Intelligent Systems with Applications 20 (2023) 200292

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

Soft prompt enhanced joint learning for cross-domain aspect-based sentiment analysis

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

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

ARTICLE INFO

Keywords:
Aspect-based sentiment analysis
Cross-domain
Soft prompt

ABSTRACT

Aspect term extraction is a fundamental task in fine-grained sentiment analysis, aiming to detect customer's opinion targets from reviews about products or services. The traditional supervised models have achieved promising results with annotated datasets. However, their performance dramatically decreases in cross-domain aspect term extraction tasks. Existing cross-domain transfer learning methods face two common limitations: (1) these works directly inject linguistic features into language models, making it challenging to transfer linguistic knowledge to the target domain; (2) they rely on the fixed predefined prompts, which is time-consuming to construct the prompts for all potential aspect term spans. To address the limitations, we propose a soft prompt-based joint learning method for cross-domain aspect term extraction in this paper. 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. Furthermore, 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 two groups of datasets and the experimental results demonstrate the effectiveness of the proposed method for cross-domain aspect terms extraction.

1. Introduction

Gaining specialist knowledge in business development requires the ability to quickly understand customers' complaints and requirements by analysing their feedback. One effective method for doing so is Aspect-Based Sentiment Analysis (ABSA), which involves extracting aspect and opinion terms and identifying their corresponding sentiments in customer reviews (Liu, 2012, Pontiki et al., 2016). The majority of the existing studies address ABSA tasks independently (Zhong et al., 2023, Liu et al., 2023) or treat some sub-tasks as one task (Luo et al., 2019, Akhtar et al., 2020). In this paper, we study a crucial sub-task of ABSA, named Aspect Term Extraction (ATE), which aims to identify opinion targets from customer reviews. For example, in Fig. 1, the task is expected to detect the aspect term *Keyboard* from the sentence "*Keyboard responds well to presses*". To generate all possible prompt templates "*The aspect is* _.", we need to enumerate all possible term spans (e.g., *keyboard*, *keyboard responds*, *presses*, etc.) in the input sentence as aspect term candidates. The computational complexity is $O(n^2)$ for the input sentence with n words.

Recently, ATE has been widely studied in literature with the emergence of pre-trained language models (PLMs), such as BERT (Devlin et al., 2019), 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 the ATE task (Xu et al., 2018, Wang et al., 2020, Wan et al., 2020, Gao, Wang 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

* Corresponding author.
E-mail addresses: jingli.shi@aut.ac.nz (J. Shi), weihua.li@aut.ac.nz (W. Li), quan.bai@utas.edu.au (Q. Bai), yyang@hfut.edu.cn (Y. Yang), jjh@jlufe.edu.cn (J. Jiang).

https://doi.org/10.1016/j.iswa.2023.200292
Received 8 March 2023; Received in revised form 16 September 2023; Accepted 17 October 2023
Available online 20 October 2023

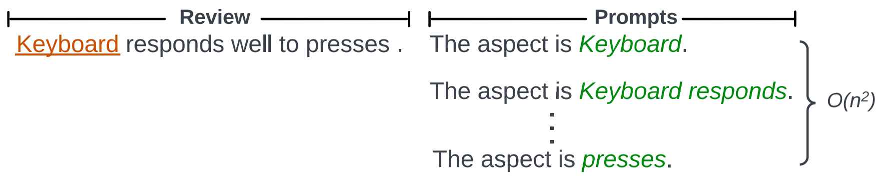

Fig. 1. The traditional inputs of the prompt tuning model for aspect term extraction.

learning method is designed, named prompt tuning, to reformulate NLP tasks as cloze-style question answering (Lester et al., 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 (Chen et al., 2022), sentiment classification (Li et al., 2021), NER (Chen et al., 2021)). By manually designing prompts, similar attempts have been made on aspect term extraction to detect aspect terms from each text span in a review (Gao et al., 2022, Li, Yang et al., 2022, Li et al., 2021). As shown in Fig. 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 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 adaptation (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.

In the existing research works of cross-domain aspect term extraction (ATE) models, several research efforts have been conducted on transferring learned knowledge from the source domain to the target domain. However, these models still suffer from challenges that require further innovation. The motivation behind this proposed method arises from the need to alleviate the limitations of existing approaches. To address the aforementioned challenges in the task of cross-domain aspect term extraction, we propose a joint learning method, which is the first to use soft prompt integrating with transferable linguistic knowledge to solve the domain adaptation problem. Unlike previous works, our method pioneers the integration of soft prompts within joint training to tackle the cross-domain ATE problem. By employing soft prompts, we overcome the time-consuming issue associated with hard prompts, which require an exhaustive enumeration of prompt queries across potential aspect spans. Additionally, our method introduces learnable linguistic features, serving as an enhancement component to bridge the gap between different domains and capture domain-invariant aspects of ATE. To effectively transfer knowledge to target domain, multiple source domains based adaptation is designed by applying domain-invariant and domain-specific features, which are proved to be relevant to the target domain (Ziser & Reichart, 2017). Through efficient knowledge transfer from source domains, our proposed method achieves outstanding performance across multiple real-world datasets of cross-domain ATE. Furthermore, our analysis of experimental results highlights the significant contributions of soft prompts and learnable syntactic features in enhancing overall performance. By introducing these innovative elements, our method addresses existing challenges and provides a novel way for further advancements in cross-domain aspect term extraction.

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

- We propose a novel joint learning method for cross-domain ATE tasks. To the best of our knowledge, we are the first attempt to solve the domain adaptation problem via soft prompts.
- The learnable soft prompts are designed with rich external syntactic knowledge to better leverage the domain-specific and domain-invariant knowledge across domains.
- We evaluate our method on both large and small datasets, demonstrating that it can be effectively applied to the other application scenarios for cross-domain ABSA tasks (e.g., opinion term extraction, aspect-opinion pair extraction, sentiment triplet extraction, etc.).

The remainder of the paper is organized as follows. In Section 2, related works are reviewed in the cross-domain aspect term extraction. Section 3 formally defines the relevant concepts and formulates the problem. The proposed method is introduced in Section 4. The experimental work and results are presented and discussed in Section 5. Finally, the conclusions and future work are described in Section 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 high 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: rule-based model, fine-tuning PLM, and prompt-tuning PLM.

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 (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 do not 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- (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 (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 (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 (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 (Li et al., 2019). Despite the outstanding performance, neural network-based methods fail to obtain the satisfactory quality of domain-invariant features and exploit the significant supervision signals in the target domains, which leads to low precision results.

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 the language model for ATE (Gong et al., 2020). Anand et al. apply the evolutionary approach to automatically learn linguistic patterns of aspect words, which mitigates 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 the source domain to generate data in the target domain for ATE model training (Li, Yu, & Xia, 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-based BERT to capture domain-invariant features. However, all the language model-based methods rely heavily on annotated resources, the performance of fine-tuning language models may be unstable on a small-scale data. Moreover, most of existing methods only integrate the linguistic features directly into language models, which cannot achieve word-level adaptation for aspect extraction.

2.3. Prompt-based model

With the release of OpenAI's Generative Pretrained Transformer (GPT), prompt engineering emerges to migrate PLMs to downstream tasks. There are two types of prompt: (1) hard prompt, which is manually designed by human; (2)soft prompt, which is created during the process of prompt tuning (Lester et al., 2021). To address the learning challenges caused by the 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 a prompt-based model for aspect-based sentiment analysis subtasks, in which sentiment knowledge prompts are constructed by integrating features from aspects, opinions, and polarities (Li et al., 2021). Gao et al. introduce a unified generative framework to solve different ABSA tasks by controlling the type of task prompts (Gao et al., 2022). By assembling prompts for 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 (Li, Yang et al., 2022). Ben et al. present an example-based prompt learning method, which can be applied to unseen domains of multiple tasks, namely rumour detection, multi-genre natural language inference, and aspect prediction (Ben-David et al., 2022). However, the 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.

To transfer knowledge from one task to another, Vu et al. propose Soft Prompt Transfer (SPoT), which first learns a prompt on one or more source tasks, and then uses it to initialise the prompt of a target task (Vu et al., 2022). SPot can boost the performance of prompt tuning across many tasks. Zhong et al. develop a novel method to leverage the knowledge distillation technique to transfer knowledge from source prompt to target prompt (Zhong et al., 2022). Moreover, they design a new metric to achieve adaptive prompt transfer. 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 (Wu & Shi, 2022). Asai et al. introduce a multi-task language model tuning method that transfers knowledge across different tasks via 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 paper, 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 capture 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 the performance improvements.

3. Problem formulation

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

Formally, the proposed method formulates the task of cross-domain aspect term extraction as a sequential tagging problem. The datasets of two domains are given, i.e., $\mathbb{D}_s$ and $\mathbb{D}_t$ which represent the source and target domains, respectively. In the source domain, $\mathbb{D}_s = \{S_s^i, y_s^i\}_{i=1}^{N_s}$ refer to N_s annotated sentences, where S_s^i is the ith sentence, and $y_s^i \in \{B, I, O\}$ denotes the corresponding aspect label. In the 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 objective of cross-domain ATE is to learn a function, which can learn both in-domain and domain-invariant knowledge between source and target domains to better predict token-level labels in the target domain.

4. Soft prompt-based joint learning model

In this section, we first describe the overview of the proposed method. Then we introduce each module from bottom to up in the

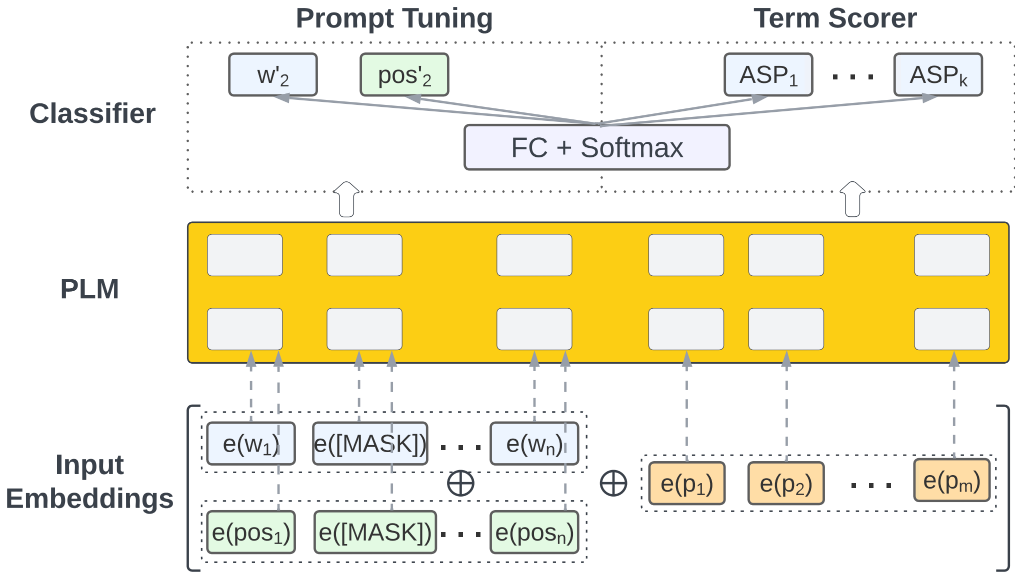

Fig. 2. Overview architecture of the soft prompt-based joint learning model.

whole architecture. Finally, we present the learning objective for cross-domain ATE.

The overall architecture for our feature-based domain adaptation component is shown in Fig. 2. The proposed model consists of three layers: (1) Input Embeddings, in which input words, syntax terms, and prompts are converted into embeddings as input of next layer. (2) PLM, refers to the pre-trained language model, we propose to apply T5 as the backbone model of PLM. (3) Classifier, is the final layer of the proposed method. The output representations from the previous layer 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.

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 types of embeddings: (a) word embedding e^w is obtained via the pre-trained language model by Equation (1). (b) syntax embedding e^{pos} is calculated in Equation (2). To leverage the domain-invariant features more effectively, 25% of the 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 (3). Inspired by previous work on prompt features (Ziser & Reichart, 2018, Ben-David et al., 2020, 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.

$$e^w = T5(\{w_1, w_2, ..., w_n\}) \tag{1}$$

$$e^{pos} = T5(\{t_1, [MASK], ..., t_n\}) \tag{2}$$

$$e^p = T5(\{p_1, ..., p_m\}) \tag{3}$$

4.2. Pre-trained language model

To automatically generate the prompts for aspect extraction, a large pre-trained text-to-text transformer, T5, is used as backbone model. During the inference, T5 is pre-trained to generate prompt spans based on the input words shown in Fig. 3. Given an input text $\{w_1, w_2, ..., w_n\}$,

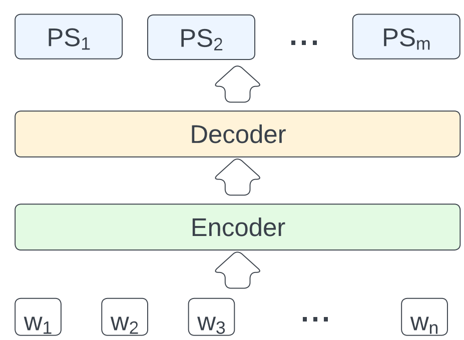

Fig. 3. The PLM module in the proposed model.

the decoder of T5 is able to generate prompt spans $\{ps_1, ps_2, ..., ps_m\}$ where m is the total number of spans in one prompt.

4.3. 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. We 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 (5).

$$\hat{y}^p = softmax(W^p * [e^w; e^{pos}; e^p] + b^p), \tag{4}$$

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).

$$\hat{y}^p = softmax(W^p * [e^w; e^{pos}; e^p] + b^p), \tag{5}$$

$$\mathcal{L}_{prompt} = \sum^{\mathbb{D}_s} \sum_i^n f(\hat{y}_i^p, y_i^p), \tag{6}$$

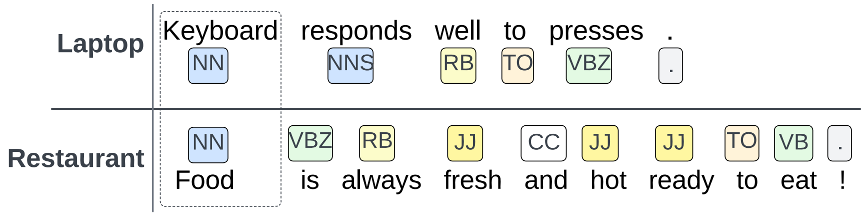

Fig. 4. POS tags of reviews from laptop and restaurant domain. The aspect terms *Keyboard* and *Food* share same POS tag *NN*.

4.4. Syntax learning

For aspect terms from different domains, their linguistic features maintain often-occurring patterns (Hu & Liu, 2004, Qiu et al., 2011, Chen & Qian, 2021). An example is given in Fig. 4. 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 (7).

$$\hat{y}^{pos} = softmax(W^{pos} * [e^{w}; e^{pos}; e^{p}] + b^{pos}), \tag{7}$$

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 (8).

$$\mathcal{L}_{syntax} = \sum^{\mathbb{D}_s} \sum_{i}^{n} I(i) * f(\hat{y}_i^{pos}, y_i^{pos}), \tag{8}$$

$$I(i) = \begin{cases} 1 & if\ token\ is\ masked \\ 0 & else \end{cases} \tag{9}$$

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.

4.5. Training objective

Given the source domain data and the target data, 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 a weighted sum of the cross-entropy losses from syntax learning and multiple domain knowledge enhanced prompt-tuning in Equation (10).

$$\mathcal{L}(\theta) = \alpha * \mathcal{L}_{prompt} + \beta * \mathcal{L}_{syntax} \tag{10}$$

where α and β are the trade-off parameters.

4.6. Cross-domain aspect extraction

To explain the process of cross-domain aspect term extraction, we provide an illustrative example using customer review from laptop domain, as shown in Fig. 5. The aspect term extraction consists four key steps during the inference stage.

(1) The customer review is converted into tokens with special tokens as beginning and ending, [CLS] and [SEP] respectively. Additional, the Extended Part-of-Speech (XPOS) tags for each word are extracted using the NLP tool (i.e., Stanza[1]).

(2) During the model inference stage, prompts are generated from the decoder of Soft Prompt Enhanced Joint Learning (SP-JL) model after incorporating both the review and syntax information.

(3) The hidden states from SP-JL are passed through a fully-connected feed forward neural network, which aims to learn task-specific features from the representations generated from SP-JL.

(4) The outputs from the previous layer are fed into a Softmax activation function, which is the final layer of the proposed method. For each token, the proposed model directly predicts one of three classes: B (Beginning), I (Inside), and O (Outside), which are a standard tagging scheme for entity extraction task (Ratinov & Roth, 2009). Finally, the aspect term *"Keyboard"* is identified in this example.

5. Experiments

In this section, extensive experiments are conducted on two groups of datasets to evaluate the performance of the proposed model. The first experiment is conducted on a large-scale dataset with four domains, which aims to show that the proposed soft prompt-based method outperforms fine-tuning models. The second experiment performed on a small-scale dataset with three domains is to reveal that our model has the potential to be used in a wide range of ABSA scenarios.

5.1. Dataset

The experiments of our method are conducted on two groups of benchmark datasets with different domains. The first group, $\mathbb{G}1$, consists of four domains: Device ($\mathbb{D}$), Laptop ($\mathbb{L}$), and Restaurant ($\mathbb{R}$), and Service ($\mathbb{S}$). The basic statistics of the first group of the dataset are presented in Table 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 2. The numbers of sentences and aspect terms are presented in training and test datasets of $\mathbb{DI}$, $\mathbb{AS}$, and $\mathbb{E}$.

- ***Device*** **($\mathbb{D}$)** (Hu & Liu, 2004) is the set of customer reviews on digital device, which is collected from Amazon and CNET. This dataset includes a variety of digital devices (e.g., camera, DVD, MP3, etc.). There are 3,836 sentences, 2,131 aspect terms. For model training, 2,557 reviews from the dataset are utilized, and the remaining 1,279 reviews are used for model testing.
- ***Laptop*** **($\mathbb{L}$)** (Pontiki et al., 2014, 2015, 2016) is the customer reviews on laptop. This annotated dataset is obtained from SemEval, which is an internal workshop on semantic evaluation. The laptop dataset consists of 3,845 sentences and 2,939 labelled aspect terms. To facilitate the experiments, the dataset is split into two subsets: a training subset including 3,045 reviews and testing subset consisting of 800 reviews.
- ***Restaurant*** **($\mathbb{R}$)** (Pontiki et al., 2014, 2015, 2016) is obtained from SemEval, which consists of 6,035 sentences with 6,656 aspect terms. To facilitate the experimentation process, the dataset is divided into two distinct subsets: a training set comprising 3,877 reviews and a testing set containing 2,158 reviews.
- ***Service*** **($\mathbb{S}$)** refers to the customer reviews specifically related to web services (Toprak et al., 2010). It was collected from the review portals at rateitall[2]. This dataset comprises 2,239 sentences, with a total of 2,734 annotated aspect terms. In our research experiments, the dataset is split into two subsets: a training subset consisting of 1,492 reviews and a testing subset comprising 747 reviews.

[1] https://stanfordnlp.github.io/stanza/.

[2] http://www.rateitall.com.

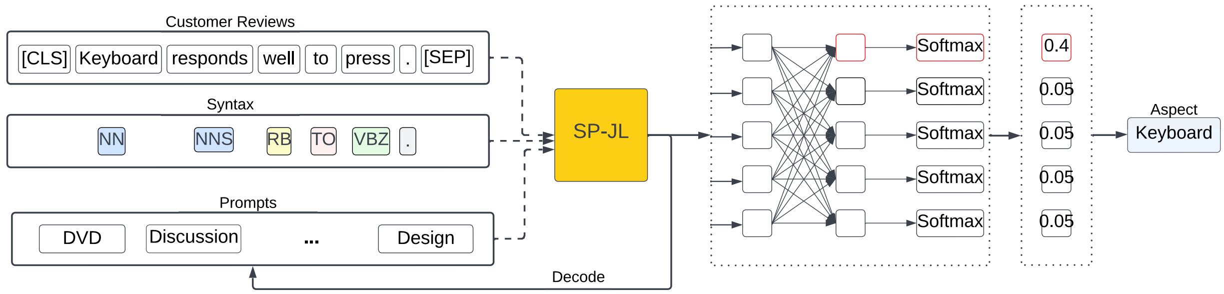

Fig. 5. The illustration of cross-domain aspect term extraction. SP-JL means the proposed soft prompt enhanced joint learning model.

Table 1
Statistics of the first group datasets.

Domain	Sentences	Aspect Terms	Train	Test
ℝ	6035	6656	3877	2158
𝕃	3845	2939	3045	800
𝔻	3836	2131	2557	1279
𝕊	2239	2734	1492	747

Table 2
Statistics of the second group datasets.

Domain	Sentences	Aspect Terms	Train	Test
𝔻𝕀	375	377	262	113
𝔸𝕊	380	394	266	114
𝔼	550	561	385	165

- ***Diapers*** **(𝔻𝕀)** is annotated on customer reviews of diaper for opinion mining (Ding et al., 2008). The reviews are collected from amazon.com, which includes 375 sentences and 377 aspect terms. The training and testing subsets are 262 and 113, respectively.
- ***Antivirus Software*** **(𝔸𝕊)** is customer reviews annotated for sentiment analysis[3] by (Ding et al., 2008). 𝔸𝕊 contains 380 sentences and 394 aspect terms. In our experiment, it is split into 266 training and 114 testing subsets.
- ***Electronics*** **(𝔼)** is annotated by our annotators, which is originally collected for cross-domain sentiment classification[4] by (Zola et al., 2019). In this dataset, 561 aspect terms are labelled in 550 sentences. The training and testing are 385 and 165, respectively.

5.2. Implementation and hyper-parameters

We utilise the Pytorch framework[5] to implement our model. T5-base[6] is used as the base LMs. We use Stanford NLP Toolkit[7] 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 dimensional hidden vectors. The Adam optimiser (Kingma & Ba Adam, 2015) is applied with an initial learning rate of 2e-3. The epoch is set to 20, and the batch size is 16.

5.3. Baselines

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

- ***CrossCRF*** (Jakob & Gurevych, 2010) is a traditional sequence labelling method, 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*** (Chen et al., 2012) is a marginalised stacked de-noising 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*** (Yang & Eisenstein, 2015) performs dense feature representation learning, which are more robust to domain shift, using neural language models to obtain low-dimensional embeddings directly.
- ***RNCRF*** (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*** (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*** (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*** (Li et al., 2019) firstly explores an unsupervised domain adaptation 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 adaptation.
- ***BERT*** directly fine-tunes base BERT (Devlin et al., 2019) to predict collapsed labels for cross-domain ATE task.
- ***ARNN-GRU*** (Wang & Pan, 2019a) is a dependency-tree-based recursive neural network for cross-domain aspect-based sentiment analysis. GRU is incorporated to reduce label noise with an auto-encoder in the auxiliary task.
- ***TRNN-GRU*** (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*** (Xu et al., 2019) post-trains base BERT (Devlin et al., 2019) on mixed datasets from Yelp and Amazon reviews, and then fine-tunes the trained model to detect aspect terms across domains.

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

- ***CrossBERT-UDA*** (Gong et al., 2020) is an end-to-end framework that performs feature and instance based adaptation 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.
- ***EXAL*** (Pereg et al., 2020) is a baseline model that includes the same size and structure as the SA-EXAL except for the syntactic module.
- ***CDRG-Indep*** (Yu et al., 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*** (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*** (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 refine the task classifier.
- ***SynBridge*** (Chen & Qian, 2021) is an active domain adaptation model that transfers aspect words by actively supplementing transferable knowledge. The syntactic bridges are constructed via recognising syntactic roles as pivots to identify transferable syntactic roles for the words across domains.
- ***SemBridge*** (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*** (Chen & Wan, 2022) is a simple but effective method based on mutual information maximization for cross-domain ABSA tasks.

Table 3
Experimental results of machine learning models for cross-domain ATE on $\mathbb{G}1$.

Model	CrossCRF	DP	mDA	FEMA	RNCRF	Ours
$\mathbb{R} \to \mathbb{L}$	0.197	0.198	0.209	0.266	0.243	**0.593**
$\mathbb{S} \to \mathbb{L}$	0.116	0.198	0.146	0.150	-	**0.480**
$\mathbb{D} \to \mathbb{L}$	0.242	-	0.257	0.268	0.406	**0.527**
$\mathbb{L} \to \mathbb{R}$	0.282	0.376	0.243	0.350	0.409	**0.655**
$\mathbb{S} \to \mathbb{R}$	0.170	0.376	0.325	0.376	-	**0.612**
$\mathbb{D} \to \mathbb{R}$	0.659	0.376	0.213	0.207	0.346	**0.630**
$\mathbb{R} \to \mathbb{D}$	0.211	0.218	0.172	0.229	0.243	**0.563**
$\mathbb{L} \to \mathbb{D}$	0.299	-	0.294	0.296	0.315	**0.434**
$\mathbb{S} \to \mathbb{D}$	0.097	0.218	0.169	0.187	-	**0.586**
$\mathbb{R} \to \mathbb{S}$	0.088	0.197	0.131	0.108	-	**0.528**
$\mathbb{L} \to \mathbb{S}$	0.086	0.197	0.131	0.148	-	**0.555**
$\mathbb{D} \to \mathbb{S}$	0.045	0.197	0.131	0.088	-	**0.570**

Table 4
Experimental results of deep learning models for cross-domain ATE on $\mathbb{G}1$ ($* \to \mathbb{L}$ and $* \to \mathbb{R}$.)

Model	$\mathbb{R} \to \mathbb{L}$	$\mathbb{S} \to \mathbb{L}$	$\mathbb{D} \to \mathbb{L}$	$\mathbb{L} \to \mathbb{R}$	$\mathbb{S} \to \mathbb{R}$	$\mathbb{D} \to \mathbb{R}$
Hier-Joint	0.317	0.300	0.362	0.467	0.520	0.504
RNSCN	0.266	0.189	-	0.356	0.332	0.346
AD-SAL	0.341	0.270	-	0.430	0.410	0.410
BERT	0.314	0.305	-	0.404	0.447	0.403
ARNN-GRU	0.404	-	0.511	0.529	-	0.484
TRNN-GRU	0.402	-	0.517	0.538	-	0.512
CrossBERT	0.397	0.350	-	0.454	0.513	0.426
CrossBERT-UDA	0.439	0.348	-	0.495	0.471	0.427
EXAL	0.440	-	0.458	0.482	-	0.538
SA-EXAL	0.476	-	0.477	0.547	-	0.545
CDRG-Indep	0.402	0.332	-	0.551	0.538	0.501
CDRG-Merge	0.466	0.395	-	0.600	0.563	0.527
AHF	0.557	0.448	-	0.646	0.591	0.597
SynBridge	0.551	0.453	-	0.653	0.584	0.628
SemBridge	0.579	0.451	-	0.662	0.593	0.636
SDAM	0.546	0.467	-	0.631	0.586	0.609
FMIM-BERT	0.494	0.424	-	0.634	0.592	0.573
Ours	**0.593**	**0.480**	**0.527**	0.655	**0.612**	0.630

5.4. Evaluation metrics

In this paper, the standard evaluation metric, F1 score (F1), is employed to evaluate the proposed model. F1 is formulated in Equation (13), where P and R are precision and recall, respectively. For the tasks of aspect-based sentiment analysis and entity extraction, the F1 score is the most widely used evaluation metric (Venugopalan & Gupta, 2022, Alshuwaier et al., 2022, Rizou et al., 2023).

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$F1 = 2 * \frac{P * R}{P + R} \tag{13}$$

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.

5.5. Experimental results and model analysis

We 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 3, 4, and 6. We can observe that our 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 in Table 3, our method is significantly superior to machine learning and based models. In Table 4, the performance of our method is lower than that of SemBridge for domain adaptation $\mathbb{L} \to \mathbb{R}$ and $\mathbb{D} \to \mathbb{R}$ (-0.007 and -0.006, respectively), indicating that SemBridge captures more syntactic and semantic knowledge of source domains and transfers the meaningful knowledge to target domain. Without semantic features, SynBridge achieves a degraded performance compared with the proposed method. Whereas, our 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 5).

Table 6 presents the results of cross-domain ATE on a small-scale of dataset. Compared with fine-tuning based models, AD-SAL and CrossBERT-UDA, our 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-scale and small-scale datasets with competitive performance. Compared with prompt-tuning models, it appears 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.

As shown in Tables 3, 4, and 6, our analysis of the experimental results on both large-scale and small-scale datasets underscores the significant contributions made by soft prompts and learnable linguistic features. The soft prompts alleviate the computational burden associated with hard prompts, enabling efficient and scalable aspect term extraction across domains. Moreover, the soft prompts are able to incorporate domain-dependent and domain-independent features into the proposed model, which benefits the performance of aspect extraction cross do-

Table 5
Experimental results of deep learning models for cross-domain ATE on 𝔾1 (* → 𝔻 and * → 𝕊).

Model	ℝ → 𝔻	𝕃 → 𝔻	𝕊 → 𝔻	ℝ → 𝕊	𝕃 → 𝕊	𝔻 → 𝕊
Hier-Joint	0.320	0.316	0.334	0.198	0.234	0.235
RNSCN	0.333	-	0.220	0.200	0.166	0.200
AD-SAL	0.354	-	0.336	0.280	0.272	0.266
BERT	0.276	-	0.339	0.195	0.258	0.303
ARNN-GRU	0.351	0.404	-	-	-	-
TRNN-GRU	0.373	0.412	-	-	-	-
CrossBERT	0.332	-	0.332	0.244	0.233	0.282
CrossBERT-UDA	0.349	-	0.321	0.331	0.279	0.280
EXAL	0.382	0.416	-	-	-	-
SA-EXAL	0.405	0.422	-	-	-	-
CDRG-Indep	0.308	-	0.349	0.417	0.441	0.371
CDRG-Merge	0.326	-	0.369	0.424	0.471	0.418
AHF	0.502	-	0.478	0.438	0.427	0.444
SynBridge	0.533	-	0.539	0.327	0.337	0.381
SemBridge	0.553	-	0.546	0.350	0.350	0.377
SDAM	0.516	-	0.580	0.456	0.453	0.552
FMIM-BERT	0.397	-	0.376	0.514	0.549	0.528
Ours	**0.563**	**0.434**	**0.586**	**0.528**	**0.555**	**0.570**

mains. Meanwhile, the learnable linguistic features serve as a promising enhancement component, enabling the extraction of domain-invariant features and facilitating robust performance in varying cross-domains.

5.6. Ablation study

To analyse the effect of each component including syntax learning and prompts, the ablation experiments are conducted on datasets 𝔾1 and 𝔾2. The experimental results are shown in Tables 7 and 8.

5.6.1. Effect of syntax learning

In this subsection, the effect of syntax learning is verified via the ablation study. Tables 7 and 8 present the experimental results on datasets 𝔾1 and 𝔾2, respectively. We find that, without the syntax learning component, the results of our method see a decrease in all target domains for both datasets (i.e., -5.5%, -3.3%, -3.8%, -4%, -1.5%, -3.3% on 𝔾2, 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 target domain for cross-domain ATE task. For example, the model trained on domain 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 aspect term *Program* in the sentence "*The program brings more problems than a virus...*" on the target domain 𝔸𝕊.

5.6.2. Effect of prompts

We present an evaluation of the effect of prompts by removing them from our method. In Tables 7 and 8, the F1 scores of w/o prompts are presented across all target domains. After removing prompts, we observe a significant performance drop on all domains for both datasets (i.e., -7.8%, -5.4%, -1.8%, -6.0%, -4.2%, -2.5% on 𝔾2, 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 tasks.

The selection of prompts may have a huge impact on the model performance (Gao, Fisch, & Chen, 2021). Therefore, we conduct experiments using different numbers of prompt tokens on domain 𝔻𝕀 → 𝔸𝕊 to further investigate the influence of soft prompts. The results are shown in Fig. 6, revealing that the length of the prompt token affects the performance of prompt tuning on domain 𝔻𝕀 → 𝔸𝕊. In our method, we set the prompt token length as 3 to achieve the best results for cross-domain ATE.

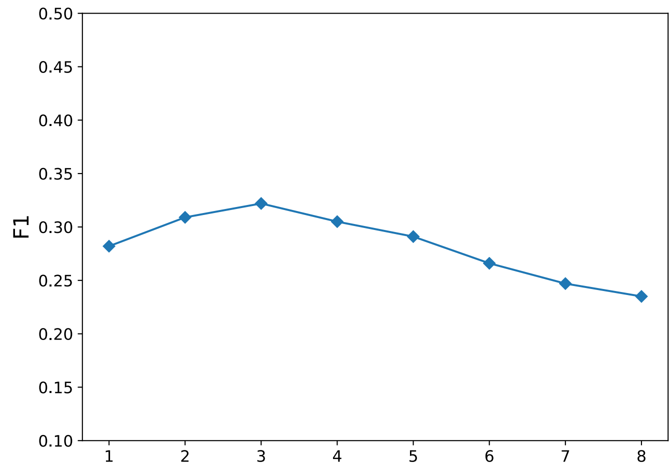

Fig. 6. Experimental results of different lengths of soft prompt tokens on 𝔻𝕀 → 𝔸𝕊.

5.7. Case study

To further demonstrate the effectiveness of the proposed model, we perform a case study on dataset 𝔾1. Table 9 presents the results of cross-domain ATE by AD-SAL, CrossBERT-UDA, and our method. In Example 1 "*Straight-forward, no surprises, very decent Japanese food.*", neither CrossBERT-UDA nor AD-SAL can 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*. In Examples 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 parts of these aspect terms. In Examples 4 and 5, baseline models are able to correctly predict part of aspect terms while our method can identify all of them correctly. The case study demonstrates that our method can accurately detect not only multiple aspect terms but also aspect terms with multiple words.

6. Conclusion and future work

In this paper, we 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 to narrow the domain gap.

Our experiments across two groups of datasets spanning a range of domains demonstrate the effectiveness of our approach over the existing models for cross-domain aspect term extraction. Our model can achieve outstanding performance when the data size is quite small, indicating that our method is effective to tackle ATE task on both small-scale and large-scale datasets. 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) (Shi et al., 2023). To explore the effectiveness of soft prompt in knowledge transfer, we only conduct experiments of ATE task. In the future, we plan to expand our model to complete more aspect-based sentiment analysis tasks (e.g., opinion term extraction, aspect-opinion term pair extraction, sentiment triplet extraction, etc.).

Table 6
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
EXAL	0.213	0.188	0.167	0.189	0.163	0.184
SA-EXAL	0.225	0.211	0.236	0.205	0.190	0.207
CrossBERT-UDA	0.244	0.201	0.195	0.213	0.160	0.196
SynBridge	0.272	0.281	0.255	0.279	0.193	0.202
SymBridge	0.303	0.297	0.260	0.286	0.205	0.211
Ours	**0.322**	**0.312**	**0.279**	**0.313**	**0.213**	**0.251**

Table 7
Ablation study over cross-domain ATE on $\mathbb{G}1$. w/o indicates without.

Model	-w/o Syntax	-w/o Prompts	only T5	Ours
$\mathbb{R} \rightarrow \mathbb{L}$	0.581	0.573	0.560	**0.593**
$\mathbb{S} \rightarrow \mathbb{L}$	0.477	0.464	0.450	**0.480**
$\mathbb{D} \rightarrow \mathbb{L}$	0.511	0.508	0.493	**0.527**
$\mathbb{L} \rightarrow \mathbb{R}$	0.641	0.632	0.627	**0.655**
$\mathbb{S} \rightarrow \mathbb{R}$	0.596	0.583	0.579	**0.612**
$\mathbb{D} \rightarrow \mathbb{R}$	0.620	0.607	0.598	**0.630**
$\mathbb{R} \rightarrow \mathbb{D}$	0.549	0.533	0.528	**0.563**
$\mathbb{L} \rightarrow \mathbb{D}$	0.428	0.419	0.411	**0.434**
$\mathbb{S} \rightarrow \mathbb{D}$	0.570	0.565	0.551	**0.586**
$\mathbb{R} \rightarrow \mathbb{S}$	0.511	0.508	0.504	**0.528**
$\mathbb{L} \rightarrow \mathbb{S}$	0.539	0.526	0.510	**0.555**
$\mathbb{D} \rightarrow \mathbb{S}$	0.565	0.522	0.515	**0.570**

However, it is challenging research work to design a unified framework to complete all ABSA sub-tasks in different domains due to the complex relations existing among aspect, opinion, and sentiment polarity. Simultaneously, the performance of the proposed method can be further improved on small-scale datasets to enhance the robustness.

CRediT authorship contribution statement

Jingli Shi: Writing – original draft, Conceptualization, Data curation, Methodology, Software. **Weihua Li:** Conceptualization, Project administration, Writing – review & editing, Supervision. **Quan Bai:** Formal analysis, Project administration, Writing – review, Supervision. **Yi Yang:** Writing – review, Formal analysis, Resources, Validation. **Jianhua Jiang:** Writing – review, Project administration, Resources.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors would like to acknowledge the financial support from Callaghan Innovation (CSITR1902, 2020), New Zealand, without which this research would not have been possible. We are grateful for their contributions to the advancement of science and technology in New Zealand. The authors would also like to thank CAITO.ai for their invaluable partnership and their contributions to the project.

References

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*, 1–167.
Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval 2016)* (pp. 19–30).
Zhong, Q., Ding, L., Liu, J., Du, B., Jin, H., & Tao, D. (2023). Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 1–14.
Liu, J., Zhong, Q., Ding, L., Jin, H., Du, B., & Tao, D. (2023). Unified instance and knowledge alignment pretraining for aspect-based sentiment analysis. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, *31*, 2629–2642.
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).
Akhtar, M. S., Garg, T., & Ekbal, A. (2020). Multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing*, *398*, 247–256.
Devlin, J., Chang, M.-W., Lee, K., & Toutanova Bert, K. (2019). 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).
Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer Bart, L. (2020). 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).
Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*, 1–67.
Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. Improving language understanding by generative pre-training (2018).
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*, 9.
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.
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).
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).
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).
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).
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*, Article 108668.
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).
Chen, X., Zhang, N., Xie, X., Deng, S., Yao, Y., Tan, C., Huang, F., Si, L., & 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).
Li, C., Gao, F., Bu, J., Xu, L., Chen, X., Gu, Y., Shao, Z., Zheng, Q., Zhang, N., Wang, Y., et al. (2021). Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis, arXiv preprint. arXiv:2109.08306.
Chen, X., Zhang, N., Li, L., Xie, X., Deng, S., Tan, C., Huang, F., Si, L., & Chen Lightner, H. (2021). A lightweight generative framework with prompt-guided attention for low-resource ner. arXiv preprint, arXiv:2109.00720.
Gao, T., Fang, J., Liu, H., Liu, Z., Liu, C., Liu, P., Bao, Y., & 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).
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* (pp. 1–8). IEEE.

Table 8
Ablation study over cross-domain ATE on $\mathbb{G}2$. w/o indicates without.

Model	$\mathbb{DI} \to \mathbb{AS}$	$\mathbb{E} \to \mathbb{AS}$	$\mathbb{DI} \to \mathbb{E}$	$\mathbb{AS} \to \mathbb{E}$	$\mathbb{E} \to \mathbb{DI}$	$\mathbb{AS} \to \mathbb{DI}$
-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**

Table 9
The prediction results of aspect term on domain $\mathbb{R}$ by AD-SAL, CrossBERT-UDA, and our 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]}

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* (pp. 2324–2335).

Ziser, Y., & Reichart, R. (2017). Neural structural correspondence learning for domain adaptation. In *Proceedings of the 21st conference on computational natural language learning (CoNLL 2017)* (pp. 400–410).

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*.

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).

Chernyshevich, M., & Belarus, I. (2014). Cross-domain extraction of product features using conditional random fields. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 14)* (pp. 309–313).

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* (pp. 2171–2181).

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

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.

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* (pp. 4590–4600).

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* (pp. 4129–4138).

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).

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* (pp. 7035–7045).

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

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). Springer.

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).

Howard, P., Ma, A., Lal, V., Simoes, A. P., Korat, D., Pereg, O., Wasserblat, M., & 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).

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).

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.

Vu, T., Lester, B., Constant, N., Al-Rfou, R., & Cer, D. (2022). Spot: Better frozen model adaptation through soft prompt transfer. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 5039–5059).

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao Panda, D. (2022). Prompt transfer meets knowledge distillation for efficient model adaptation. arXiv preprint, arXiv:2208.10160.

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* (pp. 2438–2447).

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.

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* (pp. 1241–1251).

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

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

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

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

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* (pp. 317–327).

Ratinov, L., & Roth, D. (2009). Design challenges and misconceptions in named entity recognition. In *Proceedings of the thirteenth conference on computational natural language learning (CoNLL-2009)* (pp. 147–155).

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).

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).

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).

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).

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*, 1469–1499.

Kingma, D. P., & Ba Adam, J. (2015). A method for stochastic optimization. In *Proceedings of the 3rd international conference on learning representations*.

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

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).

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).

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).

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).

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.

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

Alshuwaier, F., Areshey, A., & Poon, J. (2022). Applications and enhancement of document-based sentiment analysis in deep learning methods: Systematic literature review. *Intelligent Systems with Applications*, *15*, Article 200090.

Rizou, S., Theofilatos, A., Paflioti, A., Pissari, E., Varlamis, I., Sarigiannidis, G., & Chatzisavvas, K. C. (2023). Efficient intent classification and entity recognition for university administrative services employing deep learning models. *Intelligent Systems with Applications*, Article 200247.

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* (pp. 3816–3830).

Shi, J., Li, W., Bai, Q., Yang, Y., & Jiang, J. (2023). Syntax-enhanced aspect-based sentiment analysis with multi-layer attention. *Neurocomputing*, Article 126730.