0, 1]$ is a trade-off coefficient between loss of concern and concern relation, and the larger value means the greater influence of concern relation on the proposed method.

4.5 Experiments

In this section, extensive experiments are conducted to evaluate the proposed approach by using COVID-19 Twitter datasets. First, COVID-19 dataset collection and pre-processing are described. Second, the proposed approach is compared against six state-of-the-art baselines in terms of precision, recall, and F1 score. Third, quantitative analytical results and conduct ablation studies are presented following the experimental results. Finally, a case study is given to illustrate the effectiveness of the proposed approach.

4.5.1 Dataset and Experiment Setting

Twitter is one of the largest social media platforms, providing a rich source for evidence. It is easy for people to obtain the tweets associated with COVID-19 by using API. The experiments are conducted by using a public large-scale Twitter dataset about COVID-19, which contains English language-specific tweets from 204 different countries and territories (Lamsal, 2021). The dataset is proposed in the scientific literature for research with topics related to COVID-19.

The dataset has been pre-processed in two ways, i.e., manual annotation and auto-annotation. In the former, the annotators label the tweets according to the concern definitions and formulations. While, in the latter, tweets are annotated by using the approach proposed in my past research work (J. Shi et al., 2021).

Many prior research works have explored people's reactions and attempted to discover wide-spreading topics about COVID-19 (L. Li et al., 2020; Killeen et al., 2020; Hou et al., 2020; Kaveh-Yazdy & Zarifzadeh, 2020; L. Li et al., 2020). Based on the findings and conclusion of these works, the most popular topics are extracted and eight types of concerns are defined, i.e., Finance (FIN), Government (GOV), Disease (DIS), Medicine (MED), Person (PER), Location (LOC), Food (FOD), and Date and Time (DAT). On top of that, two types of relations among the concerns, i.e., co-occurrence and cause-effect, are investigated. This is because both types of relations are capable of capturing implicit information about public concerns, demonstrating their associations and potential causes. For instance, by analysing the tweet "... due to the locked transportation..., farmers forced to dump green chilli ...", it is important to know the concern "green chilli" is dumped due to the concern "locked transportation" in the time of COVID-19 pandemic. The statistics of concerns and the relations are listed in Table 4.1 and Table 4.2, respectively.

The statistics of the datasets are listed in Table 4.3. The dataset is divided into 2

Table 4.2: Statistics of Concern Relation Categories

Type	Tweets	Concern Relation	
		CO_OCC	CA_EFF
Manual-labelled	1761	932	829
		53%	47%
Auto-labelled	40068	19485	20583
		49%	51%

Table 4.3: Statistics of the manual-labelled and auto-labelled dataset

	Train	Test
Manual-labelled	1418	343
Auto-labelled	32264	7804

sub-datasets: train dataset and test dataset, occupying 80% and 20%, respectively.

Evaluation Metrics

In this chapter, three standard evaluation metrics, i.e., precision, recall, and F1 score, are employed to evaluate the proposed model.

The outcome of predicted concerns is considered correct only when both of the concerns in one tweet are predicted correctly. In other words, $(c1, c2)$ is recognised as a correct concern pair if $c1$ and $c2$ are correctly predicted at the same time. Correspondingly, the relation prediction is considered valid only when the associated concern pair is correctly predicted.

Hyper-parameters

The language model BERT has been proven to be effective for many natural language processing tasks. In the experiments, the pre-trained BERT-base [1] is utilised to obtain word representations of tweet corpus, and the hidden dimension of embedding H is set as 768. Because BERT uses WordPiece tokeniser to generate word tokens, some concern words may break into several pieces. To detect concern in one word instead of

[1]https://github.com/huggingface/pytorch-pretrained-BERT

sub-word pieces, the corresponding representations of sub-word tokens are averaged to get one concern representation. For example, the representation of the concern "covid-19" is the average of representations of three-word pieces "co", "##vid", "##19". The proposed network is regularised by using dropout at the embedding layer, with a dropout ratio of 0.2. Bi-LSTM and GCN are adopted as the encoding layer, with 300 LSTM units. I employ the full dependency tree of sentences as the adjacency matrix of GCN.

4.5.2 Baselines

The proposed approach is evaluated by comparing it against the following baselines.

- ***Joint Model*** (Zheng et al., 2017) is a joint extraction method to detect both entity and relation in one tweet by using a novel tagging scheme. It is an end-to-end model consisting of a Bi-LSTM encoder layer and an LSTM decoder layer.
- ***Copy Mechanism Model*** (X. Zeng et al., 2018) is a state-of-the-art model for jointly extracting relation triplets from a sentence. It is also an end-to-end model based on seq-to-seq learning with a decoder layer, having two different decoding methods, i.e., one-decoder and multi-decoder. Both different strategies are used as counterparts in the experiments.
- ***SPTree***(Miwa & Bansal, 2016) is a novel end-to-end recurrent neural network model aiming at extracting entities and relations by capturing word sequence and dependency tree substructure features. The stacked bidirectional tree-structured LSTM-RNN models are applied to sequential Bi-LSTM-RNN models to detect both entities and relations with shared parameters.
- ***JointER***(B. Yu et al., 2020) is a joint entity and relation extraction model which can address the limitations, including redundant entity pairs and ignoring the

important inner structure of entities. The model decomposes a joint extraction task into Head-Entity (HE) extraction and Tail-Entity-Relation (TER) extraction to detect head-entity, tail-entity, and relations.

- ***SPERT***(Eberts & Ulges, 2020) is introduced as a span-based model, which can jointly extract entity and relation by conducting lightweight reasoning on BERT embedding and relation classification based on localised and marker-free context features.

- ***CopyMTL***(D. Zeng, Zhang & Liu, 2020) is a multi-task learning framework with copy mechanisms to predict multi-token entities and relations. It is an extremely effective model which can address two existing problems of the entity and relation extraction: (1) inaccurate entity extraction caused by failing to differ the head and tail entity; (2) failing to predict multi-token entities.

4.5.3 Experimental Results and Model Analysis

In this section, I present and 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, I select all the counterparts, which incorporate a Bi-LSTM encoder layer.

The experimental results are demonstrated in Table 4.4, which presents the predicted outcomes, i.e., Precision, Recall, and F1, of the proposed approach as well as the state-of-the-art methods on manual-labelled and auto-labelled datasets. As can be observed from the table, the proposed approach outperforms the others in terms of F1 score, which proves its effectiveness. Specifically, in Figures 4.5 - 4.7, the CG-based model outperforms One-decoder, Multi-decoder, NovelTagging, SPTree, and CopyMTL models on both manual-labelled and auto-labelled Twitter datasets. Although JointER and SPERT achieve better performance than that of ours in terms of precision and recall

Table 4.4: Evaluation results of different models on COVID-19 Tweets Dataset

Model	Manual-labelled Tweets			Auto-labelled Tweets		
	Precision	Recall	F1	Precision	Recall	F1
One-Decoder	0.167	0.161	0.164	0.328	0.321	0.326
Multi-Decoder	0.159	0.152	0.156	0.399	0.347	0.373
NovelTagging	0.273	0.336	0.302	0.570	0.593	0.582
SPTree	0.424	0.349	0.383	0.434	0.366	0.397
JointER	**0.644**	0.369	0.469	0.405	0.314	0.354
SPERT	0.239	**0.675**	0.339	0.310	**0.839**	0.421
CopyMTL-One	0.427	0.393	0.412	0.461	0.413	0.447
CopyMTL-Mul	0.538	0.515	0.530	0.594	0.551	0.573
Proposed Model	0.545	0.630	**0.567**	**0.638**	0.642	**0.592**

on the manual-labelled dataset, SPERT leverages the pre-trained BERT model to obtain contextual features of sentences, but the inner structure of entities is neglected, which inevitably hinders the performance of entity and relation extraction. The embeddings in JointER are initialised using the shallow representatives model, i.e., Glove (Pennington et al., 2014), without context-specific information, which is critical for entity and relation extraction models.

The promising performance of the proposed approach mainly attributes to its structural design. First, the interaction of the CG structure captures the inner dependency between concerns. Second, the shared state passing from the concern extraction module to relation extraction module, provides important concern features for relation extraction. It is worth noting that baselines can achieve state-of-the-art results on high-quality datasets, e.g., NYT and WebNLG, but the performance significantly degrades on the noisy and imbalanced social media data. The grammatical mistakes of tweets make it difficult to capture relations between concerns. NovelTagging and SPTree utilise novel tagging but cannot carry out promising results. Other baselines, including One-Decoder, Multi-Decoder, JointER, and CopyMTL, apply Bi-LSTM to capture sequential features of concerns, but they fail to detect the relation features and concerns due to the

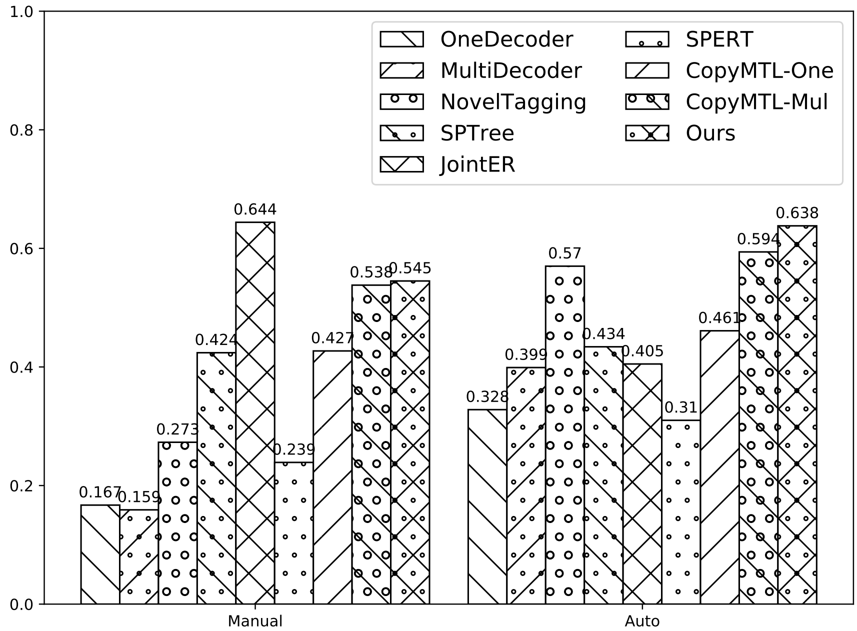


Figure 4.5: Experiment results (Precision) on COVID-19 tweets dataset.

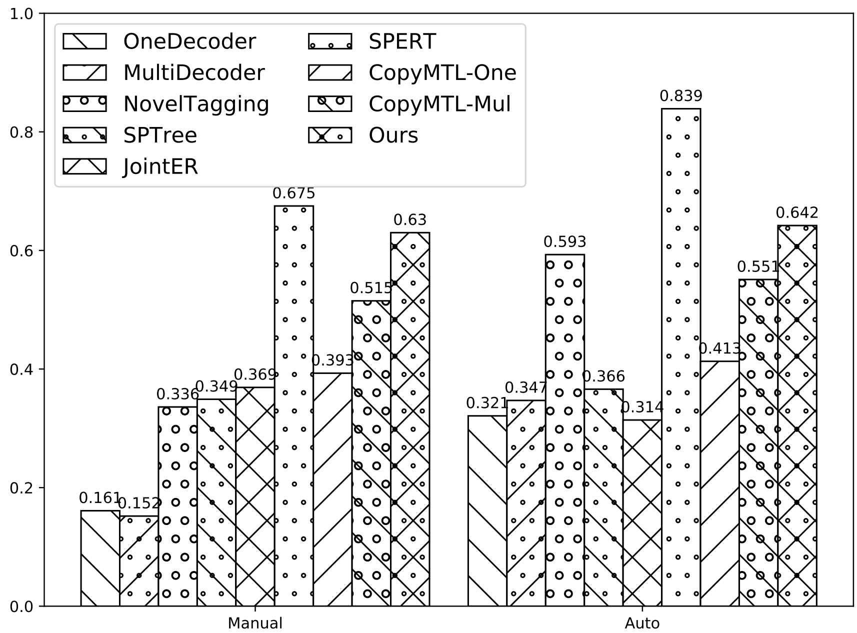


Figure 4.6: Experiment results (Recall) on COVID-19 tweets dataset.

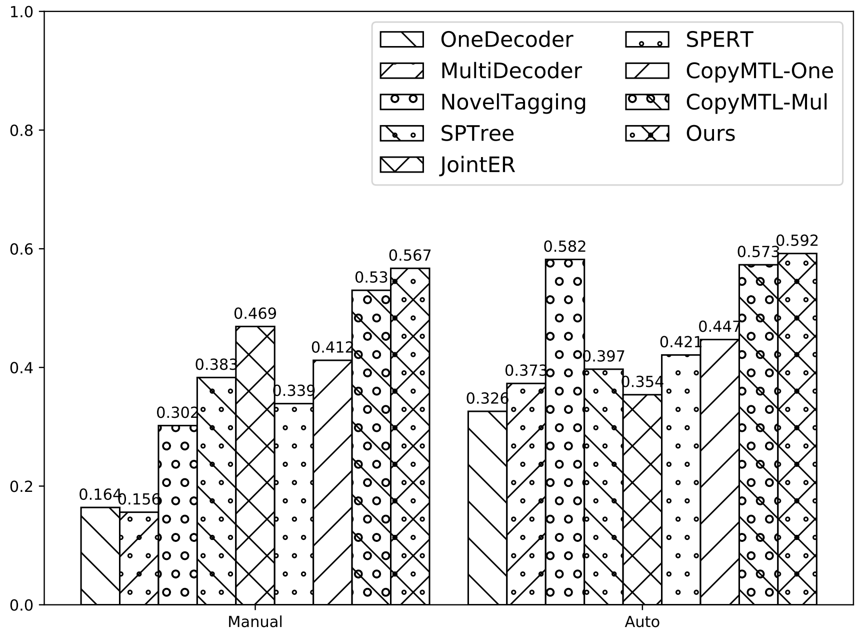


Figure 4.7: Experiment results (F1 score) on COVID-19 tweets dataset.

Table 4.5: The performance of the proposed method with different training and testing data distribution

Test-Train (%)	Manual-labelled Tweets			Auto-labelled Tweets		
	Precision	Recall	F1	Precision	Recall	F1
30-70	0.497	0.573	0.488	0.569	0.586	0.536
20-80	0.545	0.630	0.567	0.638	0.642	0.592
10-90	0.533	0.614	0.550	0.621	0.618	0.571

unstructured sentences in the tweet dataset.

To better understand the experimental results, some examples are presented, which are obtained by applying the proposed method to COVID-19 tweets. The examples are demonstrated in Figure 4.8. The proposed method can detect two concerns (e.g., "food shortage", "corona virus") and the concern types (e.g., "FOD", "DIS"). Moreover, the relation (e.g., "CA_EFF") between concerns is further extracted from the tweets. Incorporating with concern relation, the proposed method boosts in revealing meaningful information about public concerns instead of only knowing isolated concerns from tweets.

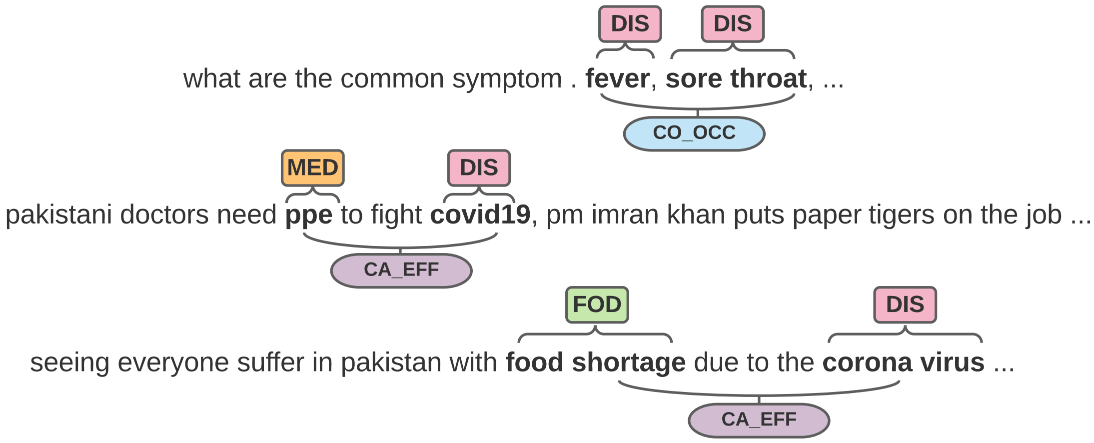


Figure 4.8: Examples of the proposed method on the COVID-19 tweets.

To study the effect of training and test data distribution on the proposed model performance, some experiments are conducted by dividing tweets into training and testing data with different training-testing ratios. Both manual and auto labelled data are divided into three groups with three types of training-testing ratios, 70%-30%, 80%-20%, and 90%-10%. The experimental results are summarised in Table 4.5. As can be seen, the proposed method achieves the best performance on manual and auto labelled tweets when datasets are split into 80% training and 20% test data. when the training data is increased to 90% of total tweets, the proposed method begins over-fitting, and the F1 score drops by 2.1%. The increasing of training data can be helpful in increasing the performance of the proposed method, while it can also decrease the proposed method's performance due to the over-fitting of the model.

In Table 4.4, the proposed method is compared against other models which can detect entities and extract relations simultaneously, and the experimental results demonstrate the superiority of the proposed method in concern and relation extraction. Moreover, I explore the effect of the embedding dropout on the performance of the proposed method in Table 4.6. The experimental results show that the performance (e.g., F1 score) will be decreased with low embedding dropout values on both manual and auto labelled datasets, for example, the F1 score decreases 2.1% and 1.9% with embedding dropout

Table 4.6: The performance of the proposed method with different embedding dropout values

Embedding	Manual-labelled Tweets			Auto-labelled Tweets		
Dropout	Precision	Recall	F1	Precision	Recall	F1
0	0.540	0.612	0.546	0.618	0.642	0.573
0.1	0.538	0.615	0.554	0.623	0.644	0.579
0.2	0.533	0.626	0.561	0.629	0.650	0.587
0.3	0.545	0.630	0.567	0.638	0.642	0.592
0.4	0.524	0.602	0.553	0.611	0.625	0.581

0 and 0.3 on both datasets, respectively. The results also report that the performance improvement is not able to obtain when dropout values are reduced to 0.2 and 0.3.

4.5.4 Ablation Study

The ablation study in this section aims to investigate the impact of CG and shared state components in the proposed approach.

Since manual labelling a large-scale dataset turns out to be a tedious and non-trivial task, sufficient manual-labelled training data sets are usually not available to conduct public concern extraction and analysis for an emergency event. Furthermore, the public concern coverage in datasets also appears imbalanced, which prevents the existing models from generalisation, subsequently impacting the performance to a large extent. The proposed approach can mitigate this issue, giving an outstanding performance on both manual-labelled and auto-labelled datasets.

Table 4.7 lists the results of the ablation study. The approach has been re-evaluated by comparing the performance against that without CG component and shared state components. It can be seen from the table that, in manual-labelled dataset, CG-CRE with CG and shared state outperforms the models without CG and shared state by 11% and 7%, respectively. While in auto-labelled datasets, it surpasses 6% and 1%, respectively. The results explicitly reveal that CG and shared state components play a significant role in jointly identifying concerns and relations.

Table 4.7: Ablation study of CG-CRE model on manual-labelled and auto-labelled Tweets Dataset

Dataset	Method	Precision	Recall	F1
Manual-labelled Tweets	CG-CRE (without CG)	0.416	0.482	0.457
	CG-CRE (without shared state)	0.463	0.516	0.494
	CG-CRE (with all components)	**0.545**	**0.630**	**0.567**
Auto-labelled Tweets	CG-CRE (without CG)	0.551	0.583	0.536
	CG-CRE (without shared state)	0.615	0.624	0.586
	CG-CRE (with all components)	**0.638**	**0.642**	**0.592**

Table 4.8: Outputs from different models on tweets. "pred:[]" means the model predicts null for this tweet. NovelTagging only predicts "c1" and "c2" without concern types.

Models	Tweet
NovelTagging	[**seeing everyone**]$_{c1,r:co\_occ}$ suffer [**in pakistan**]$_{c2,r:co\_occ}$ with food shortage due to the corona virus i have made bag which contain rice
JointER	seeing everyone suffer in pakistan with food shortage due to the corona virus i have made bag which contain rice [**pred:[]**]
CG-CRE	seeing everyone suffer in pakistan with [**food shortage**]$_{c1:FOD,r:ca\_eff}$ due to the [**corona virus**]$_{c2:DIS,r:ca\_eff}$ i have made bag which contain rice
NovelTagging	a greeting from the heart to [**doctors**]$_{c1,r:co\_occ}$, nurses, [**paramedics**]$_{c2,r:co\_occ}$, ... who stand together to tackle the corona epidemic.
JointER	a greeting from the heart to [**doctors**]$_{c1:MED,r:co\_occ}$, [**nurses**]$_{c2:MED,r:co\_occ}$, paramedics, ... who stand together to tackle the corona epidemic.
CG-CRE	a greeting from the heart to [**doctors**]$_{c1:MED,r:co\_occ}$,[**nurses**]$_{c2:MED,r:co\_occ}$, paramedics, ... who stand together to tackle the corona epidemic.
NovelTagging	[**coronavirus**]$_{c1,r:co\_occ}$ could double number of people going hungry. the risk of major interruptions to [**food supplies**]$_{c21,r:co\_occ}$ over the coming months is growing.
JointER	[**coronavirus**]$_{c1:DIS,r:ca\_eff}$ could double number of people [**going hungry**]$_{c2:FOD,r:ca\_eff}$. the risk of major interruptions to food supplies over the coming months is growing.
CG-CRE	[**coronavirus**]$_{c1:DIS,r:ca\_eff}$ could double number of people [**going hungry**]$_{c2:FOD,r:ca\_eff}$. the risk of major interruptions to food supplies over the coming months is growing.
NovelTagging	breaking one of somalia 's greatest artist ha [**died**]$_{c1,r:co\_occ}$ in london after contracting [**corona virus**]$_{c2,r:co\_occ}$ ...
JointER	breaking one of somalia 's greatest artist ha died in london after contracting corona virus ... [**pred:[]**]
CG-CRE	breaking one of somalia 's greatest [**artist**]$_{c1:PER,r:ca\_eff}$ ha died in london after contracting [**corona virus**]$_{c2:DIS,r:ca\_eff}$ ...
NovelTagging	what are the [**common**]$_{c1,r:co\_occ}$ [**symptom**]$_{c2,r:co\_occ}$. fever , sore throat ...
JointER	what are the common symptom . fever , sore throat ... [**pred:[]**]
CG-CRE	what are the common symptom . [**fever**]$_{c1:DIS,r:co\_occ}$, [**sore throat**]$_{c2:DIS,r:co\_occ}$ ...
NovelTagging	social distancing, stay home, [**naija people**]$_{c1,r:co\_occ}$ will not hear. this corona thing has just started with us in this [**country**]$_{c2,r:co\_occ}$, we ...
JointER	social distancing, stay home, naija people will not hear. this corona thing has just started with us in this country, we ... [**pred:[]**]
CG-CRE	[**social distancing**]$_{c1:GOV,r:co\_occ}$, [**stay home**]$_{c2:GOV,r:co\_occ}$, naija people will not hear. this corona thing has just started with us in this country, we ...

4.5.5 Case Study

In this section, I conduct case studies, presenting some representative public concern extraction examples, to further prove the effectiveness and validity of the proposed approach. Table 4.8 shows the outputs from three models, including NovelTagging, JointER, and the proposed CG-CRE. In the first case, both concerns and concern relation are identified incorrectly by NovelTagging, and JointER predicts nothing. By contrast, CG-CRE can extract both concerns correctly. Similar outputs are presented in the fifth and the sixth case. As for the second and third cases, NovelTagging only detects one concern correctly and cannot extract the second concern and relation. However, JointER and CG-CRE can accurately identify concerns and concern relations. JointER is not able to carry out the prediction results. In the fourth case, NovelTagging can identify only one concern correctly. JointER is able to obtain accurate predictions, but still remains to be improved in eliminating null prediction. NovelTagging is weak at extracting relations from Twitter datasets.

Based on the experimental results and case studies, I can conclude that the proposed CG-CRE model can yield better performance on both entity recognition and relation extraction than the state-of-the-art models.

4.5.6 Discussion

In this chapter, an end-to-end model is presented to simultaneously extract concern and concern relations from the social media dataset of COVID-19. GCN and Bi-LSTM are jointly combined to learn sequential and regional dependency features from tweets. In order to capture more features of model input, the influence of graph structure for concern and relation extraction is explored. The sequential and regional features from the dataset are concatenated, enabling the embedding vectors to represent rich contextual information of both concerns and relations. The proposed model is evaluated on manual-labelled and auto-labelled datasets. The experimental results show that the proposed model can outperform the existing entity and relation extraction models, which demonstrates the effectiveness of the proposed method. Furthermore, the previous methods only work on hand-crafted datasets, while the proposed model turns out to be applicable to both manual-labelled and auto-labelled datasets. Therefore, the proposed method can be easily transferred and applied to other pandemic situations, e.g., Zika, Dengue Fever, and Yellow Fever.

4.6 Summary

This chapter presented a cutting-edge, deep learning-based approach for identifying people's concerns and their related relationships through the integration of Graph Convolutional Networks, Bi-directional Long Short-Term Memory, and Concern Graphs. This model harnesses both sequential features obtained from BERT embeddings and regional features of tweets obtained from the Concern Graph module, leading to improved the concern detection and increased resistance to noise. This solution overcomes the challenge of limited manually labelled data. Thorough experiments were performed on real-world datasets, using both manually and automatically labelled tweets, to evaluate the proposed model. The results show that the proposed model outperforms state-of-the-art models in terms of F1 score.

This chapter mainly answers Research Question 1 mentioned in Chapter 1. The model and results of this chapter have been published in (J. Shi, Li, Yongchareon et al., 2022).

Chapter 5

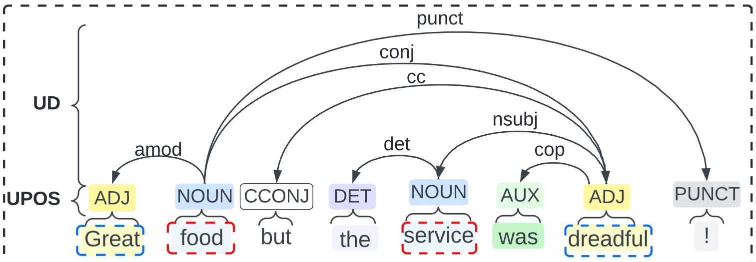
Unified Framework for Aspect-based Sentiment Analysis

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 chapter, I 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-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.

5.1 Overview

Sentiment Analysis (SA) aims to analyse people's attitudes, opinions, and sentiment distributions toward certain products, services, or opinions (B. Liu, 2012). 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) (Pontiki et al., 2014; Yan et al., 2021). 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



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 5.1: 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.

terms are "*great*" and "*dreadful*", and sentiment polarities are "*positive*" and "*negative*". Figure 5.1 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 of SA (Yan et al., 2021). In the early works, ATE (Q. Liu et al., 2016; H. Xu et al., 2018; Luo, Li, Liu, Wang & Unger, 2019; X. Wang, Xu, Sun & Tao, 2020) and OTE (C. Wu et al., 2018; H. Dai & Song, 2019; W. Wang & Pan, 2019b) have been recognised as two common tasks in ABSA. Recently, many studies have been dedicated on other sub-tasks, such as AOE (M. Wu, Wang & Pan, 2020; Veyseh et al., 2020; Z. Wu, Zhao, Dai, Huang & Chen, 2020; Y. Feng, Rao, Tang, Wang & Liu, 2021; J. Jiang, Wang & Aizawa, 2021), ALSC (X. Li, Bing, Lam & Shi, 2018; He, Lee, Ng & Dahlmeier, 2018; L. Xu, Bing, Lu & Huang, 2020; Sungheetha & Sharma, 2020), and AESC (D. Ma et al., 2018; X. Li, Bing, Li & Lam, 2019; M. Hu et al., 2019). Due to the limitation of annotated datasets, AOPE (Fan et al., 2019; H. Zhao et al., 2020; Z. Wu, Ying et al., 2020; S. Chen et al., 2020; L. Gao et al., 2021) and ASTE (Z. Wu, Ying et al., 2020; Peng et al., 2020; Yan et al., 2021; L. Xu, Li et al., 2020; C. Zhang, Li, Song & Wang, 2020; Z. Chen, Huang, Liu, Shi & Jin, 2021; Mao et al., 2021; S. Chen et al., 2021; H. Chen, Zhai, Feng, Li & Wang, 2022) 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 (Yan et al., 2021). A few recent works explored the unified model (Peng et al., 2020; Mao et al., 2021; Yan et al., 2021). 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 (Hochreiter & Schmidhuber, 1997), GRU (Cho, van Merriënboer, Bahdanau & Bengio, 2014), Transformer (Vaswani et al., 2017), and ELMo (Peters et al., 2018), have been extensively applied to ABSA (M. Yang et al., 2019; Luo, Li, Liu & Zhang, 2019; S. Chen et al., 2020; Z. Wu, Ying et al., 2020; Y. Zhang et al., 2021; Z. Chen et al., 2021; W. Li, Shao, Ji & Cambria, 2022; Z. Li, Li, Zhou & Lu, 2021). Recently, pre-trained language models, e.g., BERT (Devlin et al., 2019a), RoBERTa (Y. Liu et al., 2019), and BART (Lewis et al., 2020), achieved superior performances in various ABSA subtasks without explicit consideration of syntactic information (Du, Sun, Wang, Qi & Liao, 2020; Yan et al., 2021; Mao et al., 2021; S. Chen et al., 2021; J. Dai et al., 2021). 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 5.1. To alleviate this problem, the hierarchical tree models, e.g., TreeLSTM (Socher et al., 2013) and Graph Convolutional Network (GCN) (Kipf & Welling, 2017), 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 (Q. Liu et al., 2016; H. Chen et al., 2022; Z. Chen et al., 2021).

Although these works have achieved state-of-the-art performances, serval 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 proposed models. Instead of developing multiple models, an integrated model turns out to be more effective in solving each task of ABSA in a unified manner. **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 aspect and opinion terms extraction and improve the performance for ABSA sub-tasks (Peng et al., 2020; Mao et al., 2021; Yan et al., 2021). Although these unified frameworks are based on pre-trained language models that can capture implicit syntactic information from a sentence, the limitation still exists due to the absence of explicit syntactical features for enhancing 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 positively impact model performance (C. Zhang et al., 2019; J. Jiang et al., 2021; Z. Chen et al., 2021; Tian, Chen & Song, 2021; Liang et al., 2021; R. Li et al., 2021). For example, in Figure 5.1, 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 $\{food \leftarrow great|amod\}$, where *amod* refers to the adjectival modifier.

In this chapter, to tackle the challenges mentioned above, I 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. I further design two types of GCNs, named Semantic and Syntax GCN, which play different roles in the proposed framework. The semantic GCN is able to learn the representation via adjacency neighbourhood of context, and syntax GCN 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 a graph passes through the multilayer semantic and syntax GCNs. On top of that, the hidden features from the multilayer GCNs 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 extraction and classification tasks.

To sum up, the contributions are listed as follows:

- Firstly, I 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 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 chapter is organised as follows. Related works are reviewed in Section 5.2. In Section 5.3, relevant concepts are formally defined, and the problem is formulated. After briefly describing the proposed model in Section 5.4, the experiments and analysis of experimental results are presented in Section 5.5. Finally, the contributions and future works are summarised in Section 5.5.7.

5.2 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.

5.2.1 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 (M. Hu & Liu, 2004a). Liu et al. propose an unsupervised framework based on lifelong learning to improve opinion and aspect extraction (Q. Liu et al., 2016). 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 (H. Xu et al., 2018). Based on an Inward-Outward LSTM, a target-fused neural network model is proposed to perform target-oriented opinion word extraction (Fan et al., 2019). A transferable network is introduced for fine-grained opinion extraction in (W. Wang & Pan, 2019b). 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 (X. Li, Bing, Lam & Shi, 2018). In (C. Zhang et al., 2019), a GCN layer is designed over the dependency tree of a sentence for aspect-specific 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, 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 (L. Xu, Liu, Lai, Chen & Zhao, 2013). Similarly, Li et al. develop an LSTM-based deep multi-task learning framework that handles aspect and opinion extraction jointly (X. Li & Lam, 2017). 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 (M. Hu et al., 2019). Aspect-opinion pair extraction is first introduced in (Fan et al., 2019), 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

(S. Chen et al., 2020). 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 (Peng et al., 2020). Chen et al. adopt a multi-turn machine reading comprehension framework using BERT to extract the triplet (S. Chen et al., 2021). 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 (Mao et al., 2021). 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 (Yan et al., 2021). 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 (J. Dai et al., 2021).

5.2.2 Syntax-based Aspect Level Sentiment Analysis

In customer reviews, explicit syntactic relations are usually presented between opinion words and the corresponding aspect words (Q. Liu et al., 2016). 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 (C. Zhang et al., 2019). 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 (K. Sun, Zhang, Mensah, Mao & Liu, 2019). Veyseh et al. incorporate the syntactic structures of the sentences into LSTM and GCN models for targeted opinion word extraction (Veyseh et al., 2020). To enhance the task of aspect-opinion pair extraction, a label-ware GCN is introduced for modelling rich syntactic knowledge in (S. Wu, Fei, Ren, Ji & Li, 2021). By leveraging dependency relations and types, a type-ware GCN is developed to address ABSA sub-tasks (Tian et al., 2021). 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 (H. Chen et al., 2022). 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.

5.2.3 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 (C. Zhang et al., 2019). 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 (W. Wang et al., 2017). Zhang et al. present a novel retrieval-based attention mechanism to retrieve important semantic features for aspect extraction (C. Zhang et al., 2019). Wu et al. employ an attention layer to enhance the connection between aspect and opinion words for aspect-oriented opinion extraction (Z. Wu, Ying et al., 2020). Chen et al. propose a supervised self-attention mechanism to extract opinion entities and relations for aspect-opinion pair extraction (S. Chen et al., 2020). 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 (J. Jiang et al., 2021), 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 (Y. Feng et al., 2021). 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 chapter, 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 (R. Li et al., 2021), in this chapter, 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.

5.3 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.

5.3.1 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: Aspect Term Extraction (ATE) aims to extract all aspect terms $\{a_i|a_i \in A\}$ from a review R.

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

Definition 3: Aspect-Level Sentiment Classification (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: Aspect-oriented Opinion Extraction (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: Aspect Extraction and Sentiment Classification (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: Aspect-Opinion Pair Extraction (AOP) detects the aspect-opinion pairs $\{(a_i, o_j)|a_i \in A, o_i \in O\}$ in the review R.

Definition 7: Aspect Sentiment Triplet Extraction (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.

5.3.2 Problem Formulation

In this chapter, I 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 (ATE or OTE):

  $Y_{ATE/OTE} = \{w_i^{si}, w_i^{ei}\}_{i=1}^{|Y|}$

- Single task with oriented targets (ALSC or AOE):

  $Y_{ALSC} = \{s_i|a_i\}_{i=1}^{|Y|}$ or

  $Y_{AOE} = \{w_i^{si}, w_i^{ei}|a_i\}_{i=1}^{|Y|}$

- Pair task (AESC or AOP):

  $Y_{AESC} = \{(w_i^{si}, w_i^{ei}, s_i)\}_{i=1}^{|Y|}$ or

  $Y_{AOP} = \{(\dot{w}_i^{si}, \dot{w}_i^{ei}, \ddot{w}_i^{si}, \ddot{w}_i^{ei})\}_{i=1}^{|Y|}$

- Triplet task (ASTE):

  $Y_{AOP} = \{(\dot{w}_i^{si}, \dot{w}_i^{ei}, \ddot{w}_i^{si}, \ddot{w}_i^{ei}, s_i)\}_{i=1}^{|Y|}$,

  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.

5.4 Unified Syntax-Enhanced Network

The proposed framework involves five layers: input layer, embedding layer, MSS-GCN fusion encoder layer, decoder layer, and output layer. Figure 5.2 shows the overview architecture of the proposed network.

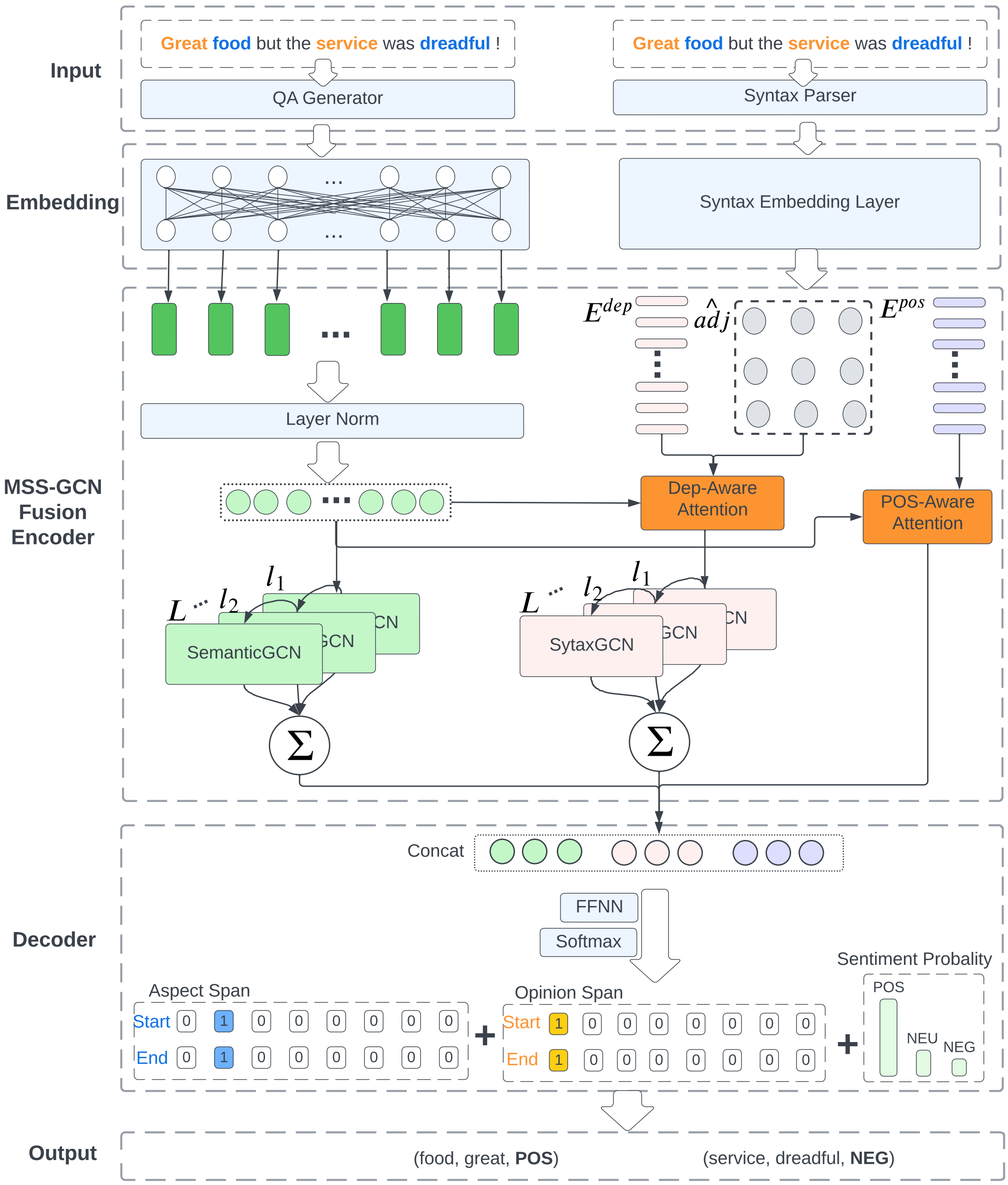


Figure 5.2: The overview architecture of the proposed network. This shows an example of a triplet where the aspect and opinion terms are *food*, and *great*, respectively. The corresponding sentiment polarity is *positive*.

5.4.1 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 (S. Chen et al., 2021). 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 the aspect food/service and opinion great/dreadful? A: POS/NEG}*.

5.4.2 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 (H. Xu et al., 2019). Specifically, the pre-trained BERT model is less task awareness and domain awareness (Y. Lin, Tan & Frank, 2019; H. Xu et al., 2019). Thus, the domain BERT model is applied in the proposed framework. Given a sentence $\{w_1, w_2, ..., w_n\}$, semantic embedding $E^s = \{e^s_i \in \mathbb{R}^{d^s}\}^n_{i=1}$ can be obtained using Equation 5.1, 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 self, a self-loop adjacency matrix $\hat{J} = \{\hat{adj}_{i,j}\}_{n \times n}$ is calculated using Equation 5.2. A dependency matrix $T = \{t_{i,j}\}_{n \times n}$ is utilised to record

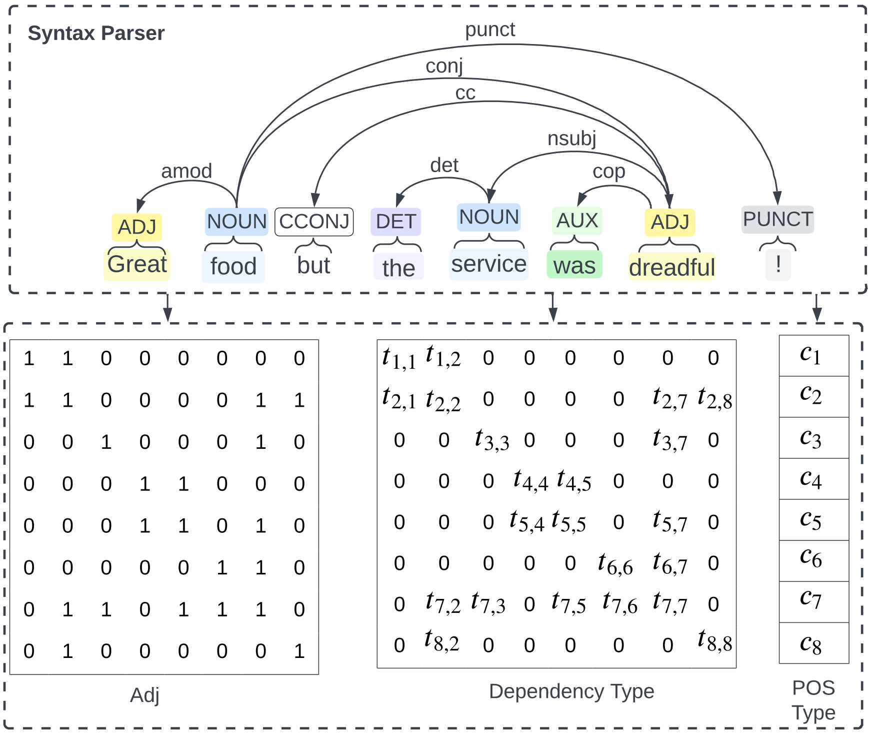


Figure 5.3: The syntax input generator.

the dependency types, and the embedding of dependency types 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 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 5.3 illustrates the example of matrix generation for dependency relation, dependency type, and POS category.

$$E^s(\{e_1^s, e_2^s, ..., e_n^s\}) = DomBERT(\{w_1, w_2, ..., w_n\}) \tag{5.1}$$

$$\hat{J} = J + I, \tag{5.2}$$

where I describes n x n identity matrix.

5.4.3 MSS-GCN fusion encoder 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 5.3, layer normalisation is applied to the contextual representations from domain BERT (Ba, Kiros & Hinton, 2016).

$$\hat{E}^s(\{\hat{e}_1^s, \hat{e}_2^s, ..., \hat{e}_n^s\}) = LayerNorm(E^s)(\{e_1^s, e_2^s, ..., e_n^s\}) \tag{5.3}$$

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 $\alpha_{i,j}^s$ is computed by Equations 5.4 - 5.6.

$$\dot{e}_i^s = \dot{W}\hat{e}_i^s + \dot{b} \tag{5.4}$$

$$\ddot{e}_j^s = \ddot{W}\hat{e}_j^s + \ddot{b} \tag{5.5}$$

$$\alpha_{i,j}^s = \frac{exp(\dot{e}_i^s \ddot{e}_j^s)}{\sum_{j=1}^n exp(\dot{e}_i^s \ddot{e}_j^s)}, \tag{5.6}$$

where $\dot{W}$ and $\ddot{W}$ denote trainable weight matrices. $\dot{b}$ 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 5.7.

$$h_i^{s(l^s+1)} = \sigma(\sum_{j=1}^n \alpha_{i,j}^s (W^{s(l^s+1)} h_i^{s(l^s)} + b^{s(l^s+1)})), \tag{5.7}$$

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 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 5.8.

$$\begin{aligned} h_i^{d(l^d+1)} = \sigma(\sum_{j=1}^{n} \alpha_{i,j}^{d(l^d+1)} (W_1^{d(l^d+1)} h_i^{d(l^d)} + W_2^{d(l^d+1)} e_i^{d(l^d+1)} + \\ W_3^{d(l^d+1)} \hat{e}_i^{s(l^d+1)} + b^{d(l^d+1)})), \end{aligned} \tag{5.8}$$

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 5.9 - 5.11. The details of dep-aware attention are shown in Figure 5.4.

$$\hat{e}_i^{d(l^d+1)} = W^{d(i)}[h_i^{d(l^d)}; e_i^{d(l^d+1)}; \hat{e}_i^{s(l^d+1)}] + b^{d(i)} \tag{5.9}$$

$$\hat{e}_j^{d(l^d+1)} = W^{d(j)}[h_j^{d(l^d)}; e_j^{d(l^d+1)}; \hat{e}_j^{s(l^d+1)}] + b^{d(j)} \tag{5.10}$$

$$\alpha_{i,j}^{d(l^d+1)} = \frac{\hat{adj}_{i,j} exp(\hat{e}_i^{d(l^d+1)} \hat{e}_j^{d(l^d+1)})}{\sum_{j=1}^{n} \hat{adj}_{i,j} exp(\hat{e}_i^{d(l^d+1)} \hat{e}_j^{d(l^d+1)})} \tag{5.11}$$

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 computation in attention. Specifically, in each $l^p + 1$-th layer, the POS-aware attention score is formulated in Equation 5.12 through

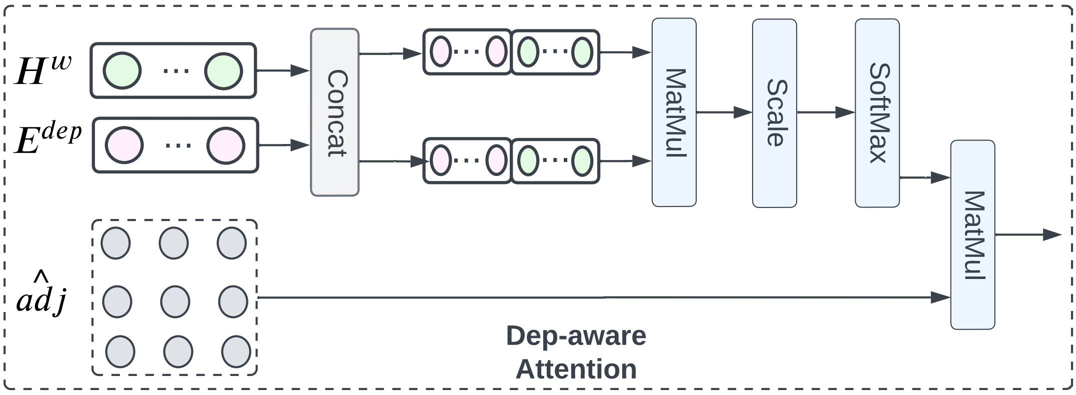


Figure 5.4: The illustration of Dep-aware attention.

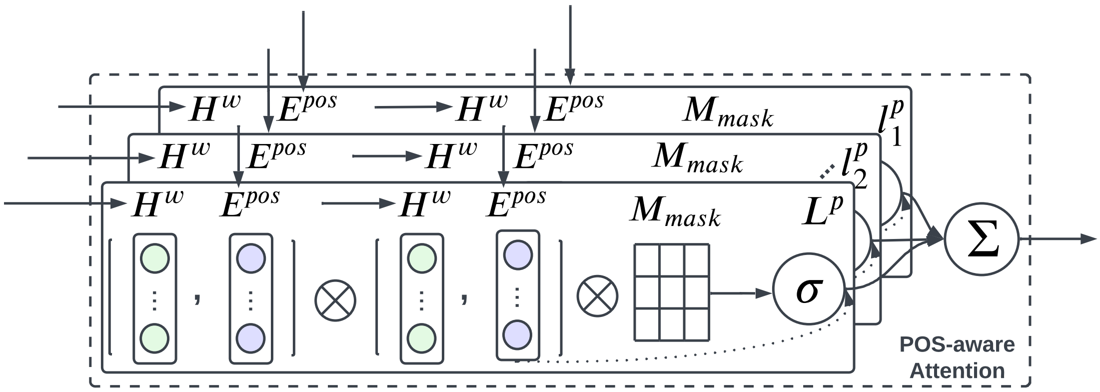


Figure 5.5: The illustration of POS-aware attention.

integrating semantic embedding and POS embedding. Figure 5.5 presents the process of POS-aware attention.

$$\alpha_{i,m}^{p(l^p+1)} = \frac{exp(W^p[h_m^{p(l^p)}; \hat{e}_m^s])}{\sum_{k=i-D}^{i+D} exp(W^p[h_k^{p(l^p)}; \hat{e}_k^s])} \tag{5.12}$$

$$h_i^{p(l^p+1)} = \sum_{m=i-D}^{i+D} \alpha_{i,m}^{p(l^p+1)} h_i^{p(l^p)} \tag{5.13}$$

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.

5.4.4 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 concatenated through a weighted average as in Equations 5.14 - 5.15. The same operation is applied to the output of POS attention in Equation 5.16.

$$\hat{h}_i^s = \sum_{l^s=1}^{L^s} \theta^{s(l^s)} h_i^{s(l^s)} \tag{5.14}$$

$$\hat{h}_i^d = \sum_{l^d=1}^{L^d} \theta^{d(l^d)} h_i^{d(l^d)} \tag{5.15}$$

$$\hat{h}_i^p = \sum_{l^p=1}^{L^p} \theta^{p(l^p)} h_i^{d(l^p)}, \tag{5.16}$$

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 5.17.

$$p(y_i^c) = softmax(W^c[\hat{h}_i^s; \hat{h}_i^d; \hat{h}_i^p] + b^c), \tag{5.17}$$

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

5.4.5 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 3 shows the process of triplet prediction.

Algorithm 3 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 $\hat{Y}$

5.4.6 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 5.18.

$$\mathcal{L} = \sum_{c \in C} \theta^c \mathcal{L}^c, \tag{5.18}$$

where $C = \{startindex, endindex, sentimentpolarity\}$ presents the task category. θ^c means the regularisation coefficients to balance the learning between different tasks. For each task category, the negative log-likelihood loss is formulated in Equations 5.19 - 5.21.

$$\mathcal{L}^{si} = -\sum_{i=1}^{n} p(y_i^{si}) log(\hat{p}(y_i^{si})) \tag{5.19}$$

$$\mathcal{L}^{ei} = -\sum_{i=1}^{n} p(y_i^{ei}) log(\hat{p}(y_i^{ei})) \tag{5.20}$$

$$\mathcal{L}^{sp} = -\sum_{t=1}^{T} p(y_t^{sp}) log(\hat{p}(y_t^{sp})), \tag{5.21}$$

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.

5.5 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.

5.5.1 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 5.1 - 5.3. All benchmark datasets originate from the Semantic Evaluation

(SemEval) workshops (Pontiki et al., 2014, 2015, 2016) 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 $\mathbb{D}_1$ (W. Wang et al., 2016, 2017). In $\mathbb{D}_2$, the triplet, including aspect term, opinion term, and sentiment polarity, is labelled to address the sub-task ASTE (Peng et al., 2020). As a revised variant dataset of $\mathbb{D}_2$, some missing triplets are corrected in $\mathbb{D}_3$ (L. Xu, Li et al., 2020). For each group dataset, the ratio of training, validating, and testing datasets are shown in Tables 5.1 - 5.3. Besides, I further list the ratio of each sentiment polarity to all polarities.

Table 5.1: The statistics of datasets $\mathbb{D}_1$. Notations #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 5.2: 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 5.3: 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%

Evaluation Metrics

In this chapter, three standard evaluation metrics, i.e., precision (P), recall (R), and F1 score (F1), are adopted to evaluate the proposed model. Specifically, P, R, and F1 are defined in Equations 5.22 - 5.24.

$$P = \frac{T_P}{T_P + F_P} \tag{5.22}$$

$$R = \frac{T_P}{T_P + F_N} \tag{5.23}$$

$$F1 = \frac{2 * P * R}{P + R}, \tag{5.24}$$

where T_P denotes the correct number of predictions for the element in each sub-task. For example, it is necessary to correctly predict all elements in (a, o) and (a, o, s) for sub-task AOP and ASTE, respectively. F_P refers to the number of prediction errors where additional items are predicted beyond the actual targets. F_N indicates the number of missed predictions that the number of predicted targets is lower than the actual number. For AOP and ASTE, F_N denotes the number of predicted pairs and triplets is less than the actual number in the dataset.

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 NLP Toolkit [2](i.e., Stanza

[1]https://pytorch.org/
[2]https://stanfordnlp.github.io/stanza/

(Qi, Zhang, Zhang, Bolton & Manning, 2020)). 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 (Diederik & Jimmy, 2015) is applied with an initial learning rate of 1e-3. The epoch is set to 40, and the batch size is 10.

5.5.2 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 5.4 presents the core module, selected datasets, and solved ABSA sub-tasks for each baseline.

- ***DP***(Qiu et al., 2011) 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***(X. Zhang, Jiang, Peng, Tu & Goldwasser, 2017) is an end-to-end neural auto-encoder model for sequential structured prediction problems. The

[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 consists of an encoder, a CRF (Lafferty et al., 2001) model enhanced by deep neural networks, and a decoder, a generative model to reconstruct the input.

- ***LSTM-RNN***(P. Liu et al., 2015) applies Recurrent Neural Network (RNN) 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 model can be improved even further by incorporating some linguistic features, e.g., POS and phrasal information, into RNNs.

- ***RNCRF***(W. Wang et al., 2016) 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***(C. Zhang et al., 2020) 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+***(X. Li et al., 2019) 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+***(H. Dai & Song, 2019) 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***(Peng et al., 2020) 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+***(W. Wang et al., 2017) 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***(M. Hu et al., 2019) 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.

- ***SPAN-ASTE***(L. Xu, Chia & Bing, 2021) 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***(He, Lee, Ng & Dahlmeier, 2019) 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***(Z. Chen & Qian, 2020) 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***(L. Xu, Li et al., 2020) 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***(Mao et al., 2021) solves all sub-tasks of ABSA in a unified end-to-end framework by joint training two BERT Machine Reading Comprehension (MRC) models with parameter sharing.

- ***BMRC***(S. Chen et al., 2021) transforms ASTE into a multi-turn machine reading comprehension problem, and comprehensively identifies triplets by a bidirectional MRC structure.

- ***BART-ABSA***(Yan et al., 2021) converts all ABSA sub-tasks into a unified generative formulation, where the pre-trained model BART (Lewis et al., 2020) is utilised to solve sub-tasks in an end-to-end framework.

Table 5.4: 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
SPAN-ASTE	BERT	$\mathbb{D}_3$	Y	Y	Y	N	N	N	Y
IMN-BERT	BERT	$\mathbb{D}_1$	Y	Y	Y	N	Y	N	N
RACL-BERT	BERT	$\mathbb{D}_1$	Y	Y	Y	N	Y	N	N
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
BART-ABSA	BART	$\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	Y	Y	Y	Y
Ours	Domain-BERT	$\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$	Y	Y	Y	Y	Y	Y	Y

5.5.3 Experimental Results and Model Analysis

In this subsection, I present extensive experimental results using three groups of datasets, i.e., $\mathbb{D}_1$, $\mathbb{D}_2$, and $\mathbb{D}_3$. Then, I 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 5.5 - 5.7.

Firstly, Table 5.5 presents the comparison results for 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 the proposed model. Thus the performance is close to the models based on BERT, e.g., SPAN-BERT, RACL-BERT, and DMRC. As another unified framework for ABSA, BART-ABSA obtains competitive results on *Res14* and *Lap14*

Table 5.5: 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	-	-
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**

for ATE and OTE sub-tasks. However, the proposed model performs better in other sub-tasks due to the assistance of domain knowledge and syntactic information.

Secondly, the sub-tasks, including OTE, AESC, AOP, and ASTE, are addressed in dataset $\mathbb{D}_2$, and the results are demonstrated in Table 5.6. It is evident that some baselines can achieve better results for OTE than the proposed model. However, the proposed model can outperform all baselines on all datasets for AESC, AOP, and ASTE sub-tasks. Because OTE is a single target extraction task, the dependency relation does not bring many contributions to such a task. For pair and triplet extraction tasks, the proposed model performs the best of all, explicitly demonstrating the effectiveness of the proposed network in capturing direct and indirect interactions among targets. Although the proposed model's performance for OTE on Res14 and Res16 appears slightly lower than RACL-BERT, 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 (H. Xu et al., 2019). Therefore, domain BERT is applied in the proposed model to address such challenges and improve the performance of

Table 5.6: 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
	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
Lap14	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
	BART-ABSA	-	-	-	-	-	68.17	-	-	66.11	-	-	57.59
	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
	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**
	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
Res14	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
	BART-ABSA	-	-	-	-	-	78.47	-	-	77.68	-	-	72.46
	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
	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**
	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
Res15	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
	BART-ABSA	-	-	-	-	-	69.95	-	-	67.98	-	-	60.11
	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
	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**
	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
Res16	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
	BART-ABSA	-	-	-	-	-	75.69	-	-	77.38	-	-	69.98
	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
	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**

sentiment classification.

Thirdly, the proposed model is compared with the latest baselines for ASTE on dataset $\mathbb{D}_3$, shown in Table 5.7. It can be observed that SPAN-ASTE shows the 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, the proposed model demonstrates a consistent improvement in terms of F1 score on all other datasets. I note that either GloVe (Pennington et al., 2014) or Word2Vec (Mikolov, Sutskever, Chen, Corrado & Dean, 2013) is applied to obtain word embeddings in Li-Unified+, RINANTE+, and TS, and there is a big gap between their performances and the ones with BERT-based models. Thus, the pre-trained BERT can capture more informative contextual features than GloVe and Word2Vec. With the assistance of domain BERT and syntactic structure, the proposed model is able to upgrade the performance further.

Table 5.7: 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
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**

5.5.4 Further Analysis

To better understand the embedding difference between BERT and domain BERT, I plot the distributions of aspect-opinion and sentiment embedding in Figures 5.6 and 5.7. As can be observed from Figure 5.6, 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 5.7. The embedding points 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.

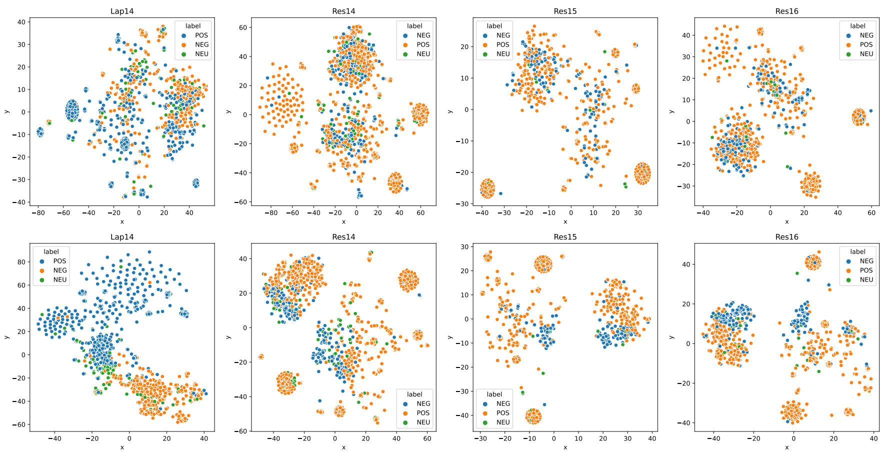


Figure 5.7: Sentiment embedding analysis on dataset $\mathbb{D}_2$. The first row is aspect-opinion embeddings from BERT, and the second row is from domain BERT.

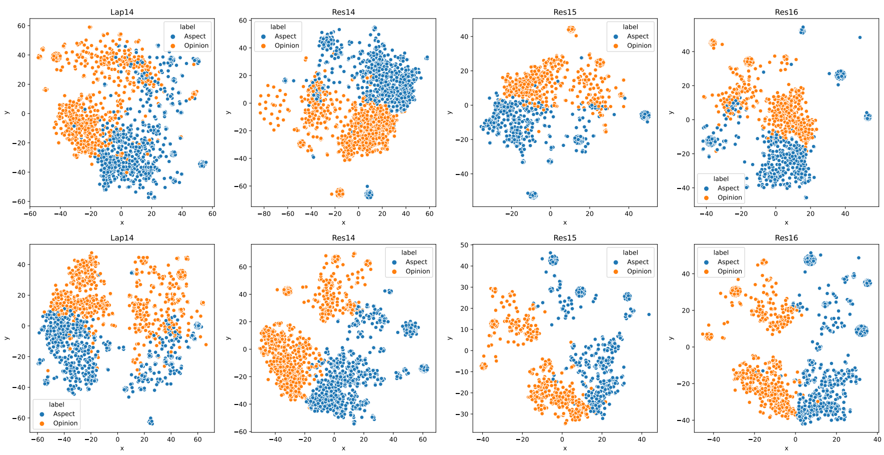


Figure 5.6: Aspect-Opinion embedding analysis on dataset $\mathbb{D}_2$. The first row is aspect-opinion embeddings from BERT, and the second row is from domain BERT.

5.5.5 Ablation Study

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

- +*BERT*, the vanilla BERT (Devlin et al., 2019a) is utilised for encoding the context;
- +*DomainBERT*, the context embedding is obtained by two domain BERT models;
- +*MSS*, the module SyntaxGCN with dep-aware and POS-aware attention are incorporated in the proposed model;
- *Full*, all components are applied in the proposed model.

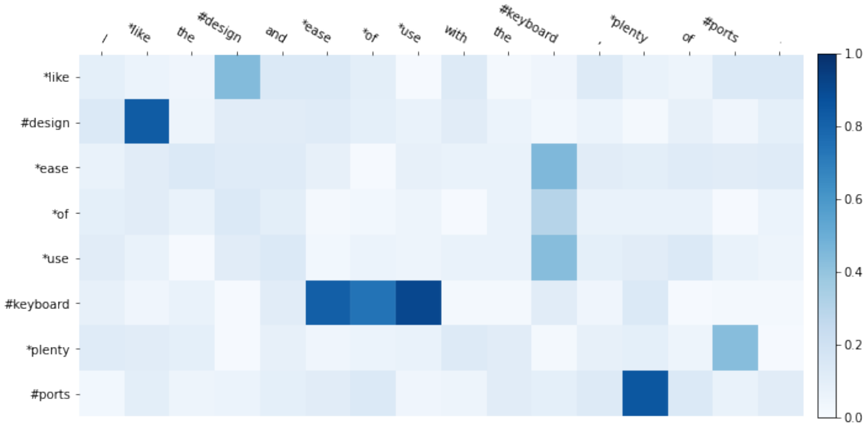
The experimental results are shown in Table 5.8. By comparing models with different embeddings, I notice that domain BERT contributes to the performance improvement for ABSA sub-tasks, which proves that the learned knowledge in domain BERT is beneficial for sub-tasks of ABSA. ABSA turns out to be a very domain-specific task. Meanwhile, the model can achieve better performance by adding component attention to the four datasets, demonstrating the validity of leveraging the syntactic structure via syntax GCN. It can be observed from Table 5.8 that the proposed model outperforms all variants in terms of F1 score. Therefore, both domain knowledge and syntactic structure information can enhance semantic embeddings, capture the syntactic correlations between terms, and explain the performance improvements accordingly.

Table 5.8: Ablation study of the proposed model on Dataset $\mathbb{D}_3$.

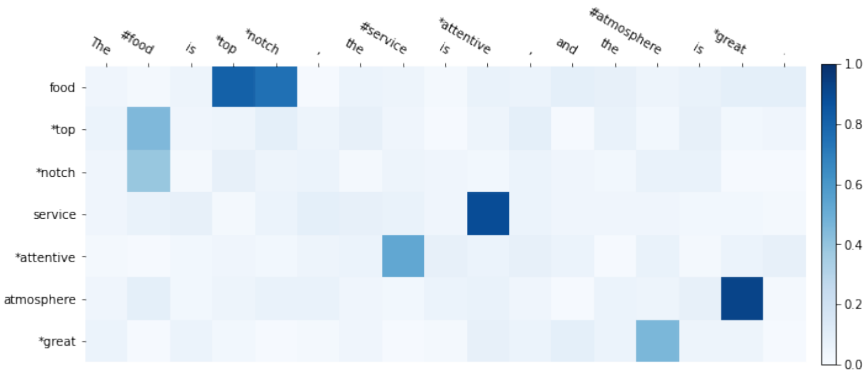
Datasets		Lap14			Res14			Res15			Res16		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
AESC	+BERT	60.76	62.22	61.48	71.70	64.74	68.04	67.01	60.37	63.52	64.94	72.60	68.55
	+DomainBERT	74.99	56.33	64.34	75.52	66.70	70.84	68.89	66.82	67.84	70.06	73.49	71.73
	+SytaxGCN	73.03	60.91	66.42	71.61	75.92	73.70	73.14	66.35	69.58	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
	+SytaxGCN	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
	+SytaxGCN	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**

5.5.6 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 5.9 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. The proposed 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



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



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

Figure 5.8: Visualisation of dep-aware attention mechanism. Marker * indicates the aspect term, and marker # refers to the opinion term.

long usually expresses a positive opinion of the corresponding aspect. The proposed 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 syntax GCN with dep-aware attention, which benefits the extraction of both opinion terms *SMALL* and *below average*. Figure 5.8 shows two examples of attention scores from the proposed dep-aware attention mechanism. For both sentences, the aspect terms have high attention scores with opinion terms, implying that the proposed attention mechanism can capture important information and improve the performance of target extraction.

Table 5.9: 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)]

5.5.7 Discussion

In this chapter, 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 more informative syntactic structure, further enhancing the semantic representations. Moreover, a couple of multi-layer attention mechanisms are designed 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 the proposed model. Finally, a case study is presented to exhibit the performance of the proposed network.

5.6 Summary

In this chapter, I introduced a unified framework for addressing all defined sub-tasks in aspect-based sentiment analysis. The framework combines a multi-layer semantic model based on graph convolutional networks and a multi-layer syntax model to capture semantic connections and explicit dependency relations, respectively. The semantic features learned by the semantic model are integrated with the syntax model to provide semantic guidance and enhance the learning of comprehensive syntactic representations. The framework also features two innovative attention mechanisms for modelling dependency relations, types, and part-of-speech tags for detecting aspect and opinion term boundaries. Different from the conventional syntax-based models, the proposed method integrates semantic and syntax through a joint learning mechanism. I conduct extensive experiments to verify the effectiveness of the proposed unified framework. The results of these experiments, conducted on four sets of real-world datasets, clearly demonstrate the superiority of the proposed framework over several baseline models.

This chapter mainly answers Research Question 2 mentioned in Chapter 1. The research work of this chapter has been published in (J. Shi, Li, Bai, Yang & Jiang, 2022).

Chapter 6

Soft Prompt-Tuning for Cross-Domain Sentiment Analysis

Aspect term extraction is a fundamental task in fine-grained sentiment analysis, which aims at detecting customer opinion targets from reviews on products or services. The traditional supervised models can achieve promising results with annotated datasets, however, the performance dramatically decreases when they are applied to the task of cross-domain aspect term extraction. Existing cross-domain transfer learning methods either directly inject linguistic features into Language models, making it difficult to transfer linguistic knowledge to the target domain, or rely on the fixed predefined prompts, which is time-consuming to construct the prompts over all potential aspect term spans.

To resolve the limitations, I propose a soft prompt-based joint learning method for cross-domain aspect term extraction in this chapter. Specifically, by incorporating external linguistic features, the proposed method learns domain-invariant representations between source and target domains via multiple objectives, which bridges the gap between domains with varied distributions of aspect terms. Further, the proposed method interpolates a set of transferable soft prompts consisting of multiple learnable vectors that are beneficial to detect aspect terms in the target domain. Extensive experiments are conducted on the benchmark datasets and the experimental results demonstrate the effectiveness of the proposed method for cross-domain aspect terms extraction.

6.1 Overview

To develop specialist knowledge for business development, it's crucial to rapidly understand customer complaints or requirements by analysing their feedback. As an outstanding method of review analysis, aspect-based sentiment analysis (ABSA) aims to extract the aspect and opinion terms, and identify their corresponding sentiments from customer reviews(B. Liu, 2012; Pontiki et al., 2016). In this chapter, I focus on a crucial sub-task for ABSA, named aspect term extraction (ATE), which is to identify opinion targets from customer review sentences. For the example in Figure 6.1, the task is expected to detect the aspect term *Keyboard* from the sentence "*Keyboard responds well to presses*".

Recently, ATE has been well studied in literature with the emergence of pre-trained language models (PLMs), such as BERT (Devlin et al., 2019a), BART (Lewis et al., 2020), T5 (Raffel et al., 2020), GPT v1-3 (Radford et al., 2018, 2019; Brown et al., 2020). By fine-tuning PLMs, remarkable results are achieved for ATE task (H. Xu et al., 2018; X. Wang et al., 2020; Wan et al., 2020; L. Gao et al., 2021; Venugopalan & Gupta, 2022). However, a large number of annotated data is required to fine-tune PLMs for downstream tasks. The data annotation work is labour-intensive and time-consuming, which can lead to the lack of training data for fine-tuning PLMs (Le Scao & Rush, 2021). Moreover, fine-tuning PLMs has become more and more difficult for real-world applications due to the exponentially increased trainable parameters. To overcome the fine-tuning challenges, a new learning method is designed, named prompt tuning, to

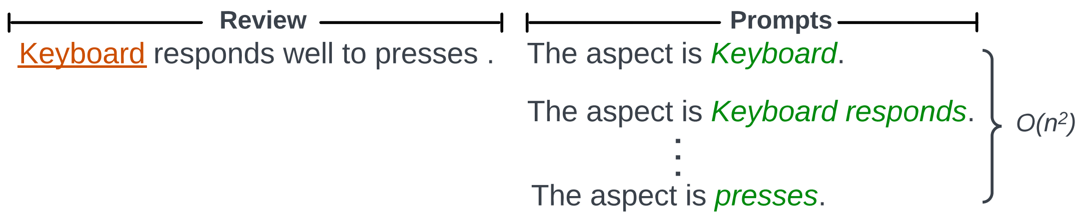


Figure 6.1: The traditional inputs of the prompt tuning model for aspect term extraction.

reformulate NLP tasks as cloze-style question answering (Lester, Al-Rfou & Constant, 2021). Without updating any of the parameters of PLMs, the prompt-based learning has achieved outstanding results on many NLP tasks (e.g., relation classification (X. Chen et al., 2022), sentiment classification (C. Li et al., 2021), NER (X. Chen et al., 2021)). By manually designing prompts, similar attempts have been made on aspect term extraction to detect aspect term from each text span in a review (T. Gao et al., 2022; H. Li et al., 2022; C. Li et al., 2021). As shown in Figure 6.1, to extract aspect term *Keyboard*, some prompts are designed using the template *The aspect is* __. Despite the success of fine-tuning and prompt-tuning methods, both of them suffer from the domain challenges. For PLMs in fine-tuning methods, they are pre-trained on universal datasets without specific domains, which leads to task-agnostic and poor performance for domain adaption (H. Xu et al., 2019). For prompt-based learning methods, they suffer from high costs in enumerating all possible spans of aspect terms, and the existing models fail to achieve a robust performance on cross-domain datasets due to the varied distributions of aspect terms in different domains and the complexity of constructing prompts.

To address the aforementioned challenges in the task of cross-domain aspect term extraction, I propose a joint learning method, which is the first to use soft prompt integrating with transferable linguistic knowledge to solve the domain adaptation problem. The designed soft prompts facilitate pre-trained language models better fit the aspect term distributions in the target domain, which reduce the high cost due to enumeration of all possible aspect term spans. Following, the linguistic knowledge learning module is designed to learn the domain-invariant representation across domains. By appending learnable prompts and linguistic knowledge with context features, the prompt method can generate better token representations to further improve the performance of cross-domain ATE.

The main contributions of this chapter can be summarised as follows:

- To the best of my knowledge, it's the first to resolve cross-domain ATE task by a soft prompt-based joint learning method.
- The learnable prompts are designed on multiple source domains to enable efficient knowledge transfer.
- The experimental results demonstrate that the proposed method can outperform the state-of-the-art fine-tuning and prompt-tuning models.

The remainder of this chapter is organised as follows. In Section 6.2, related works are reviewed in the cross-domain aspect term extraction. Section 6.3 formally defines the relevant concepts and formulates the problem. The proposed method is introduced in Section 6.4. The experimental work and results are presented and discussed in Section 6.5. Finally, the conclusions and future work are described in Section 6.5.6.

6.2 Related Work

Aspect term extraction is a fine-grained sentiment analysis task, which receives a lot of attention. However, only a few studies attempt to focus on domain adaptation for ATE. Cross-domain ATE aims to transfer the learned knowledge from the source domain to the target domain which labelled data is limited for ATE task. Due to the complexity of this task and the scarcity of labelled data in target domains, cross-domain ATE has become a challenging task. The existing methods can be grouped into three categories: neural network-based model, language model-based model, and prompt-based model.

6.2.1 Neural Network-based Model

Early research works about the cross-domain ATE mainly focus on hand-crafted domain-independent features and neural network models (Y. Ding et al., 2017). Jakob et al. formulate the ATE problem as an information extraction task, and propose a Conditional Random Field (CRF) based method for single- and cross-domain ATE (Jakob & Gurevych, 2010). Chernyshevich designs a CRF-based system, which is trained on a mixture of annotated training data, to detect aspect terms on all domain-specific test datasets (Chernyshevich & Belarus, 2014). For CRF-based methods, they cannot work well if training datasets are from different domains from the test datasets. To overcome this problem, Ding et al. propose a long short-term memory network (LSTM) based method by utilising the domain-independent syntactic rules (Y. Ding et al., 2017). To bridge the gap between different domains, domain-invariant dependency relations are used as pivot information to reduce domain shift by a novel recursive neural network (W. Wang & Pan, 2018). In the following research, Wang et al. extend the previous work, in which word representations and syntactic head relations are fed into a conditional domain adversarial network(W. Wang & Pan, 2019a). In another study, Wang et al. exploit local and global memory interactions of an interactive memory network to capture intra-correlations among aspect or opinion terms themselves, as well as between aspect and opinion terms (W. Wang & Pan, 2019b). The auxiliary task and domain adversarial networks are utilised to align source and target space for cross-domain ATE. Marcacini et al. present a transductive learning method to combine features of labelled aspect terms, unlabelled aspect terms, and linguistic information from both source and target domains (Marcacini et al., 2018). The proposed method can overcome the issue of model inconsistency for cross-domain ATE due to different feature spaces. To reduce the reliance on external linguistic resources, an adversarial learning method is presented to learn an alignment weight for each word by aligning the inferred correlation vectors of aspect and opinion terms (Z. Li et al., 2019). Despite the outstanding performance, neural network-based methods fail to obtain satisfactory quality of domain-invariant features and exploit the significant supervision signals in the target domains, which leads to low precision results.

6.2.2 Language Model-based Model

Recent research works found that fine-tuning language models with sophisticated task-specific layers can obtain word sense and geometrical dependency parse relations, which benefit the cross-domain ATE task (Hewitt & Manning, 2019). Pereg et al. incorporate external linguistic information into the language model with a self-attention mechanism for cross-domain ATE (Pereg et al., 2020). The proposed method is able to leverage the intrinsic knowledge of language models with externally introduced syntactic features to bridge the gap between source and target domains. Based on BERT, Gong et al. propose an end-to-end framework integrating feature-based adaptation and instance-based adaptation, which significantly improves the performance of language model for ATE (Gong et al., 2020). Anand et al. apply evolutionary approach to automatically learn linguistic patterns of aspect words, which mitigate the problem of manual engineering pattern rules (Anand & Mampilli, 2021). Mampilli et al. combine language models with attention mechanism for ATE, and this method achieves good results in-domain and unseen-domain datasets (Mampilli & Anand, 2022). Li et al. propose a new generative cross-domain data augmentation framework, which exploits the annotated data from source domain to generate data in target domain for ATE model training (J. Li et al., 2022). To solve the model extensibility and robustness on target domain datasets, Howard et al. introduce a novel method to automatically construct domain-specific knowledge graphs of aspect terms, and inject features from these graphs into language models for ATE in target domains (Howard et al., 2022). Klein et al. utilise syntactic relations connecting opinion and the related aspect words to transfer learned knowledge from language model (Klein et al., 2022). Their analyses and experiments prove that the syntactic relations transfer well across domains. To transfer knowledge of aspect terms and sentiment, Dong et al. propose a syntax-base BERT to capture domain-invariant features. However, all there language model-based methods rely heavily on annotated resources, the performance of fine-tuning language models may be unstable on small-scale data. Moreover, most existing methods only integrate the linguistic features directly into language models, which cannot achieve word-level adaption for aspect extraction.

6.2.3 Prompt-based Model

To address the learning challenges caused by increasing size of LMs, prompt-based methods are proposed to leverage language prompts and task descriptions as context to make ABSA similar to language modelling. Early studies explore hard templates, which are defined manually for ABSA tasks in a single domain. Li et al are the first to incorporate prompt-based model for aspect-based sentiment analysis subtasks, in which sentiment knowledge prompts are constructed by integrating features from aspects, opinions, and polarities (C. Li et al., 2021). Gao et al. introduce a unified generative framework to solve different ABSA tasks by controlling the type of task prompts (T. Gao et al., 2022). By assembling prompts of simple tasks, their method can transfer learned knowledge to difficult tasks. Li et al. propose a prompt-based teacher-student network to alleviate the problem of over-fitting existing in the basic prompt-based models (H. Li et al., 2022). Ben et al. present an example-based prompt learning method, which can be applied to unseen domains with multiple tasks, namely rumour detection, multi-genre natural language inference, and aspect prediction (Ben-David et al., 2022). These methods suffer from error propagation induced by entity span detection, high cost due to enumeration of all possible text spans, and omission of inter-dependencies among token labels in a sentence. However, domain knowledge is required to design a prompt manually. Therefore, soft prompts are constructed to allow LMs to effectively perform specific tasks, which are several learnable vectors instead of human-interpretable natural language.

Wu et al. adopt soft prompts instead of fixed predefined templates to learn different representations for different domains, then a novel domain adversarial training mechanism to learn domain-invariant features between the source domain and target domain for sentiment classification task (H. Wu & Shi, 2022). Asai et al. introduce a multi-task language model tuning method that transfers knowledge across different tasks via the soft prompts (Asai et al., 2022). Such a model is highly parameter-efficient and achieves promising performance using knowledge from high-resource datasets for sentiment classification and other NLP tasks. The existing hard and soft prompt-based methods either focus on a single domain or can be only applied to sentiment classification instead of aspect term extraction.

In this chapter, to alleviate the challenges of cross-domain aspect term extraction in the existing models, a joint learning method is proposed to integrate high-quality transferable knowledge from source domains via a mixture of trainable soft prompts and domain-invariant and learnable linguistic features. Different from the previous works, the proposed method is the first work that incorporates soft prompts into joint training to solve the cross-domain ATE problem. The soft prompts can overcome the time-consuming issue caused by hard prompts for enumerating the prompt queries over all potential aspect spans. The learnable linguistic features can serve as an enhancement component to bridge the gap between different domains and further capturing domain-invariant features for ATE task. The proposed method enables efficient knowledge transfer from source domains and achieves outstanding performance on multiple datasets for cross-domain ATE. Furthermore, the analysis of experimental results shows that the soft prompts and learnable syntactic features largely contribute to performance improvements.

6.3 Preliminaries

In this section, the formal definitions related to cross-domain ATE are presented, and then the problem is formally formulated based on these definitions.

6.3.1 Problem Formulation

Formally, the proposed method formulates the task of cross-domain aspect term extraction as a sequence tagging problem. Two domain datasets are given, $\mathbb{D}_s$ and $\mathbb{D}_t$ which represent the source and target domain, respectively. For the source domain dataset, $\mathbb{D}_s = \{S_s^i, y_s^i\}_{i=1}^{N_s}$ are N_s annotated sentences, where S_s^i is the ith sentence, $y_s^i \in \{B, I, O\}$ denotes the corresponding aspect label. In target domain dataset, $\mathbb{D}_t = \{S_t^i\}_{i=1}^{N_t}$ consists of N_t unlabelled sentences, where S_t^i indicates the ith sentence. The goal of cross-domain ATE is to learn a function, which can learn both in-domain and domain-invariant knowledge between source and target domain to better predict token-level labels on the test set from the target domain.

6.4 Soft Prompt-Based Joint Learning Model

In this section, I first describe the overview of the proposed method. Then I introduce each module from bottom to up in the whole architecture. Finally, I present the learning objective for cross-domain ATE.

The overall architecture for the proposed feature-based domain adaptation component is shown in Figure 6.3. Together with word and syntax embeddings, generated prompts are encoded as features and are fed into one learning layer. The output representations are as input of Softmax layer. Except for the aspect term extractor, a syntax learning module is designed to learn structural correspondence between domains. Each module is described in the following sub-sections.

6.4.1 Input Embedding

Given a sentence $s = \{w_1, w_2, ..., w_n\}$ with n words, the word sequences are converted into continuous embedding $E_s = \{e_1, e_2, ..., e_n\}$. For each embedding e_i, it consists of three type embeddings: (a) word embedding e^w is obtained via pre-trained language model by Equation 6.1. (b) syntax embedding e^{pos} is calculated in Equation 6.2. To leverage the domain-invariant features more effectively, 25% of original POS tags are randomly replaced with a special token [MASK], and a syntax learning module is designed to predict the masked POS tags. (c) soft prompt embeddings are computed in Equation 6.3. Inspired by previous work on prompt features (Ziser & Reichart, 2018; Ben-David, Rabinovitz & Reichart, 2020; Ben-David, Oved & Reichart, 2021), Mutual Information is applied to automatically extract prompts. To select prompts that are related to all source domains, the Euclidean distance is computed on T5 embeddings of prompts and the aspect tokens to generate m features for each training input.

$$e^w = T5(\{w_1, w_2, ..., w_n\}) \tag{6.1}$$

$$e^{pos} = T5(\{t_1, [MASK], ..., t_n\}) \tag{6.2}$$

$$e^p = T5(\{p_1, ..., p_m\}) \tag{6.3}$$

6.4.2 Soft Prompt Learning

Prompt tuning is a method integrating extra information into pre-trained language models by converting downstream tasks into cloze questions. The prompt is the primary component of the prompt tuning model. In the proposed method, prompts are aspect terms encoding domain-specific semantics. I leverage the prompts from various domains to span the shared semantic space, and reflect the similarities and differences between different domains.

The prediction of aspect term is formalised with designed prompts in Equation 6.4.

$$\hat{y}^p = softmax(W^p * [e^w; e^{pos}; e^p] + b^p), \tag{6.4}$$

where W^p is the training weights and b^p is the bias vector. The training objective of soft prompt tuning is calculated using cross-entropy loss in Equation 6.5.

$$\mathscr{L}_{prompt} = \sum^{\mathbb{D}_s} \sum_i^n f(\hat{y}_i^p, y_i^p), \tag{6.5}$$

6.4.3 Syntax Learning

For aspect terms from different domains, their linguistic features maintain often-occurring patterns (M. Hu & Liu, 2004a; Qiu et al., 2011; Z. Chen & Qian, 2021). Given an example in Figure 6.2, The aspect *Keyboard* from the domain **Laptop** shares the same POS tag *NN* with the aspect *Food* in the domain **Restaurant**, indicating that these aspect terms are similar in syntax. To learn the syntax knowledge, the encoded masked feature e^{pos} is fed into a Softmax layer. The predicted POS tag can be calculated in Equation 6.6.

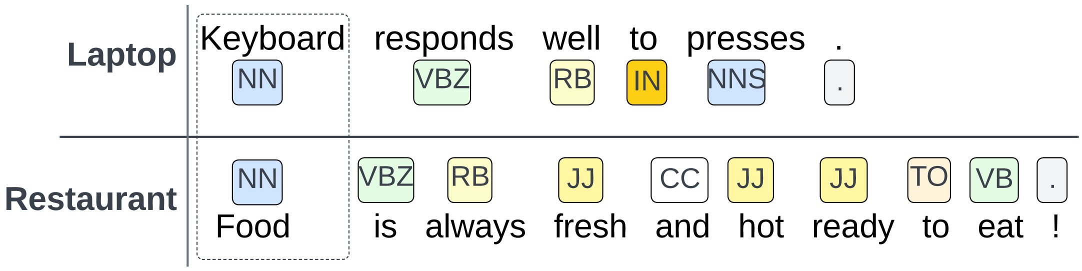


Figure 6.2: POS tags of reviews from laptop and restaurant domain. The aspect terms *Keyboard* and *Food* share the same POS tag *NN*.

$$\hat{y}^{pos} = softmax(W^{pos} * [e^{w}; e^{pos}; e^{p}] + b^{pos}), \tag{6.6}$$

where $\hat{y}^{pos} \in \mathbb{R}^{N^{pos}}$, N^{pos} is the number of total POS tags. W^{pos} is the learnable weight, and b^{pos} is the bias tensor. To optimise the learning process, the cross-entropy loss is calculated in Equation 6.7.

$$\mathscr{L}_{syntax} = \sum^{\mathbb{D}_s} \sum_{i}^{n} I(i) * f(\hat{y}_i^{pos}, y_i^{pos}), \tag{6.7}$$

$$I(i) = \begin{cases} 1 & if\, token\, is\, masked \\ 0 & else \end{cases} \tag{6.8}$$

where $I(i)$ is the indicator to filter the masked tokens. y_i^{pos} is the real POS tag of ith token in the input sentence.

6.4.4 Training Objective

Given the source domain datasets and the target dataset, the aspect term extraction and syntax discriminator are jointly trained for optimising the soft prompt embeddings, syntax embedding, and aspect term predictor. The final training objective is obtained by the weighted sum of the cross-entropy losses from syntax learning and multiple domain knowledge enhanced prompt-tuning in Equation 6.9.

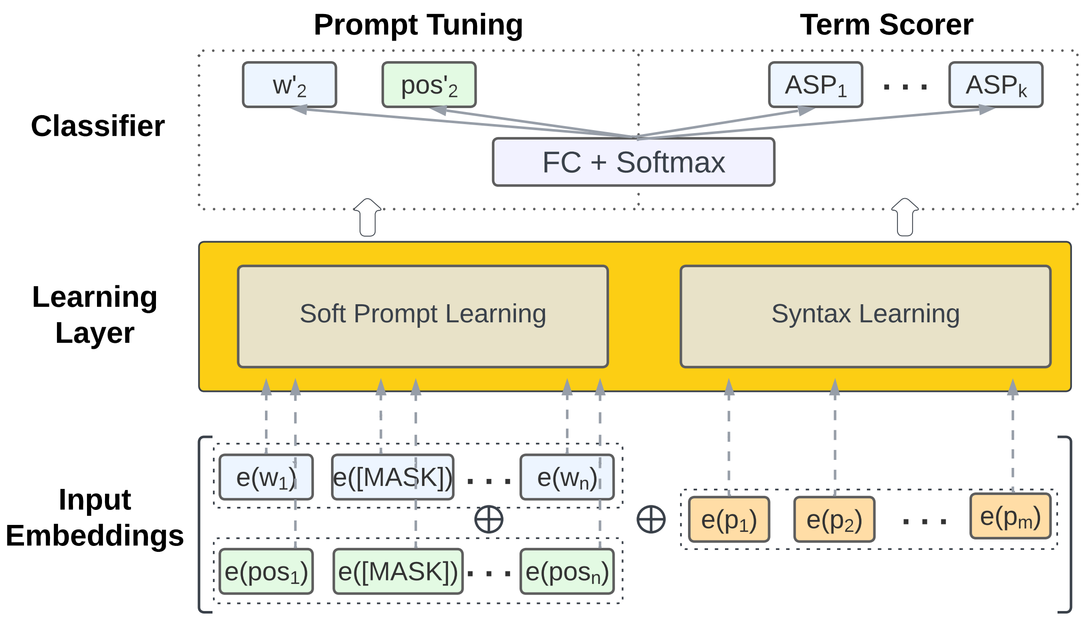


Figure 6.3: Overview architecture of the soft prompt-based joint learning model.

$$\mathscr{L}(\theta) = \alpha * \mathscr{L}_{prompt} + \beta * \mathscr{L}_{syntax} \tag{6.9}$$

where α and β are the trade-off parameter.

6.5 Experiments

In this section, extensive experiments are conducted on two groups of datasets to evaluate the proposed model

6.5.1 Dataset

The experiments of the proposed method are conducted on two groups of benchmark datasets with different domains. For the first group of dataset $\mathbb{G}1$, four domains are included: Device ($\mathbb{D}$) is the set of all the digital device reviews (Toprak, Jakob & Gurevych, 2010). Laptop ($\mathbb{L}$) and Restaurant ($\mathbb{R}$) are from SemEval ABSA challenges

Table 6.1: Statistics of the first group datasets.

Domain	Sentences	Train	Test
$\mathbb{R}$	6035	3877	2158
$\mathbb{L}$	3845	3045	800
$\mathbb{D}$	3836	2557	1279
$\mathbb{S}$	2239	1492	747

Table 6.2: Statistics of the second group datasets.

Domain	Sentences	Train	Test
$\mathbb{DI}$	375	262	113
$\mathbb{AS}$	380	266	114
$\mathbb{E}$	550	385	165

(Pontiki et al., 2014, 2015, 2016), which contain customer reviews of laptop and restaurant. Service ($\mathbb{S}$) refers to the customer reviews of web services (M. Hu & Liu, 2004a). The basic statistics of the first group of dataset are presented in Table 6.1. The second group of datasets $\mathbb{G}2$ contains three domains: Diapers($\mathbb{DI}$), Antivirus Software ($\mathbb{AS}$), and Electronics($\mathbb{E}$), shown in Table 6.2. $\mathbb{DI}$ and $\mathbb{AS}$ are prepared for opinion mining[1] by (X. Ding, Liu & Yu, 2008). $\mathbb{E}$ is annotated by the annotator, which is originally collected for cross-domain sentiment classification[2] by (Zola, Cortez, Ragno & Brentari, 2019).

Implementation and Hyper-parameters

In the proposed method, the Pytorch framework [3] is utilised to implement the proposed model. T5-base [4] is used as the base LMs. I use Stanford NLP Toolkit [5](i.e., Stanza (Qi et al., 2020)) to obtain the syntax structures of all datasets. All experiments are conducted on a single NVIDIA RTX A6000 GPU accelerator.

The default settings are used for T5-base, e.g., 24 layers of self-attention with 1024

[1]https://www.cs.uic.edu/ liub/FBS/Reviews-9-products.rar
[2]https://github.com/paolazola/Cross-source-cross-domain-sentiment-analysis
[3]https://pytorch.org/
[4]https://huggingface.co/t5-base
[5]https://stanfordnlp.github.io/stanza/

dimensional hidden vectors. The Adam optimiser (Diederik & Jimmy, 2015) is applied with an initial learning rate of 2e-3. The epoch is set to 20, and the batch size is 16.

6.5.2 Baselines

To verify the effectiveness of the proposed method, several competitive baselines are utilised to compare with the proposed model.

- ***CrossCRF*** (Jakob & Gurevych, 2010) is a traditional sequence labelling method, in which linguistic features (i.e., word type, POS tag, and dependency relation) are applied to detect aspect terms using CRF.
- ***DP*** (Qiu et al., 2011) addresses two problems, i.e., opinion lexicon expansion and opinion target extraction using a semi-supervised method based on bootstrapping. The dependency relations linking opinion terms and targets are extracted using a dependency parser, and then the identified relations are used to expand the initial opinion lexicons and detect aspect terms.
- ***mSDA*** (M. Chen, Xu, Weinberger & Sha, 2012) is a marginalised stacked denoising auto-encoder, which uses linear denoisers to build blocks for learning feature representations. This method can address the issues of high computational cost and lack of scalability to high-dimensional features.
- ***FEMA*** (Y. Yang & Eisenstein, 2015) performs dense feature representation learning, which is more robust to domain shift, using neural language models to obtain low-dimensional embeddings directly.
- ***RNCRF*** (W. Wang et al., 2016) integrates recursive neural networks and CRFs into a joint model to detect aspect and opinion terms. The unified framework can propagate bidirectional information between aspect and opinion terms, and learn high-level discriminative features.

- ***Hier-Joint*** (Y. Ding et al., 2017) combines rule-based, unsupervised aspect term extraction with neural network-based supervised methods to learn a hidden representation for different domains.

- ***RNSCN*** (W. Wang & Pan, 2018) is a novel recursive neural network, which can reduce the issue of domain shift in word level by dependency relations. The syntactic relations can be used as invariant pivot information across different domains between source and target datasets.

- ***AD-SAL*** (Z. Li et al., 2019) firstly explores an unsupervised domain adaption setting for joint extraction of aspect and opinion terms. Moreover, a selective adversarial learning method is proposed to learn an alignment weight for each word to achieve fine-grained domain adaption.

- ***BERT*** directly fine-tunes base BERT (Devlin et al., 2019a) to predict collapsed labels for cross-domain ATE task.

- ***TRNN-GRU*** (W. Wang & Pan, 2019a) introduces a conditional domain adversarial network to improve the knowledge transferability across different domains. Furthermore, the recursive neural network with a sequence labelling classifier is integrated to model contextual influence to predict the aspect terms in target datasets.

- ***CrossBERT*** (H. Xu et al., 2019) post-trains base BERT (Devlin et al., 2019a) on mixed datasets from Yelp and Amazon reviews, and then fine-tune the trained model to detect aspect terms across domains.

- ***CrossBERT-UDA*** (Gong et al., 2020) is an end-to-end framework that performs feature and instance based adaption for cross-domain ABSA tasks. This method can learn domain-invariant features via linguistic information, and perform word-level instance weighting based on BERT.

- ***SA-EXAL*** (Pereg et al., 2020) incorporates external linguistic information into a self-attention mechanism with BERT, which can bridge the gap across domains by leveraging the intrinsic knowledge from BERT with external syntactic information.

- ***CDRG-Indep*** (J. Yu, Gong & Xia, 2021) aims to generate target-domain data with fine-grained annotation based on labelled data in source domain, and then directly train a sequence labelling model on the generated dataset by adopting BERT model.

- ***CDRG-Merge*** (J. Yu et al., 2021) is similar to CDRG-Indep except for the training strategy, which merges the labelled source data with generated data as training examples.

- ***AHF*** (Y. Zhou et al., 2021) integrates pseudo-label based semi-supervised learning and adversarial training in a unified network for cross-domain ABSA tasks. The target data is utilised for training domain discriminator and refining the task classifier.

- ***SynBridge*** (Z. Chen & Qian, 2021) is an active domain adaptation model that transfers aspect words by actively supplementing transferable knowledge. The syntactic bridges are constructed by recognising syntactic roles as pivots to identify transferable syntactic roles for the words across domains.

- ***SemBridge*** (Z. Chen & Qian, 2021) is a similar model to SynBridge, but SemBridge retrieves transferable prototypes to link aspect words across domains.

- ***SDAM*** (Dong et al., 2022) is a syntax-guided domain adaptation method that exploits syntactic structure similarities to build pseudo training data.

- ***FMIM-BERT*** (X. Chen & Wan, 2022) is a simple but effective method based on mutual information maximization for cross-domain ABSA tasks.

6.5.3 Experimental Results and Model Analysis

I conduct experiments of cross-domain ATE on two groups of datasets $\mathbb{G}1$ and $\mathbb{G}2$, and the overall comparison results are shown in Tables 6.3 and 6.4. I can observe that the proposed method outperforms all baselines on most domains in dataset $\mathbb{G}1$ and all domains in dataset $\mathbb{G}2$. Compared with the previous approaches, the proposed method is significantly superior to machine learning and based models. However, the performance of the proposed method is lower than that of SemBridge for domain adaption $\mathbb{L} \rightarrow \mathbb{R}$ and $\mathbb{D} \rightarrow \mathbb{R}$ (-0.007 and -0.006, respectively), indicating that SemBridge captures more syntactic and semantic knowledge of source domains and transfers this meaningful knowledge to target domain. Without semantic features, SynBridge achieves a degraded performance compared with the proposed method. In contrast, the proposed method can outperform most of the fine-tuning based models, which shows that soft prompt tuning based method can learn domain-dependent features, but also domain-invariant knowledge. The outstanding performance demonstrates the prompt-tuning based method is able to solve the problem of cross-domain ATE.

Table 6.4 presents the results of cross-domain ATE on a small-scale dataset. Compared with fine-tuning based models, AD-SAL and CrossBERT-UDA, the proposed method achieves the best performance on all three domains (over 5% absolute improvement). The improvement demonstrates that the prompt-tuning based mode can be applied to both big and small-scale datasets with competitive performance. Compared with prompt tuning models, it's more difficult to train the domain-specific model on source domains for fine-tuning approaches. While prompt tuning models can activate some prior knowledge in language models by the feature distribution of prompts.

Table 6.3: Experimental results for cross-domain ATE on $\mathbb{G}1$

Model	$\mathbb{R} \rightarrow \mathbb{L}$	$\mathbb{S} \rightarrow \mathbb{L}$	$\mathbb{D} \rightarrow \mathbb{L}$	$\mathbb{L} \rightarrow \mathbb{R}$	$\mathbb{S} \rightarrow \mathbb{R}$	$\mathbb{D} \rightarrow \mathbb{R}$	$\mathbb{R} \rightarrow \mathbb{D}$	$\mathbb{L} \rightarrow \mathbb{D}$	$\mathbb{S} \rightarrow \mathbb{D}$	$\mathbb{R} \rightarrow \mathbb{S}$	$\mathbb{L} \rightarrow \mathbb{S}$	$\mathbb{D} \rightarrow \mathbb{S}$
CrossCRF	0.197	0.116	0.242	0.282	0.170	0.659	0.211	0.299	0.097	0.088	0.086	0.045
DP	0.198	0.198	-	0.376	0.376	0.376	0.218	-	0.218	0.197	0.197	0.197
mDA	0.209	0.146	0.257	0.243	0.325	0.213	0.172	0.294	0.169	0.131	0.131	0.131
FEMA	0.266	0.150	0.268	0.350	0.376	0.207	0.229	0.296	0.187	0.108	0.148	0.088
RNCRF	0.243	-	0.406	0.409	-	0.346	0.243	0.315	-	-	-	-
Hier-Joint	0.317	0.300	0.362	0.467	0.520	0.504	0.320	0.316	0.334	0.198	0.234	0.235
RNSCN	0.266	0.189	-	0.356	0.332	0.346	0.333	-	0.220	0.200	0.166	0.200
AD-SAL	0.341	0.270	-	0.430	0.410	0.410	0.354	-	0.336	0.280	0.272	0.266
BERT	0.314	0.305	-	0.404	0.447	0.403	0.276	-	0.339	0.195	0.258	0.303
TRNN-GRU	0.402	-	0.517	0.538	-	0.512	0.373	0.412	-	-	-	-
CrossBERT	0.397	0.350	-	0.454	0.513	0.426	0.332	-	0.332	0.244	0.233	0.282
CrossBERT-UDA	0.439	0.348	-	0.495	0.471	0.427	0.349	-	0.321	0.331	0.279	0.280
SA-EXAL	0.476	-	0.477	0.547	-	0.545	0.405	0.422	-	-	-	-
CDRG-Indep	0.402	0.332	-	0.551	0.538	0.501	0.308	-	0.349	0.417	0.441	0.371
CDRG-Merge	0.466	0.395	-	0.600	0.563	0.527	0.326	-	0.369	0.424	0.471	0.418
AHF	0.557	0.448	-	0.646	0.591	0.597	0.502	-	0.478	0.438	0.427	0.444
SynBridge	0.551	0.453	-	0.653	0.584	0.628	0.533	-	0.539	0.327	0.337	0.381
SemBridge	0.579	0.451	-	0.662	0.593	0.636	0.553	-	0.546	0.350	0.350	0.377
SDAM	0.546	0.467	-	0.631	0.586	0.609	0.516	-	0.580	0.456	0.453	0.552
FMIM-BERT	0.494	0.424	-	0.634	0.592	0.573	0.397	-	0.376	0.514	0.549	0.528
Ours	0.593	0.480	0.527	0.655	0.612	0.630	0.563	0.434	0.586	0.528	0.555	0.570

Table 6.4: Experimental results for cross-domain ATE on $\mathbb{G}2$.

Model	$\mathbb{DI} \rightarrow \mathbb{AS}$	$\mathbb{E} \rightarrow \mathbb{AS}$	$\mathbb{DI} \rightarrow \mathbb{E}$	$\mathbb{AS} \rightarrow \mathbb{E}$	$\mathbb{E} \rightarrow \mathbb{DI}$	$\mathbb{AS} \rightarrow \mathbb{DI}$
AD-SAL	0.208	0.181	0.172	0.191	0.156	0.176
CrossBERT-UDA	0.244	0.201	0.195	0.213	0.160	0.196
Ours	0.322	0.312	0.279	0.313	0.213	0.251

6.5.4 Ablation Study

To analyse the effect of each component including syntax learning and prompts, the ablation experiments are conducted on dataset $\mathbb{G}2$ and the experimental results are shown in Table 6.5.

Effect of Syntax Learning

In this subsection, the effect of syntax learning is verified via the ablation study. Table 6.5 presents the experimental results on dataset $\mathbb{G}2$. I find that, without syntax learning

Table 6.5: Ablation study over cross-domain ATE on $\mathbb{G}2$. w/o indicates without.

Model	$\mathbb{DI} \rightarrow \mathbb{AS}$	$\mathbb{E} \rightarrow \mathbb{AS}$	$\mathbb{DI} \rightarrow \mathbb{E}$	$\mathbb{AS} \rightarrow \mathbb{E}$	$\mathbb{E} \rightarrow \mathbb{DI}$	$\mathbb{AS} \rightarrow \mathbb{DI}$
-w/o Syntax	0.267	0.279	0.241	0.273	0.198	0.218
-w/o Prompts	0.244	0.258	0.261	0.253	0.171	0.226
only T5	0.196	0.218	0.133	0.236	0.119	0.164
Ours	0.322	0.312	0.279	0.313	0.213	0.251

component, the results of the proposed method see a decrease in all target domains (i.e., -5.5%, -3.3%, -3.8%, -4%, -1.5%, -3.3%, respectively). This shows that linguistic features are necessary to capture domain-invariant information between domains. The domain-independent features can bridge the gap over domains and facilitate the prediction in the target domain for cross-domain ATE task. For example, the model trained on domain $\mathbb{DI}$ with one input sentence "*The Diaper Champ is the best we found!*", and the aspect term is *Diaper Champ* with POS tag *NN*. After learning this syntactic pattern in the source domain, it can be easier for the model to predict the aspect term *Program* in the sentence "*The program brings more problems than a virus...*" on target domain $\mathbb{AS}$.

Effect of Prompts

I present an evaluation of the effect of prompts by removing them from the proposed method. In Table 6.5, the F1 scores of w/o prompts are presented across all target domains. After removing prompts, I observe a significant performance drop on all domains (-7.8%, -5.4%, -1.8%, -6.0%, -4.2%, -2.5%, respectively), suggesting the designed prompts can be leveraged to span the semantic space of source domains. The shared semantic knowledge can further promote the performance of cross-domain ATE task.

As stated in the previous research study (T. Gao, Fisch & Chen, 2021), the selection of prompts may have a huge impact on the model performance. Therefore, I conduct experiments using different numbers of prompt tokens on domain $\mathbb{DI} \rightarrow \mathbb{AS}$ to further investigate the influence of soft prompts. The results are shown in Figure 6.4, which demonstrates that the length of prompt token effects the performance of prompt tuning on domain $\mathbb{DI} \rightarrow \mathbb{AS}$. In the proposed method, I set the prompt token length as 3 to achieve the best results for cross-domain ATE.

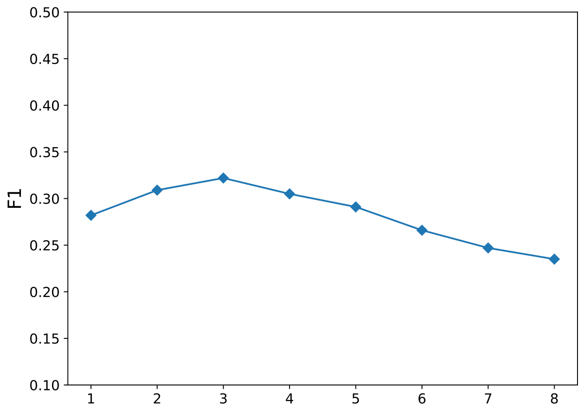


Figure 6.4: Experimental results of different lengths of soft prompt tokens on $\mathbb{DI} \rightarrow \mathbb{AS}$.

6.5.5 Case Study

To further demonstrate the effectiveness of the proposed model, I perform a case study on dataset $\mathbb{G}1$. Table 6.6 present the results of cross-domain ATE by AD-SAL, CrossBERT-UDA, and the proposed method. In example 1 "*Straight-forward, no surprises, very decent Japanese food.*", both CrossBERT-UDA and AD-SAL can not identify the aspect term *Japanese food*. CrossBERT-UDA extracts one of three aspect terms, while AD-SAL only predicts one wrong aspect term in example 2 "*While there's a decent menu, it shouldn't take ten minutes to get your drink and 45 for a dessert pizza*", in which there are three aspect terms *menu*, *drink*, and *dessert pizza*. For example 3 and 6, there are multiple aspect terms and most of them include more than one word. AD-SAL and CrossBERT-UDA can only detect part of words for these aspect terms. In examples 4 and 5, baseline models are able to correctly predict part of aspect terms while the proposed method identifies all of them. The case study demonstrates that the proposed method can accurately detect not only multiple aspect terms but also aspect terms with multiple words.

Table 6.6: The prediction results of aspect term on domain $\mathbb{R}$ by AD-SAL, CrossBERT-UDA, and the proposed method. Incorrect predictions are indicated by marker ✗.

Input ($\mathbb{S}$->$\mathbb{R}$)	**AD-SAL**	**CrossBERT-UDA**	**Ours**
Straight-forward, no surprises, very decent [Japanese food].	{[ ]}	{[ ]}	{[Japanese food]}
While there's a decent [menu], it shouldn't take ten minutes to get your [drink] and 45 for a [dessert pizza].	{[pizza]✗}	{[menu]}	{[menu], [drink], [dessert pizza]}
I've had the [jellyfish], [horse mackerel], the [blue fin tuna] and the [sake ikura roll] among others, and they were all good.	{[jellyfish], [horse]✗, [tuna]✗}	{[jellyfish], [erel]✗, [sake]✗, [ura roll]✗}	{[jellyfish], [horse mackerel], [blue fin tuna], [sake ikura roll]}
The [food] is top notch , the [service] is attentive , and the [atmosphere] is great.	{[food], [service]}	{[food], [service]}	{[food], [service], [atmosphere]}
Try the [ribs] , sizzling [beef] and couple it with [coconut rice].	{[ribs], [beef]}	{[ribs], [beef], [rice]✗}	{[ribs], [beef], [coconut rice]}
They have a very good [chicken with avocado] and good [tuna] as well .	{[chicken]✗, [tuna]}	{[avocado]✗, [tuna]}	{[chicken with avocado], [tuna]}

6.5.6 Discussions

In this chapter, I propose a novel soft prompt-based joint learning method for cross-domain aspect term extraction. The existing approaches are either machine learning or deep learning-based, or hard prompt-based methods, which suffer from low-quality domain-invariant features or unstable performance on small-scale target datasets. Different from previous methods, soft prompts are applied to learn in-domain knowledge of different domains to enhance the domain-invariant feature representations. Instead of directly integrating syntax information, a self-supervised learning of syntactic features is designed to learn the structural correspondence between domains for narrowing the domain gap. The experiments across two groups of datasets spanning a range of domains demonstrate the effectiveness of the proposed approach over the existing models for cross-domain aspect term extraction.

6.6 Summary

In this chapter, I studied the problem of prompt construction and the integration of linguistic features for cross-domain aspect term extraction, and then proposed an innovative approach to cross-domain aspect term extraction through the use of a soft prompt-based joint learning method. This method combines external linguistic features and multiple objectives to learn domain-invariant representations between source and target domains, effectively closing the gap between domains with differing distributions of aspect terms. The method also incorporates a set of transferable soft prompts, consisting of multiple learnable vectors, to improve the detection of aspect terms in the target domain. Overall, the proposed soft prompt-based joint learning method is a groundbreaking solution for cross-domain aspect term extraction. To verify the effectiveness of the proposed method, Extensive experiments are conducted on the benchmark datasets and the experimental results demonstrate that the proposed model achieved significantly improved new state-of-the-art F1 scores.

This chapter mainly answers Research Question 3 mentioned in Chapter 1. The model and results of this chapter have been published in (J. Shi et al., 2023).

Chapter 7

Conclusion

7.1 Introduction

The findings of aspect extraction, aspect and relation extraction, a unified framework, and prompt-tuning models are summarised in this chapter. The thesis proposes four novel aspect-based sentiment analysis models, each targeting a unique aspect of aspect-based sentiment analysis.

The research contributions are summarised in Section 7.2. The limitations of the models and possible directions for future work are covered in Section 7.3.

7.2 Research Contributions

This thesis contributes to the field from the following four aspects.

7.2.1 Aspect Term Extraction Model

- I proposed a novel feature selection-based framework to explore the most relevant features for aspect term extraction, where both BERT embeddings and relevant linguistic features are integrated.

- I designed a novel feature selection method by extending the Artificial Bee Colony with an adaptive threshold, which can address the high sparsity and dimensionality issue of training datasets.

- I analysed and evaluated the effectiveness of the proposed framework by conducting extensive experiments on real-world datasets and explicitly prove the selected implicit features can improve the performance of aspect term extraction.

7.2.2 Aspect and Relation Extraction Model

- I first applied NLP and deep learning algorithms to explore public concerns and deeply investigate the corresponding relations using online social media data in the context of COVID-19 in (J. Shi et al., 2021).

- I extended the proposed model in (J. Shi et al., 2021) and then developed an end-to-end model with Concern Graph (CG) and concern states to identify public concerns as aspect terms and corresponding relations simultaneously.

- I formally define public concern with a consideration of its type and degree, and construct the concern graph to represent the regional features, improving the concern identification effectiveness.

- I further integrate concern states with Graph Convolutional Network to provide important concern features for concern relations extraction.

7.2.3 Unified Framework for Aspect-based Sentiment Analysis

- I proposed a novel neural network model to integrate explicit syntactic information with semantic features for all 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.

- I designed a multi-layer semantic GCN to learn the representation via adjacency neighbourhood of context.
- I designed a multi-layer syntax GCN to encode syntactic structure information through structural connections.
- To incorporate explicit syntax information, I propose a Multiple Syntactic Structure (MSS) fusion encoder is proposed, leveraging syntax information to enrich the semantic features of user review.

7.2.4 Soft Prompt-based Model

- I presented a new joint learning approach for the cross-domain aspect term extraction task. To the best of my knowledge, I am pioneering the solution to the domain adaptation issue through the use of soft prompts.
- I designed The learnable soft prompt with rich external syntactic knowledge to better leverage the domain-specific and domain-invariant knowledge across domains.
- I design a linguistic knowledge learning module to learn the domain-invariant representation across domains.

7.3 Limitations and Future Directions

Throughout my Ph.D. research, I proposed and developed various innovative models aimed at improving cross-domain aspect-based sentiment analysis. In this chapter, I aim to assess the limitations of these models and explore potential avenues for future advancements in this field.

Firstly, the ABC algorithm requires more computing resources than other swarm intelligence-based methods (e.g., particle swarm optimization, genetic algorithm, etc.) due to its long search time and slow convergence speed. To ensure its practicality and effectiveness, I conducted experiments on four groups of public datasets to verify the applicability of the proposed framework. In the future, I plan to enhance the proposed approach in two ways: (1) use parallel computing technology to speed up the computation of the ABC algorithm. (2) apply the proposed framework to other domain datasets, such as Amazon and Yelp reviews, to uncover additional features.

Secondly, a limited types of public concerns are defined during my investigation of the concern and relation extraction problem. The time factor was not considered in the model at the outbreak of the COVID-19 pandemic due to dataset limitations. In the future, I plan to enhance the proposed model in two ways: (1) predict more types of concerns and concern relations to gain insight into what people are focused on and how these concerns are related, (2) incorporate the time factor to track the evolution of a specific concern over time.

Thirdly, in the MSS-GCN Fusion Encoder, only the output representations of the semantic GCN and syntax GCN are combined and fed into a feed-forward neural network. The interaction between the two GCNs has not been studied. Additionally, the proposed integrated framework has only been tested on single-domain datasets. Therefore, I plan to enhance the proposed model in two ways: (1) utilise message passing between the semantic and syntactic GCNs, (2) transfer the learned knowledge from one domain to other domain datasets.

Lastly, in my research on cross-domain sentiment analysis, I focus only on aspect term extraction using a soft prompt-based model. I intend to extend the proposed model to perform more aspect-based sentiment analysis tasks such as opinion term extraction, aspect-opinion term pair extraction, sentiment triplet extraction, etc. Additionally, a potential research direction is designing a unified framework to complete all ABSA subtasks in different domains by incorporating the relationships between aspect, opinion, and sentiment polarity.

References

Agarwal, O., Kale, M., Ge, H., Shakeri, S. & Al-Rfou, R. (2020). Machine translation aided bilingual data-to-text generation and semantic parsing. In *Proceedings of the 3rd international workshop on natural language generation from the semantic web (webnlg+)* (pp. 125–130).

Akbik, A., Blythe, D. & Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. In *Proceedings of the 27th international conference on computational linguistics* (pp. 1638–1649).

Akhtar, M. S., Gupta, D., Ekbal, A. & Bhattacharyya, P. (2017). Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems*, *100*(125), 116–135.

Alsahaf, A., Petkov, N., Shenoy, V. & Azzopardi, G. (2022). A framework for feature selection through boosting. *Expert Systems with Applications*, *187*, 115895.

Anand, D. & Mampilli, B. S. (2021). A novel evolutionary approach for learning syntactic features for cross domain opinion target extraction. *Applied Soft Computing*, *102*, 107086.

Asai, A., Salehi, M., Peters, M. E. & Hajishirzi, H. (2022). Attentional mixtures of soft prompt tuning for parameter-efficient multi-task knowledge sharing. *arXiv preprint arXiv:2205.11961*.

Azhar, A. N., Khodra, M. L. & Sutiono, A. P. (2019). Multi-label aspect categorization with convolutional neural networks and extreme gradient boosting. In *Proceedings of the 2019 international conference on electrical engineering and informatics* (pp. 35–40).

Ba, J. L., Kiros, J. R. & Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.

Bagheri, A., Saraee, M. & De Jong, F. (2014). Adm-lda: An aspect detection model based on topic modelling using the structure of review sentences. *Journal of Information Science*, *40*(5), 621–636.

Battaglia, P., Pascanu, R., Lai, M., Rezende, D. J. & kavukcuoglu, K. (2016). Interaction networks for learning about objects, relations and physics. In *Proceedings of the 30th international conference on neural information processing systems* (pp. 4509–4517).

Bekoulis, G., Deleu, J., Demeester, T. & Develder, C. (2018). Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, *114*, 34–45.

Ben-David, E., Oved, N. & Reichart, R. (2021). Pada: A prompt-based autoregressive approach for adaptation to unseen domains. *arXiv preprint arXiv:2102.12206*.

Ben-David, E., Oved, N. & Reichart, R. (2022). Pada: Example-based prompt learning for on-the-fly adaptation to unseen domains. *Transactions of the Association for Computational Linguistics*, *10*, 414–433.

Ben-David, E., Rabinovitz, C. & Reichart, R. (2020). Perl: Pivot-based domain adaptation for pre-trained deep contextualized embedding models. *Transactions of the Association for Computational Linguistics*, *8*, 504–521.

Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V. & Jurafsky, D. (2004). Automatic extraction of opinion propositions and their holders. In *2004 aaai spring symposium on exploring attitude and affect in text* (Vol. 2224).

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brody, S. & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. In *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics* (pp. 804–812).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... others (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877–1901.

Cai, D. & Lam, W. (2020). Graph transformer for graph-to-sequence learning. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 34, pp. 7464–7471).

Cai, H., Tu, Y., Zhou, X., Yu, J. & Xia, R. (2020). Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th international conference on computational linguistics* (pp. 833–843).

Chandrasekaran, R., Mehta, V., Valkunde, T. & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. *Journal of Medical Internet Research*, *22*(10), e22624.

Chatterji, S., Varshney, N. & Rahul, R. K. (2017). Aspectframenet: a framenet extension for analysis of sentiments around product aspects. *The Journal of Supercomputing*, *73*, 961–972.

Chen, H., Zhai, Z., Feng, F., Li, R. & Wang, X. (2022). Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th annual meeting of the association for computational linguistics* (pp. 2974–2985).

Chen, H.-Y. & Chen, H.-H. (2016). Implicit polarity and implicit aspect recognition in opinion mining. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 20–25).

Chen, L., Lyu, H., Yang, T., Wang, Y. & Luo, J. (2020). *In the eyes of the beholder: Sentiment and topic analyses on social media use of neutral and controversial terms for COVID-19.* (arXiv preprint arXiv:2004.10225)

Chen, M., Xu, Z., Weinberger, K. & Sha, F. (2012). Marginalized denoising autoencoders for domain adaptation. *arXiv preprint arXiv:1206.4683*.

Chen, S., Liu, J., Wang, Y., Zhang, W. & Chi, Z. (2020). Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 6515–6524).

Chen, S., Wang, Y., Liu, J. & Wang, Y. (2021). Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the 2021 aaai conference on artificial intelligence* (Vol. 35, pp. 12666–12674).

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H. et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4.2*, *1*(4).

Chen, T., Xu, R., He, Y. & Wang, X. (2017). Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. *Expert Systems with Applications*, *72*, 221–230.

Chen, X. & Wan, X. (2022). A simple information-based approach to unsupervised domain-adaptive aspect-based sentiment analysis. *arXiv preprint arXiv:2201.12549*.

Chen, X., Zhang, N., Li, L., Xie, X., Deng, S., Tan, C., ... Chen, H. (2021). Lightner: A lightweight generative framework with prompt-guided attention for low-resource ner. *arXiv preprint arXiv:2109.00720*.

Chen, X., Zhang, N., Xie, X., Deng, S., Yao, Y., Tan, C., ... Chen, H. (2022). Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In *Proceedings of the acm web conference 2022* (pp. 2778–2788).

Chen, Z., Huang, H., Liu, B., Shi, X. & Jin, H. (2021). Semantic and syntactic enhanced aspect sentiment triplet extraction. In *Findings of the association for computational linguistics: Acl-ijcnlp 2021* (pp. 1474–1483).

Chen, Z., Mukherjee, A. & Liu, B. (2014). Aspect extraction with automated prior knowledge learning. In *Proceedings of the 52nd annual meeting of the association for computational linguistics* (pp. 347–358).

Chen, Z. & Qian, T. (2020). Relation-aware collaborative learning for unified aspect-based sentiment analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3685–3694).

Chen, Z. & Qian, T. (2021). Bridge-based active domain adaptation for aspect term extraction. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 317–327).

Chernyshevich, M. & Belarus, I. (2014). Cross-domain extraction of product features using conditional random fields. In *Proc. 8th int'l workshop on semantic evaluation (semeval 14)* (pp. 309–313).

Chew, C. & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PloS one*, *5*(11), e14118.

Cho, K., van Merriënboer, B., Bahdanau, D. & Bengio, Y. (2014). 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* (pp. 103–111).

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Culotta, A. & Sorensen, J. (2004). Dependency tree kernels for relation extraction. In *Proceedings of the 42nd annual meeting of the association for computational linguistics* (pp. 423–429).

Dai, H. & Song, Y. (2019). Neural aspect and opinion term extraction with mined rules as weak supervision. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 5268–5277).

Dai, J., Yan, H., Sun, T., Liu, P. & Qiu, X. (2021). 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* (pp. 1816–1829).

Damiano, A. & Catellier JR, A. (2020). A content analysis of coronavirus tweets in the United States just prior to the pandemic declaration. *Cyberpsychology, Behavior and Social Networking*, *23*(12), 889–893.

da Silva, J. A. T., Tsigaris, P. & Erfanmanesh, M. (2021). Publishing volumes in major databases related to COVID-19. *Scientometrics*, *126*(1), 831–842.

Defferrard, M., Bresson, X. & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th international conference on neural information processing systems* (pp. 3844–3852).

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019a). 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* (pp. 4171–4186).

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019b). 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* (pp. 4171–4186).

Diederik, P. K. & Jimmy, B. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd international conference on learning representations.*

Ding, X., Liu, B. & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 231–240).

Ding, Y., Yu, J. & Jiang, J. (2017). Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 31).

Dong, A., Gao, C., Jia, Y., Liao, Q., Wang, X., Wang, L. & Xiao, J. (2022). Syntax-guided domain adaptation for aspect-based sentiment analysis. *arXiv preprint arXiv:2211.05457*.

Dosoula, N., Griep, R., den Ridder, R., Slangen, R., Schouten, K. & Frasincar, F. (2016). Detection of multiple implicit features per sentence in consumer review data. In *Databases and information systems: 12th international baltic conference, db&is 2016, riga, latvia, july 4-6, 2016, proceedings 12* (pp. 289–303).

Du, C., Sun, H., Wang, J., Qi, Q. & Liao, J. (2020). Adversarial and domain-aware bert for cross-domain sentiment analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 4019–4028).

Eberts, M. & Ulges, A. (2020). Span-based joint entity and relation extraction with transformer pre-training. In *Proceedings of the 24th european conference on artificial intelligence* (pp. 2006–2013).

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211.

Fan, Z., Wu, Z., Dai, X., Huang, S. & Chen, J. (2019). Target-oriented opinion words extraction with target-fused neural sequence labeling. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 2509–2518).

Feng, J., Cai, S. & Ma, X. (2019). Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm. *Cluster Computing*, *22*, 5839–5857.

Feng, Y., Rao, Y., Tang, Y., Wang, N. & Liu, H. (2021). 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* (pp. 1805–1815).

Fu, K.-W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P. & Fung, I. C.-H. (2016). How people react to Zika virus outbreaks on Twitter? a computational content analysis. *American Journal of Infection Control*, *44*(12), 1700–1702.

Fu, T.-J., Li, P.-H. & Ma, W.-Y. (2019). Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 1409–1418).

Gao, L., Wang, Y., Liu, T., Wang, J., Zhang, L. & Liao, J. (2021). Question-driven span labeling model for aspect–opinion pair extraction. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 12875–12883).

Gao, T., Fang, J., Liu, H., Liu, Z., Liu, C., Liu, P., ... Yan, W. (2022). Lego-absa: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th international conference on computational linguistics* (pp. 7002–7012).

Gao, T., Fisch, A. & Chen, D. (2021). Making pre-trained language models better few-shot learners. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 3816–3830).

Geetha, M. & Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model. *International Journal of Intelligent Networks*, *2*, 64–69.

Glowacki, E. M., Lazard, A. J., Wilcox, G. B., Mackert, M. & Bernhardt, J. M. (2016). Identifying the public's concerns and the centers for disease control and prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. *American Journal of Infection Control*, *44*(12), 1709–1711.

Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, *57*, 345–420.

Gong, C., Yu, J. & Xia, R. (2020). Unified feature and instance based domain adaptation for aspect-based sentiment analysis. In *Proceedings of the 2020 conference on empirical methods in natural language processing (emnlp)* (pp. 7035–7045).

Hai, Z., Chang, K. & Cong, G. (2012). One seed to find them all: mining opinion features via association. In *Proceedings of the 21st acm international conference on information and knowledge management* (pp. 255–264).

Hamilton, W., Ying, Z. & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 1025–1035).

Hang, T., Feng, J., Wu, Y., Yan, L. & Wang, Y. (2021). Joint extraction of entities and overlapping relations using source-target entity labeling. *Expert Systems with Applications*, *177*, 114853.

Hatzivassiloglou, V. & McKeown, K. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics* (pp. 174–181).

He, R., Lee, W. S., Ng, H. T. & Dahlmeier, D. (2018). Exploiting document knowledge for aspect-level sentiment classification. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 579–585).

He, R., Lee, W. S., Ng, H. T. & Dahlmeier, D. (2019). An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 504–515).

Hewitt, J. & Manning, C. D. (2019). A structural probe for finding syntax in word representations. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 4129–4138).

Hoang, M., Bihorac, O. A. & Rouces, J. (2019). Aspect-based sentiment analysis using bert. In *Proceedings of the 22nd nordic conference on computational linguistics* (pp. 187–196).

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hong, Y., Liu, Y., Yang, S., Zhang, K. & Hu, J. (2020). Joint extraction of entities and relations using graph convolution over pruned dependency trees. *Neurocomputing*, *411*, 302–312.

Hou, Z., Du, F., Jiang, H., Zhou, X. & Lin, L. (2020). *Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: Social media surveillance in China.* (Preprint at *SSRN* `http://dx.doi.org/10.2139/ssrn.3551338`)

Howard, P., Ma, A., Lal, V., Simoes, A. P., Korat, D., Pereg, O., … Singer, G. (2022). Cross-domain aspect extraction using transformers augmented with knowledge graphs. In *Proceedings of the 31st acm international conference on information & knowledge management* (pp. 780–790).

Hu, M. & Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings* *of the tenth acm sigkdd international conference on knowledge discovery and data mining* (pp. 168–177).

Hu, M. & Liu, B. (2004b). Mining opinion features in customer reviews. In *Aaai* (Vol. 4, pp. 755–760).

Hu, M., Peng, Y., Huang, Z., Li, D. & Lv, Y. (2019). Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 537–546).

Hu, Y., Bai, Q. & Li, W. (2019). Context-aware influence diffusion in online social networks. In *Pacific rim knowledge acquisition workshop* (pp. 153–162).

Huang, Z., Xu, W. & Yu, K. (2015). *Bidirectional LSTM-CRF models for sequence tagging.* (arXiv preprint arXiv:1508.01991)

Jahanbin, K., Rahmanian, V. et al. (2020). Using twitter and web news mining to predict covid-19 outbreak. *Asian Pacific Journal of Tropical Medicine*, *13*(8), 378.

Jakob, N. & Gurevych, I. (2010). Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 1035–1045).

Jelodar, H., Wang, Y., Orji, R. & Huang, S. (2020). Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. *IEEE Journal of Biomedical and Health Informatics*, *24*(10), 2733–2742.

Jiang, J., Wang, A. & Aizawa, A. (2021). 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* (pp. 1986–1997).

Jiang, L., Yu, M., Zhou, M., Liu, X. & Zhao, T. (2011). Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies* (pp. 151–160).

Jin, W. & Ho, H. H. (2009). A novel lexicalized hmm-based learning framework for web opinion mining. In *Proceedings of the 26th annual international conference on machine learning* (pp. 465–472).

Jin, W., Ho, H. H. & Srihari, R. K. (2009). Opinionminer: a novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1195–1204).

Jo, Y. & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth acm international conference on web search and data mining* (pp. 815–824).

Kahla, M., Yang, Z. G. & Novák, A. (2021). Cross-lingual fine-tuning for abstractive arabic text summarization. In *Proceedings of the international conference on recent advances in natural language processing (ranlp 2021)* (pp. 655–663).

Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization* (Tech. Rep.). Citeseer.

Kassab, L., Kryshchenko, A., Lyu, H., Molitor, D., Needell, D. & Rebrova, E. (2020).

On nonnegative matrix and tensor decompositions for COVID-19 Twitter dynamics. (arXiv preprint arXiv:2010.01600)

Katiyar, A. & Cardie, C. (2017). Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In *Proceedings of the 55th annual meeting of the association for computational linguistics* (pp. 917–928).

Kaveh-Yazdy, F. & Zarifzadeh, S. (2020). Track Iran's national COVID-19 response committee's major concerns using two-stage unsupervised topic modeling. *International Journal of Medical Informatics*, *145*, 104309.

Killeen, B. D., Wu, J. Y., Shah, K., Zapaishchykova, A., Nikutta, P., Tamhane, A., ... Thies, M. (2020). A county-level dataset for informing the United States' response to COVID-19. *Hospital*, *4000*(6000), 8000.

Killgore, W. D., Cloonen, S. A., Taylor, E. C. & Dailey, N. S. (2020). Loneliness: A signature mental health concern in the era of COVID-19. *Psychiatry Research*, *290*, 113117.

Kim, E. H.-J., Jeong, Y. K., Kim, Y., Kang, K. Y. & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, *42*(6), 763–781.

Kim, S.-M. & Hovy, E. (2004). Determining the sentiment of opinions. In *Proceedings of the 20th international conference on computational linguistics* (pp. 1367–1373).

Kipf, T. N. & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 2017 conference on learning representations.*

Klein, A., Pereg, O., Korat, D., Lal, V., Wasserblat, M. & Dagan, I. (2022). Opinion-based relational pivoting for cross-domain aspect term extraction. In *Proceedings of the 12th workshop on computational approaches to subjectivity, sentiment & social media analysis* (pp. 104–112).

Koncz, P. & Paralic, J. (2011). An approach to feature selection for sentiment analysis. In *Proceedings of the 15th ieee international conference on intelligent engineering systems* (pp. 357–362).

Ku, L.-W., Liang, Y.-T. & Chen, H.-H. (2006). Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of aaai* (pp. 100–107).

Kumar, H. K. & Harish, B. (2018). A new feature selection method for sentiment analysis in short text. *Journal of Intelligent Systems*, *29*(1), 1122–1134.

Kuo, R.-J., Huang, S. L., Zulvia, F. E. & Liao, T. W. (2018). Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction. *Knowledge and Information Systems*, *55*(1), 253–274.

Lafferty, J. D., McCallum, A. & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th international conference on machine learning* (pp. 282–289).

Lai, C.-H., Liu, D.-R. & Lien, K.-S. (2021). A hybrid of xgboost and aspect-based review mining with attention neural network for user preference prediction. *International Journal of Machine Learning and Cybernetics*, *12*(5), 1203–1217.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K. & Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 260–270).

Lamsal, R. (2021). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*, *51*(5), 2790–2804.

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B. & Suran, M. (2015). Detecting themes of public concern: a text mining analysis of the centers for disease control and prevention's Ebola live Twitter chat. *American Journal of Infection Control*, *43*(10), 1109–1111.

Le Scao, T. & Rush, A. M. (2021). How many data points is a prompt worth? In *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 2627–2636).

Lester, B., Al-Rfou, R. & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 3045–3059).

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... Zettlemoyer, L. (2020). 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* (pp. 7871–7880).

Li, C., Gao, F., Bu, J., Xu, L., Chen, X., Gu, Y., ... others (2021). Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis. *arXiv preprint arXiv:2109.08306*.

Li, F., Han, C., Huang, M., Zhu, X., Xia, Y., Zhang, S. & Yu, H. (2010). Structure-aware review mining and summarization. In *Proceedings of the 23rd international conference on computational linguistics* (pp. 653–661).

Li, H., Pun, C.-M., Xu, F., Pan, L., Zong, R., Gao, H. & Lu, H. (2021). A hybrid feature selection algorithm based on a discrete artificial bee colony for parkinson's diagnosis. *ACM Transactions on Internet Technology*, *21*(3), 1–22.

Li, H., Yang, L., Li, L., Xu, C., Xia, S.-T. & Yuan, C. (2022). Pts: A prompt-based teacher-student network for weakly supervised aspect detection. In *2022 international joint conference on neural networks (ijcnn)* (pp. 1–8).

Li, J., Yu, J. & Xia, R. (2022). Generative cross-domain data augmentation for aspect and opinion co-extraction. In *Proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 4219–4229).

Li, L., Zhang, Q., Wang, X., Zhang, J., Wang, T., Gao, T.-L., ... Wang, F.-Y. (2020). Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo. *IEEE Transactions on Computational Social Systems*, *7*(2), 556–562.

Li, R., Chen, H., Feng, F., Ma, Z., Wang, X. & Hovy, E. (2021). 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* (pp. 6319–6329).

Li, S., Wang, R. & Zhou, G. (2012). Opinion target extraction using a shallow semantic parsing framework. In *Twenty-sixth aaai conference on artificial intelligence.*

Li, W., Bai, Q., Zhang, M. & Nguyen, T. D. (2018). Automated influence maintenance in social networks: An agent-based approach. *IEEE Transactions on Knowledge and Data Engineering*, *31*(10), 1884–1897.

Li, W., Qi, F., Tang, M. & Yu, Z. (2020). Bidirectional lstm with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing*, *387*, 63–77.

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

Li, X., Bing, L., Lam, W. & Shi, B. (2018). Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 946–956).

Li, X., Bing, L., Li, P. & Lam, W. (2019). A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the 2019 aaai conference on artificial intelligence* (Vol. 33, pp. 6714–6721).

Li, X., Bing, L., Li, P., Lam, W. & Yang, Z. (2018). Aspect term extraction with history attention and selective transformation. In *Proceedings of the 27th international joint conference on artificial intelligence* (pp. 4194–4200).

Li, X. & Lam, W. (2017). Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2886–2892).

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

Li, Z., Li, X., Wei, Y., Bing, L., Zhang, Y. & Yang, Q. (2019). Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 4590–4600).

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

Liao, M., Li, J., Zhang, H., Wang, L., Wu, X. & Wong, K.-F. (2019). Coupling global and local context for unsupervised aspect extraction. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 4579–4589).

Lin, C., Miller, T., Dligach, D., Bethard, S. & Savova, G. (2019). A bert-based universal model for both within-and cross-sentence clinical temporal relation extraction. In *Proceedings of the 2nd clinical natural language processing workshop* (pp. 65–71).

Lin, Y., Tan, Y. C. & Frank, R. (2019). Open sesame: Getting inside bert's linguistic knowledge. In *Proceedings of the 2019 acl workshop blackboxnlp: Analyzing and interpreting neural networks for nlp* (pp. 241–253).

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

Liu, K., Xu, H. L., Liu, Y. & Zhao, J. (2013). Opinion target extraction using partially-supervised word alignment model. In *Ijcai* (Vol. 13, pp. 2134–2140).

Liu, K., Xu, L. & Zhao, J. (2014). Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. *IEEE Transactions on knowledge and data engineering*, *27*(3), 636–650.

Liu, P., Joty, S. & Meng, H. (2015). Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1433–1443).

Liu, Q., Liu, B., Zhang, Y., Kim, D. S. & Gao, Z. (2016). Improving opinion aspect extraction using semantic similarity and aspect associations. In *Proceedings of the 30th aaai conference on artificial intelligence.*

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

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

Luo, H., Li, T., Liu, B. & Zhang, J. (2019). Doer: Dual cross-shared rnn for aspect term-polarity co-extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 591–601).

Ma, D., Li, S. & Wang, H. (2018). Joint learning for targeted sentiment analysis. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4737–4742).

Ma, J., Cheng, J. C., Xu, Z., Chen, K., Lin, C. & Jiang, F. (2020). Identification of the most influential areas for air pollution control using xgboost and grid importance rank. *Journal of Cleaner Production*, *274*, 122835.

Ma, X. & Hovy, E. (2016). End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (pp. 1064–1074).

Mampilli, B. S. & Anand, D. (2022). Cross domain aspect extraction using various embedding techniques and language models. In *Proceedings of the 2nd international conference on recent trends in machine learning, iot, smart cities and applications* (pp. 237–248).

Manek, A. S., Shenoy, P. D., Mohan, M. C. & Venugopal, K. (2017). Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier. *World Wide Web*, *20*(2), 135–154.

Mao, Y., Shen, Y., Yu, C. & Cai, L. (2021). A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the 2021 aaai conference on artificial intelligence* (Vol. 35, pp. 13543–13551).

Marcacini, R. M., Rossi, R. G., Matsuno, I. P. & Rezende, S. O. (2018). Cross-domain aspect extraction for sentiment analysis: A transductive learning approach. *Decision Support Systems*, *114*, 70–80.

Mensah, S., Sun, K. & Aletras, N. (2021). An empirical study on leveraging position embeddings for target-oriented opinion words extraction. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 9174–9179).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th international conference on neural information processing systems* (p. 3111–3119).

Mikolov, T., Yih, W.-t. & Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 746–751).

Mitchell, M., Aguilar, J., Wilson, T. & Van Durme, B. (2013). Open domain targeted sentiment. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1643–1654).

Miwa, M. & Bansal, M. (2016). End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (pp. 1105–1116).

Mohit, B. (2014). Named entity recognition. In *Natural language processing of semitic languages* (pp. 221–245). Springer.

Mowlaei, M. E., Abadeh, M. S. & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, *148*, 113234.

Mukherjee, A. & Liu, B. (2012). Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th annual meeting of the association for computational linguistics* (pp. 339–348).

Nelson, B. W., Pettitt, A., Flannery, J. E. & Allen, N. B. (2020). Rapid assessment of psychological and epidemiological correlates of COVID-19 concern, financial strain, and health-related behavior change in a large online sample. *PloS one*, *15*(11), e0241990.

Ngai, H., Park, Y., Chen, J. & Parsapoor, M. (2021). Transformer-based models for question answering on covid19. *arXiv preprint arXiv:2101.11432*.

Nguyen, D. Q., Vu, T., Pham, S. B. et al. (2014). Sentiment classification on polarity reviews: An empirical study using rating-based features. In *Proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 128–135).

Ozyurt, B. & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: Ss-lda. *Expert Systems with Applications*, *168*, 114231.

Peng, H., Xu, L., Bing, L., Huang, F., Lu, W. & Si, L. (2020). Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the 2020 aaai conference on artificial intelligence* (Vol. 34, pp. 8600–8607).

Pennington, J., Socher, R. & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543).

Pereg, O., Korat, D. & Wasserblat, M. (2020). Syntactically aware cross-domain aspect and opinion terms extraction. In *Proceedings of the 28th international conference on computational linguistics* (pp. 1772–1777).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 conference of the north American chapter of the association for computational linguistics: Human language technologies* (pp. 2227–2237).

Platt, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, *10*(3), 61–74.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., … others (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (semeval 2016)* (pp. 19–30).

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S. & Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (semeval 2015)* (pp. 486–495).

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I. & Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (semeval 2014)* (pp. 27–35).

Popescu, A.-M. & Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 339–346).

Poria, S., Cambria, E. & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, *108*, 42–49.

Poria, S., Cambria, E., Ku, L.-W., Gui, C. & Gelbukh, A. (2014). A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (socialnlp)* (pp. 28–37).

Qi, P., Zhang, Y., Zhang, Y., Bolton, J. & Manning, C. D. (2020). 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* (pp. 101–108).

Qiu, G., Liu, B., Bu, J. & Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. In *Twenty-first international joint conference on artificial intelligence.*

Qiu, G., Liu, B., Bu, J. & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, *37*(1), 9–27.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. et al. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., … others (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, *21*(140), 1–67.

Rana, T. A. & Cheah, Y.-N. (2016). Aspect extraction in sentiment analysis: comparative analysis and survey. *Artificial Intelligence Review*, *46*, 459–483.

Rana, T. A., Cheah, Y.-N. & Letchmunan, S. (2016). Topic modeling in sentiment analysis: A systematic review. *Journal of ICT Research & Applications*, *10*(1).

Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., … Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*, *74*, 634–642.

Rathan, M., Hulipalled, V. R., Venugopal, K. & Patnaik, L. (2018). Consumer insight mining: aspect based Twitter opinion mining of mobile phone reviews. *Applied Soft Computing*, *68*, 765–773.

Riquelme, F. & González-Cantergiani, P. (2016). Measuring user influence on twitter: A survey. *Information Processing & Msanagement*, *52*(5), 949–975.

Santoso, J., Setiawan, E. I., Purwanto, C. N., Yuniarno, E. M., Hariadi, M. & Purnomo, M. H. (2021). Named entity recognition for extracting concept in ontology building on indonesian language using end-to-end bidirectional long short term memory. *Expert Systems with Applications*, *176*, 114856.

Savoy, O. K. (2012). Feature selection in sentiment analysis. In *Proceedings of the 9th french information retrieval conference* (pp. 273–284).

Schouten, K., Van Der Weijde, O., Frasincar, F. & Dekker, R. (2017). Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. *IEEE transactions on cybernetics*, *48*(4), 1263–1275.

Schütze, H., Manning, C. D. & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press Cambridge.

Seewald, A. K. & Kleedorfer, F. (2007). Lambda pruning: An approximation of the string subsequence kernel for practical SVM classification and redundancy clustering. *Advances in Data Analysis and Classification*, *1*(3), 221–239.

Sharma, R., Nigam, S. & Jain, R. (2014). Opinion mining of movie reviews at document level. *arXiv preprint arXiv:1408.3829*.

Shi, J., Li, W., Bai, Q. & Ito, T. (2022). Beeae: effective aspect term extraction with artificial bee colony. *The Journal of Supercomputing*, 1–23.

Shi, J., Li, W., Bai, Q., Yang, Y. & Jiang, J. (2022). A unified syntax-enhanced network for aspect-based sentiment analysis. *Available at SSRN 4186314*.

Shi, J., Li, W., Bai, Q., Yang, Y. & Jiang, J. (2023). Soft prompt guided joint learning for cross-domain sentiment analysis. *arXiv preprint arXiv:2303.00815*.

Shi, J., Li, W., Yang, Y., Yao, N., Bai, Q., Yongchareon, S. & Yu, J. (2021). Automated concern exploration in pandemic situations-covid-19 as a use case. In *Pacific rim knowledge acquisition workshop* (pp. 178–185).

Shi, J., Li, W., Yongchareon, S., Yang, Y. & Bai, Q. (2022). Graph-based joint pandemic concern and relation extraction on twitter. *Expert Systems with Applications*, *195*, 116538.

Shi, Y., Zhu, L., Li, W., Guo, K. & Zheng, Y. (2019). Survey on classic and latest textual sentiment analysis articles and techniques. *International Journal of Information Technology & Decision Making*, *18*(04), 1243–1287.

Shunmugapriya, P., Kanmani, S., Supraja, R., Saranya, K. et al. (2013). Feature selection optimization through enhanced artificial bee colony algorithm. In *Proceedings of the 2013 international conference on recent trends in information technology* (pp. 56–61).

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y. & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631–1642).

Sun, K., Zhang, R., Mensah, S., Mao, Y. & Liu, X. (2019). Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 5679–5688).

Sun, L., Li, S., Li, J. & Lv, J. (2014). A novel context-based implicit feature extracting method. In *2014 international conference on data science and advanced analytics (dsaa)* (pp. 420–424).

Sun, X., Zhou, J., Wang, S., Li, X., Zheng, B. & Liu, D. (2022). Linguistic dependency guided graph convolutional networks for named entity recognition. In *Advanced data mining and applications: 17th international conference, adma 2021, sydney, nsw, australia, february 2–4, 2022, proceedings, part ii* (pp. 237–248).

Sungheetha, A. & Sharma, R. (2020). Transcapsule model for sentiment classification. *Journal of Artificial Intelligence*, *2*(03), 163–169.

Szomszor, M., Kostkova, P. & St Louis, C. (2011). Twitter informatics: Tracking and understanding public reaction during the 2009 swine flu pandemic. In *2011 ieee/wic/acm international conferences on web intelligence and intelligent agent technology* (pp. 320–323).

Tan, X., Cai, Y., Xu, J., Leung, H.-F., Chen, W. & Li, Q. (2020). Improving aspect-based sentiment analysis via aligning aspect embedding. *Neurocomputing*, *383*, 336–347.

Tang, D., Qin, B., Feng, X. & Liu, T. (2016). Effective lstms for target-dependent sentiment classification. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 3298–3307).

Tang, D., Qin, B. & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1422–1432).

Tang, D., Qin, B. & Liu, T. (2016). Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900.*

Tian, Y., Chen, G. & Song, Y. (2021). Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 2910–2922).

Toh, Z. & Wang, W. (2014). Dlirec: Aspect term extraction and term polarity classification system. In *Proceedings of the 8th international workshop on semantic evaluation (semeval 2014)* (pp. 235–240).

Too, J. & Mirjalili, S. (2021). A hyper learning binary dragonfly algorithm for feature selection: A covid-19 case study. *Knowledge-Based Systems*, *212*, 106553.

Toprak, C., Jakob, N. & Gurevych, I. (2010). Sentence and expression level annotation of opinions in user-generated discourse. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 575–584).

Tulkens, S. & van Cranenburgh, A. (2020). Embarrassingly simple unsupervised aspect extraction. *arXiv preprint arXiv:2004.13580*.

Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424).

Van Der Vegt, I. & Kleinberg, B. (2020). Women worry about family, men about the economy: Gender differences in emotional responses to COVID-19. In *International conference on social informatics* (pp. 397–409).

Van Lent, L. G., Sungur, H., Kunneman, F. A., Van De Velde, B. & Das, E. (2017). Too far to care? measuring public attention and fear for Ebola using Twitter. *Journal of Medical Internet Research*, *19*(6), e7219.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st conference on neural information processing systems* (pp. 5998–6008).

Venugopalan, M. & Gupta, D. (2022). An enhanced guided lda model augmented with bert based semantic strength for aspect term extraction in sentiment analysis. *Knowledge-Based Systems*, *246*, 108668.

Veyseh, A. P. B., Nouri, N., Dernoncourt, F., Dou, D. & Nguyen, T. H. (2020). Introducing syntactic structures into target opinion word extraction with deep learning. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 8947–8956).

Wahbeh, A., Nasralah, T., Al-Ramahi, M. & El-Gayar, O. (2020). Mining physicians' opinions on social media to obtain insights into COVID-19: Mixed methods analysis. *Journal of Medical Internet Research*, *6*(2), e19276.

Wan, H., Yang, Y., Du, J., Liu, Y., Qi, K. & Pan, J. Z. (2020). Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 34, pp. 9122–9129).

Wang, B. & Wang, H. (2008). Bootstrapping both product features and opinion words from chinese customer reviews with cross-inducing. In *Proceedings of the third international joint conference on natural language processing: Volume-i*.

Wang, T., Cai, Y., Leung, H.-f., Lau, R. Y., Li, Q. & Min, H. (2014). Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Systems*, *71*, 86–100.

Wang, T., Lu, K., Chow, K. P. & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. *IEEE Access*, *8*, 138162–138169.

Wang, W. & Pan, S. J. (2018). Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 2171–2181).

Wang, W. & Pan, S. J. (2019a). Syntactically meaningful and transferable recursive neural networks for aspect and opinion extraction. *Computational Linguistics*, *45*(4), 705–736.

Wang, W. & Pan, S. J. (2019b). Transferable interactive memory network for domain adaptation in fine-grained opinion extraction. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 33, pp. 7192–7199).

Wang, W., Pan, S. J., Dahlmeier, D. & Xiao, X. (2016). Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 616–626).

Wang, W., Pan, S. J., Dahlmeier, D. & Xiao, X. (2017). Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of the 2017 aaai conference on artificial intelligence* (Vol. 31).

Wang, X., Xu, H., Sun, X. & Tao, G. (2020). Combining fine-tuning with a feature-based approach for aspect extraction on reviews. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 34, pp. 13951–13952).

Wei, Y., Zhang, H., Fang, J., Wen, J., Ma, J. & Zhang, G. (2021). Joint aspect terms extraction and aspect categories detection via multi-task learning. *Expert Systems with Applications*, *174*, 114688.

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

Wu, H. & Shi, X. (2022). Adversarial soft prompt tuning for cross-domain sentiment analysis. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 2438–2447).

Wu, M., Wang, W. & Pan, S. J. (2020). Deep weighted maxsat for aspect-based opinion extraction. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 5618–5628).

Wu, S., Fei, H., Ren, Y., Ji, D. & Li, J. (2021). 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* (p. 3957–3963).

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … others (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Wu, Y., Zhang, Q., Huang, X.-J. & Wu, L. (2009). Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 1533–1541).

Wu, Z. & Ong, D. C. (2021). Context-guided bert for targeted aspect-based sentiment analysis. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 14094–14102).

Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X. & Xia, R. (2020). Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the association for computational linguistics: Emnlp 2020* (pp. 2576–2585).

Wu, Z., Zhao, F., Dai, X.-Y., Huang, S. & Chen, J. (2020). Latent opinions transfer network for target-oriented opinion words extraction. In *Proceedings of the 2020 aaai conference on artificial intelligence* (Vol. 34, pp. 9298–9305).

Xu, H., Liu, B., Shu, L. & Philip, S. Y. (2018). Double embeddings and cnn-based sequence labeling for aspect extraction. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 592–598).

Xu, H., Liu, B., Shu, L. & Philip, S. Y. (2019). Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 2324–2335).

Xu, H., Zhang, F. & Wang, W. (2015). Implicit feature identification in chinese reviews using explicit topic mining model. *Knowledge-Based Systems*, *76*, 166–175.

Xu, L., Bing, L., Lu, W. & Huang, F. (2020). Aspect sentiment classification with aspect-specific opinion spans. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 3561–3567).

Xu, L., Chia, Y. K. & Bing, L. (2021). 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* (pp. 4755–4766).

Xu, L., Li, H., Lu, W. & Bing, L. (2020). Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 2339–2349).

Xu, L., Liu, K., Lai, S., Chen, Y. & Zhao, J. (2013). Mining opinion words and opinion targets in a two-stage framework. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 1764–1773).

Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y. & Zhu, T. (2020). Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. *Journal of Medical Internet Research*, *22*(11), e20550.

Xue, K., Zhou, Y., Ma, Z., Ruan, T., Zhang, H. & He, P. (2019). Fine-tuning bert for joint entity and relation extraction in chinese medical text. In *2019 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 892–897).

Xue, Y., Xue, B. & Zhang, M. (2019). Self-adaptive particle swarm optimization for large-scale feature selection in classification. *ACM Transactions on Knowledge Discovery from Data*, *13*(5), 1–27.

Yan, H., Dai, J., Ji, T., Qiu, X. & Zhang, Z. (2021). 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* (pp. 2416–2429).

Yang, B. & Cardie, C. (2013). Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st annual meeting of the association for computational* *linguistics (volume 1: Long papers)* (pp. 1640–1649).

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

Yang, Y. & Eisenstein, J. (2015). Unsupervised multi-domain adaptation with feature embeddings. In *Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics: human language technologies* (pp. 672–682).

Yao, L., Mao, C. & Luo, Y. (2019). Graph convolutional networks for text classification. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 33, pp. 7370–7377).

Yin, H., Yang, S. & Li, J. (2020). Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. In *International conference on advanced data mining and applications* (pp. 610–623).

Yin, Y., Wang, C. & Zhang, M. (2020). Pod: Positional dependency-based word embedding for aspect term extraction. In *Proceedings of the 28th international conference on computational linguistics* (pp. 1714–1719).

Yin, Y., Wei, F., Dong, L., Xu, K., Zhang, M. & Zhou, M. (2016). Unsupervised word and dependency path embeddings for aspect term extraction. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence* (pp. 2979–2985).

Yu, B., Zhang, Z., Shu, X., Wang, Y., Liu, T., Wang, B. & Li, S. (2020). Joint extraction of entities and relations based on a novel decomposition strategy. In *Proceedings of the 24th european conference on artificial intelligence* (pp. 2282–2289).

Yu, J., Gong, C. & Xia, R. (2021). Cross-domain review generation for aspect-based sentiment analysis. In *Findings of the association for computational linguistics: Acl-ijcnlp 2021* (pp. 4767–4777).

Yu, J., Zha, Z.-J., Wang, M., Wang, K. & Chua, T.-S. (2011). Domain-assisted product aspect hierarchy generation: towards hierarchical organization of unstructured consumer reviews. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 140–150).

Zainuddin, N., Selamat, A. & Ibrahim, R. (2018). Hybrid sentiment classification on twitter aspect-based sentiment analysis. *Applied Intelligence*, *48*, 1218–1232.

Zeng, D., Liu, K., Lai, S., Zhou, G. & Zhao, J. (2014). Relation classification via convolutional deep neural network. In *Proceedings of the 25th international conference on computational linguistics: Technical papers* (pp. 2335–2344).

Zeng, D., Zhang, H. & Liu, Q. (2020). Copymtl: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *Proceedings of the 34th aaai conference on artificial intelligence* (pp. 9507–9514).

Zeng, X., Zeng, D., He, S., Liu, K. & Zhao, J. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 506–514).

Zhang, C., Li, Q. & Song, D. (2019). Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 conference on* *empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 4568–4578).

Zhang, C., Li, Q., Song, D. & Wang, B. (2020). A multi-task learning framework for opinion triplet extraction. In *Findings of the association for computational linguistics: Emnlp 2020* (pp. 819–828).

Zhang, L., Liu, B., Lim, S. H. & O'Brien-Strain, E. (2010). Extracting and ranking product features in opinion documents. In *Coling 2010: posters* (pp. 1462–1470).

Zhang, M., Palade, V., Wang, Y. & Ji, Z. (2021). Attention-based word embeddings using artificial bee colony algorithm for aspect-level sentiment classification. *Information Sciences*, *545*, 713–738.

Zhang, M., Zhang, Y. & Vo, D. T. (2015). Neural networks for open domain targeted sentiment. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 612–621).

Zhang, X., Jiang, Y., Peng, H., Tu, K. & Goldwasser, D. (2017). Semi-supervised structured prediction with neural crf autoencoder. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 1701–1711).

Zhang, Y., Qi, P. & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2205–2215).

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

Zhao, H., Huang, L., Zhang, R., Lu, Q. & Xue, H. (2020). Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3239–3248).

Zhao, X., Jiang, J., Yan, H. & Li, X. (2010). Jointly modeling aspects and opinions with a maxent-lda hybrid..

Zhao, Z., Rao, G. & Feng, Z. (2017). Dfds: a domain-independent framework for document-level sentiment analysis based on rst. In *Web and big data: First international joint conference, apweb-waim 2017, beijing, china, july 7–9, 2017, proceedings, part i 1* (pp. 297–310).

Zheng, S., Wang, F., Bao, H., Hao, Y., Zhou, P. & Xu, B. (2017). Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th annual meeting of the association for computational linguistics* (pp. 1227–1236).

Zhou, G. & Su, J. (2002). Named entity recognition using an HMM-based chunk tagger. In *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 473–480).

Zhou, G.-D. & Zhu, Q.-M. (2011). Kernel-based semantic relation detection and classification via enriched parse tree structure. *Journal of Computer Science and Technology*, *26*(1), 45–56.

Zhou, Y., Zhu, F., Song, P., Han, J., Guo, T. & Hu, S. (2021). An adaptive hybrid framework for cross-domain aspect-based sentiment analysis. In *Proceedings of* *the aaai conference on artificial intelligence* (Vol. 35, pp. 14630–14637).

Zhuang, L., Jing, F. & Zhu, X.-Y. (2006). Movie review mining and summarization. In *Proceedings of the 15th acm international conference on information and knowledge management* (pp. 43–50).

Ziser, Y. & Reichart, R. (2018). Pivot based language modeling for improved neural domain adaptation. In *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long papers)* (pp. 1241–1251).

Zola, P., Cortez, P., Ragno, C. & Brentari, E. (2019). Social media cross-source and cross-domain sentiment classification. *International Journal of Information Technology & Decision Making*, *18*(05), 1469–1499.

Zorarpacı, E. & Özel, S. A. (2016). A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications*, *62*, 91–103.

Appendix A

Glossary

NLP Natural Language Processing

SA Sentiment Analysis

ABSA Aspect-Based Sentiment Analysis

ATE Aspect Term Extraction

OTE Opinion Term Extraction

ALSC Aspect-Level Sentiment Classification

AOE Aspect-oriented Opinion Extraction

AESC Aspect Extraction and Sentiment Classification

AOPE Aspect-Opinion Pair Extraction

ASTE Aspect Sentiment Triplet Extraction

ELMO Embeddings from Language Models

BERT Bidirectional Encoder Representations from Transformers

T5 Text-To-Text Transfer Transformer

RNN Recurrent Neural Network

LSTM Long Short-Term Memory

GRU Gated Recurrent Unit

CNN Convolutional Neural Network

GCN Graph Convolutional Network

NMT Neural Machine Translation

NER Named Entity Extraction

LM Language Model

CV Computer Vision

P Precision

TP True Positive

FP False Positive

R Recall

GRNN Gated Recurrent Neural Network

BiLSTM Bidirectional Long Short-Term Memory

DP Dependency Parse

LDA Latent Dirichlet Allocation

ME Maximum Entropy

SVM Support Vector Machine

CRF Conditional Random Field

HMM Hidden Markov Model

HSBi-GRU Hierarchical Stack Bidirectional Gated Recurrent Units

MTMRC Multi-Turn Machine Reading Comprehension

PLM Pre-trained Language Model

ABC Artificial Bee Colony

PoS Part-of-Speech

PSO Particle Swarm Optimization

DT Decision Tree

SO Semantic Orientation

ACE Automated Concern Exploration

CG Concern Graph

RE Relation Extraction

CG-CRE Concern Graph-based Concern and Relation Extraction

MSS Multiple Syntactic Structure

FFN Feed-Forward Network

SemEval Semantic Evaluation

FIN Finance

GOV Government

DIS Disease

MED Medicine

PER Person

LOC Location

FOD Food

DAT Date and Time

TER Tail-Entity-Relation