Aspect-Opinion Correlation Aware and Knowledge-Expansion Few Shot Cross-Domain Sentiment Classification

Haopeng Ren, Yi Cai, *Member, IEEE*, Yushi Zeng, Jinghui Ye, Ho-fung Leung, *Senior Member, IEEE*, and Qing Li, *Senior Member, IEEE*

Abstract—Cross-domain sentiment analysis has recently attracted significant attention, which can effectively alleviate the problem of lacking large-scale labeled data for deep neural network based methods. However, most of the existing cross-domain sentiment classification models neglect the domain-specific features, which limits their performance especially when the domain discrepancy becomes larger. Meanwhile, the relations between the aspect and opinion terms cannot be effectively modeled and thus the sentiment transfer error problem is suffered in the existing unsupervised domain-adaptation methods. To address these two issues, we propose an aspect-opinion correlation aware and knowledge-expansion few shot cross-domain sentiment classification model. Sentiment classification can be effectively conducted with only a few support instances of the target domain. Extensive experiments are conducted and the experimental results show the effectiveness of our proposed model.

Index Terms—Cross domain sentiment analysis, knowledge graph, few-shot learning

1 Introduction

Sentiment analysis (SA) is a fundamental task in natural language processing (NLP), aiming to automatically assign the sentiment polarities (i.e., positive, neutral, or negative) to the user-generated text data like the restaurant reviews. For example, the sentence review "*The food in this restaurant is delicious*" expresses the positive sentiment. Currently, the deep neural network based SA models [1], [2] are widely-used and achieve remarkable performance, nevertheless suffer from the problem of lacking large-scale labeled data. It is usually time-consuming and human-intensive to make annotations in many applications. To alleviate this problem, the task of cross-domain sentiment classification [3], [4], [5] recently attracts considerable attention, which transfers the knowledge learned from the label-rich source domain to the label-scarce target domain.

The main challenge in cross-domain sentiment classification is the discrepancy between the source and target domain (e.g., the different expressions of users' emotions across domains). Facing this challenge, one group of recent domain-adaptation methods (e.g., the adversarial learning [7], [8], [9], [10]) focus on learning the *domain-invariant* (domain-shared) features (e.g., the opinion terms "*terrible*", "*great*" and "*fast*" which are shared in both the source and target domains, as shown in Fig. 1). They are often based on a key assumption that the domain-invariant features also share the same sentiment polarities in both the source and target domains. Nevertheless, it is often violated in many realistic scenarios and causes the sentiment transfer error problem. For example shown in Fig. 1, the opinion term "*fast*" expresses the negative sentiment when describing the aspect "*battery*" in the Electronic domain, while expresses the positive sentiment for the aspect term "*pan*" in the Kitchen domain. The sentiment of the domain-invariant feature "*fast*" from the source domain (i.e., Electronic domain) is wrongly transferred as negative polarity into the target domain (i.e., the Kitchen domain). Therefore, the sentiment polarities of the domain-invariant features not only rely on the domains they are in but also depend on the aspects they describe. Inspired by the success of applying syntactic information in the aspect-opinion pairs extraction task [11], [12], we introduce the syntactic knowledge structure to capture the relational features between the aspect and opinion terms for the cross-domain learning, aiming to solve the sentiment transfer error problem. Specifically, the syntactic knowledge

- *Haopeng Ren, Yushi Zeng, and Jinghui Ye are with the School of Software Engineering, South China University of Technology, Guangzhou 510650, China, and also with the Key Laboratory of Big Data and Intelligent Robot, SCUT, Guangzhou 510335, China. E-mail: se_renhp@mail.scut.edu.cn, yushi_znn@foxmail.com, jhuiye@qq.com.*
- *Yi Cai is with the School of Software Engineering, South China University of Technology, Guangzhou 510650, China, and also with the Key Laboratory of Big Data and Intelligent Robot, SCUT and the Pazhou Lab, Guangzhou 510335, China. E-mail: ycai@scut.edu.cn.*
- *Ho-fung Leung is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong. E-mail: lhf@cuhk.edu.hk.*
- *Qing Li is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. E-mail: csqli@comp.polyu.edu.hk.*

Manuscript received 10 January 2022; revised 20 July 2022; accepted 4 September 2022. Date of publication 9 September 2022; date of current version 15 November 2022.
This work was supported in part by the National Natural Science Foundation of China under Grant 62076100, in part by the Fundamental Research Funds for the Central Universities, SCUT under Grant x2rjD2220050, in part by the Science and Technology Planning Project of Guangdong Province under Grant 2020B0101100002, in part by the Hong Kong Research Council under Grants PolyU 11204919 and C1031-18G, and in part by the Internal Research from the Hong Kong Polytechnic University under Grant 1.9B0V.
(Corresponding author: Yi Cai.)
Recommended for acceptance by E. Cambria.
Digital Object Identifier no. 10.1109/TAFFC.2022.3205358

1949-3045 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

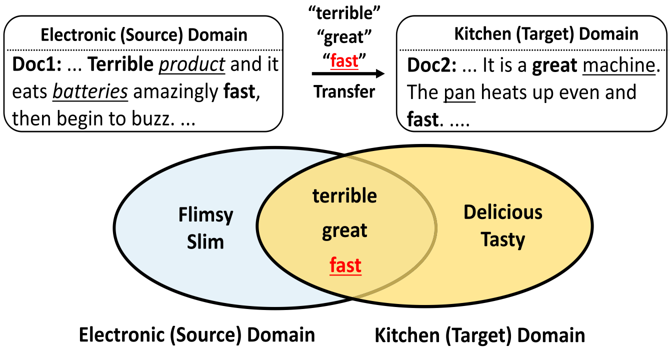


Fig. 1. An example of the sentiment transfer error in current cross-domain sentiment classification methods which focus on the domain-invariant features learning.

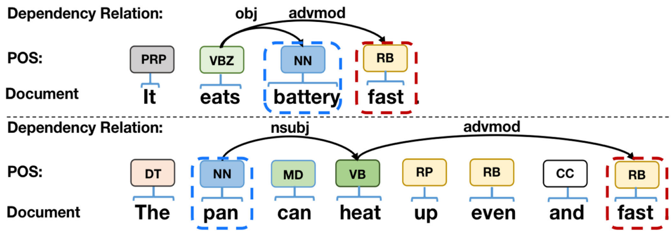


Fig. 2. Illustration of the aspect-opinion relationships in two sentences with (1) syntactic dependency structures and (2) the corresponding part-of-speech tags.

structure can provide key clues for supporting the underlying reasoning. As shown in Fig. 2, the ones with syntactic structures "*obj - advmod*" can bring the evident characteristic clues for facilitating the inference of the aspect-opinion pairs.

In addition, another group of methods [13], [14] not only focuses on capturing the domain-invariant features but also the domain-specific features which are the strong indicators for the sentiment analysis in the target domain (e.g., the opinion terms '*delicious*' and '*tasty*' in the target (Kitchen) domain, as shown in Fig. 1). To learn the domain-specific features of target domain, two kinds of solutions, i.e., *fine-tuning* [13] and *semi-supervised learning*[14] are designed by giving a small amount (e.g., 50) of target-domain training labeled data. However, these methods are still based on the deep neural networks with large-scale parameters and suffer from the lack of large-scale labeled data for the target domain, which are prone to overfitting [15]. In contrast, it is intuitive that humans can learn new knowledge after being taught just a few labeled instances [16]. Based on this intuition, the few-shot learning technique has shown effectiveness in various tasks (e.g., relation classification [17], opinion summarization [18] and image classification [16]). It encourages the model to learn the fast-learning ability from previous experience and quickly generalize to the new scenarios with a few support instances. The transferable knowledge can be extracted and propagated from a collection of meta-tasks, which enables the model to prevent the overfitting problem [19]. Motivated by this, our work in this paper explores the task of few-shot cross-domain sentiment classification, in which the cross-domain SA system can not only extract the domain-invariant features but also obtain the domain-specific features by giving only a few (e.g., 1 or 5) support instances meanwhile without encountering the overfitting problem.

Though many studies on few-shot learning obtain promising results, they still suffer from one major challenge when directly adapted to the cross-domain sentiment classification task i.e. the scarce domain-specific features contained in the few support instances from the target domain. According to our observation, the relational knowledge graph (e.g., *ConceptNet* [9], [20]) has the rich domain commonsense knowledge which benefits the domain-specific semantic understanding and becomes a potential solution to solve the problem of scarce domain-specific features. Specifically, as shown in Fig. 3, the few-shot SA model conducts the cross-domain sentiment classification based on only a few (i.e., 2) support instances which are respectively provided for the positive and negative classes in the target domain. First, the relationships between the aspect and opinion terms are built based on the dependency relations. For instance, the aspect-opinion pair "*soup* ↔ *delicious*" are connected with the dependency relation "*nsubj*", as shown in Fig. 3. Then, through the relational knowledge graph, the rich domain-specific background knowledge can be linked based on the given few support aspect-opinion pairs, which benefits the semantic understanding of aspect and opinion terms in the target domain. As we can observe, the terms with similar semantics usually share the relational knowledge structures. The sentiment features can be transferred based on the shared relational knowledge structures. For example shown in Fig. 3, the aspect terms "*soup*" and "*pizza*" share most of the neighborhood nodes (e.g., the terms "*meat*", "*restaurant*" and so on) in the relational knowledge graph. Based on the bridge of the shared relational knowledge structure, the sentiment features can be transferred from the few support instances to the query instances (e.g., both the aspect-opinion pair "*soup*↔*delicious*" in sample$_{(1)}$ and "*pizza*↔*delicious*" in query$_{(1)}$ share the positive sentiment polarity). In this way, with the help of the external relational knowledge graph, the domain-specific sentiment features can be enriched based on only a few support instances.

In this paper, we propose an aspect-opinion correlation aware and knowledge-expansion few-shot cross-domain sentiment classification model (AKFSM). As shown in Fig. 4, the framework of our proposed model consists of two phases. For the first phase named *Aspect-Opinion Correlation Aware Graph Feature Learning*, two self-supervised tasks (i.e., the relation classification task and the sentiment alignment task) are designed to pre-train the graph convolution network (GCN) encoder, aiming to capture the relational knowledge features. Then, the second phase, named *feature-fusion based few-shot learning*, conducts the sentiment classification with a few (e.g., 1 or 5) support instances by infusing the relational knowledge features and the semantic features from the domain-adapted BERT.

Our contributions are summarized as follows:

- We explore a problem of cross-domain sentiment classification in the few-shot scenario. For this problem, we propose an aspect-opinion correlation aware and knowledge-expansion few-shot cross-domain sentiment classification model. To the best of our knowledge, our work is the first study focusing on *few-shot cross-domain sentiment classification*.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

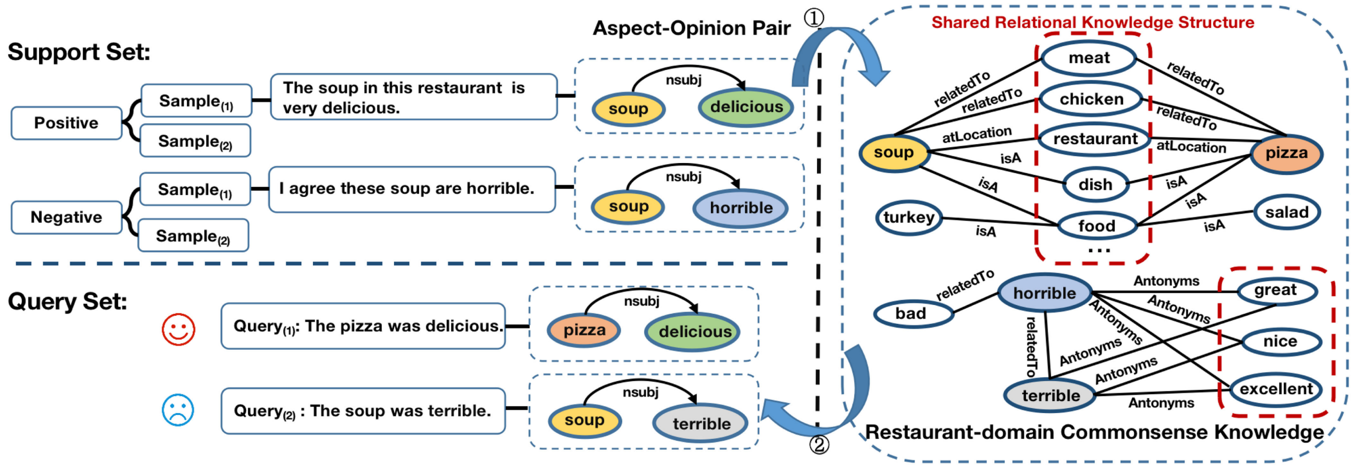


Fig. 3. A 2-shot setting example for few-shot cross-domain sentiment classification with the external commonsense knowledge graph. The "*nsubj*" denotes the dependency relation with the Standford CoreNLP libraries [6].

- We design an aspect-opinion correction aware graph feature learning method with two self-supervised pre-trained tasks to solve the sentiment transfer error problem suffered in existing unsupervised domain-adaptation methods.
- We propose a knowledge-expansion few-shot cross-domain sentiment classification model. It can effectively expand the domain-specific knowledge with only a few support instances meanwhile do not suffer from the overfitting problem.
- Extensive experiments and visualization analysis are conducted to evaluate the effectiveness of our proposed model in the few-shot cross-domain sentiment classification scenario.

2 Related Work

Sentiment classification aims to classify the sentiment polarities (e.g., positive, negative and neutral sentiment) of the given text, which is a basic task in NLP. Deep neural network based methods [1], [21], [22], [23], [24], [25], [26], [27], [28] highly rely on the large-scale labeled training data in a specific domain, but the data labeling process is often labor-intensive and time-consuming. To solve this problem, the cross-domain sentiment classification task is proposed and attracts much researchers' attention. It aims to transfer the knowledge from the label-rich source domain to the label-scarce target domain. Existing transfer-based methods for the cross-domain sentiment classification task can be summarized into three categories: pivot extraction based methods, non-pivot extraction based methods and deep transfer learning based methods.

First, the pivot extraction based method aims to capture the pivot terms and treat them as transferable features across domains. The pivot-selection strategies can be classified into two groups: *statistics-based* and *label-based* [29]. Specifically, the statistics-based methods extract the domain-shared and sentiment-indicative features based on the statistics information (e.g., the term frequencies [30] and pointwise mutual information (PMI) [31]) between the source and target domains. These methods are mostly the heuristic methods which lack the semantic understanding of the text and require the manual selection. Then, the label-based methods aim to select the pivots from the sentiment features by a supervised classifier that is trained with the labeled data of the source domain. For instance, the instance weighting method [32] is proposed to obtain the pivot feature representation by bridging the distribution of the source and target domains. The deep network based methods, such as the marginalized stacked denoising autoencoder (MSDA) [33], the HATN [34], PBLM [35] and TPT [36] are designed to extract the domain-shared sentiment features.

Second, the non-pivot extraction based method aims to capture the emotion terms which are usually the indicators for the sentiment classification in the target domain. Specifically, the HATN [34] is proposed to introduce the non-pivots by treating the pivots as the bridge. Moreover, the prior knowledge (e.g., sentiment dictionary) is added into neural networks [37], [38], [39] to capture the non-pivots. Li et al. 2020 [36] propose a Transferable Pivot Transformer (TPT) which detects both the pivot words and non-pivot words by modeling the relationships between the pivot and non-pivot words. Nevertheless, the accuracy of detecting the non-pivot words relies on the performance of pivot word extraction. Thus, these methods suffer from the error propagation problem. Moreover, as the discrepancy across domains increases, the pivot words become scarce and then the non-pivot words are difficult to be extracted.

The third category method focusing on the cross-domain sentiment classification task is based on the deep transfer learning technique. Recently, the adversarial training based methods [7], [40], [41], [42], [43] aim to automatically obtain the domain-invariant (domain-shared) features by applying the attention mechanism and adversarial training strategy. With the success of the pre-training language model (e.g., BERT [44]), the domain-aware BERT and adversarial training strategy are combined to automatically learn the domain-invariant features across domains. Moreover, the domain-adversarial framework KinGDOM [9] is proposed to learn the domain-invariant features by introducing the relational knowledge graph (i.e., ConceptNet [20]). Based on the success of graph convolutional network techniques [26], [27], [28], [45], [46], [47], [48], the graph-structure domain

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

knowledge features can be captured, which benefit the cross-domain learning. The semantics of the review documents can be enriched[20] by providing both the domain-invariant and domain-specific background concepts. Then, the domain-invariant features are captured by utilizing the adversarial training strategy. However, KinGDOM mainly learns the domain-invariant features and ignores the domain-specific features. In addition, the relationships between the aspect and opinion concepts are not built in the utilized external knowledge graph, which also causes the sentiment transfer error problem in the cross-domain sentiment classification, as shown in Fig. 1. In this paper, our work also utilizes the external knowledge graph and focuses on solving the above two problems (i.e., the domain-specific features ignoring problem and the sentiment transfer error problem).

In summary, though the existing methods based on deep transfer learning recently achieve better performance, they mainly focus on extracting the domain-invariant features for the target domain while neglecting the domain-specific features. Current two solutions for introducing the domain-specific features (i.e., the fine-tuning [13] and the semi-supervised method [14]) are prone to suffer from the overfitting problem when training with only a few labeled data of the target domain. Currently, the few-shot learning technique achieves success in many NLP tasks [15], [17], [18], [49], [50], [51], [52] and is a effective solution to avoid the overfitting problem. Inspired by the success of the adversarial training strategy [53], several cross-domain few-shot learning baselines [17] are designed. Moreover, two related works [54], [55] focusing on the cross-domain few-shot text classification task are proposed. They are also adapted to the cross-domain few-shot sentiment classification task and conduct the comparative experiments with our proposed model.

In our paper, we propose a few-shot learning based cross-domain sentiment classification model to effectively address the problem of ignoring domain-specific features. Utilizing the external knowledge graph, the rich domain-specific features can be expanded with only a few support instances. Furthermore, an aspect-opinion correlation aware graph learning method is designed to solve the sentiment transfer error problem suffered in the existing methods based on the deep transfer learning.

3 Model

To solve the sentiment transfer error problem and the domain-specific features ignoring problem suffered in existing methods, we propose a knowledge-expansion few-shot learning model for the cross-domain sentiment classification task. As shown in Fig. 4, two phases are contained in our proposed model and they are sequential relationships. First, the graph feature encoder (i.e., GCN Autoencoder) is designed and pretrained with the unlabeled data from both the source and target domains in Phase 1, aiming to capture the graph structure features of the expanded commonsense knowledge graph. In Phase 2, both the graph structure features (extracted by the GCN encoder in Phase 1) and the text semantic features (obtained by the domain-adapted encoder in Phase 2) are fused to conduct the sentiment classification with a few (1 or 5) support instances.

In the first phase, named *aspect-opinion correlation aware graph feature learning*, an aspect-opinion correlation aware knowledge graph is constructed based on the *ConceptNet* [20]. Then, two self-supervised tasks (i.e., relation classification task and sentiment alignment task) are conducted with the knowledge graph to pre-train the GCN auto-encoder, aiming to learn the relational knowledge structure of the aspect-opinion pairs. Then, the second phase, named *knowledge-expansion based few-shot learning*, aims to expand the domain-specific features by the commonsense knowledge graph based on the GCN encoder (which is obtained in Phase 1) given only a few support instances. Moreover, not all relational knowledge is beneficial for sentiment classification of the query instances. Little noise in the support set may cause a huge deviation of the feature representation in the few-shot learning scenario [51]. Motivated by this, the shared-knowledge aware attention is designed to select the transferable knowledge triplets in cross-domain learning. Finally, both the graph features from the GCN autoencoder and the text semantic features from the domain-adapted BERT are fused for the support and query instances. Then, sentiment classification is conducted based on the prototypical network with only a few support instances.

3.1 Problem Definitions

The task of few-shot cross-domain sentiment classification can be defined as follows. Given a support set S with two sentiment polarity categories $C \in \{Positive, Negative\}$, a model classifies the query instance q into the most possible sentiment polarity $c_i \in C$. S is defined as follows:

$$S = \left\{ \begin{array}{c} \{(x_1^1, c_1), (x_1^2, c_1), \ldots, (x_1^{n_1}, c_1)\}, \\ \{(x_2^1, c_2), (x_2^2, c_2), \ldots, (x_2^{n_2}, c_2)\} \end{array} \right\} \tag{1}$$

where $c_1, c_2 \in C$; and (x_i^j, c_i) denotes the support instance x_i^j belongs to the c_i sentiment polarity; x_i^j is denoted as a word sequence $\{w_1, w_2, \ldots, w_L\}$ and L is the length of the word sequence; n_i is the number of support instances belonging to the sentiment polarity c_i and is generally quite small in few-shot scenario. The few-shot cross-domain sentiment classification model is trained in the source domain and is tested in the target domain. The transferable information can be extracted and propagated from the source domain to the target domain by conducting a collection of meta-tasks [19]. Specifically, both the support set S and the query set Q are constructed in each meta task. The sentiment classification of the query instance $q \in Q$ is conducted based on the support set.

3.2 Phase 1: Aspect-Opinion Correlation Aware Graph Feature Learning

As shown in Fig. 1, the sentiment polarities of the domain-invariant features depend on both the domains they are in and the aspects they describe. Based on this observation, we design an aspect-opinion correlation aware graph learning method to capture the relational features between the aspect and opinion terms, aiming to solve the sentiment transfer error problem. Then, an aspect-opinion correlation aware sentiment knowledge graph is constructed with the large-scale unlabeled data in both source and target domain

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

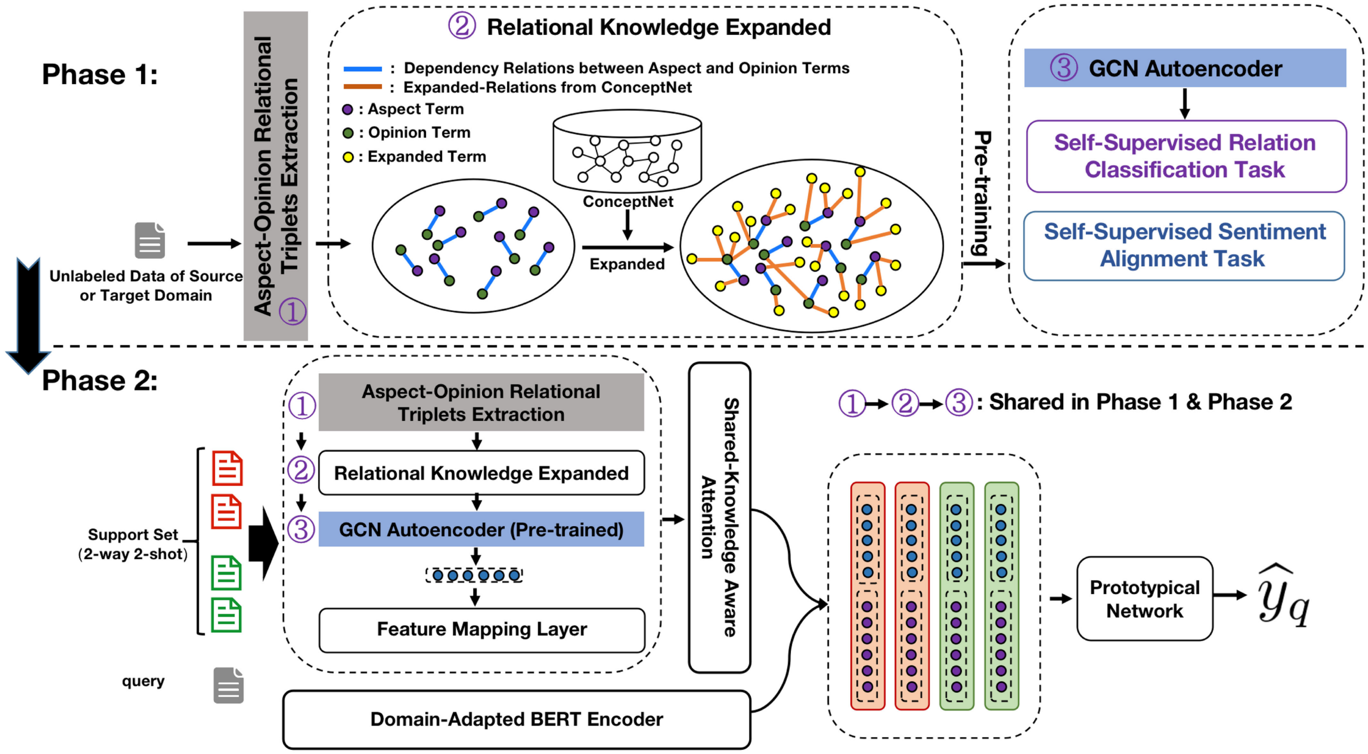


Fig. 4. The framework of our proposed knowledge-expansion few-shot cross-domain sentiment classification model.

utilizing the *ConceptNet*. The relationship between the aspect and opinion terms is built based on the syntactic information. Then, two self-supervised tasks are designed to pre-train the GCN-encoder, which aims to learn the relational knowledge features for the aspect and opinion terms.

3.2.1 Knowledge Graph Construction

The relational knowledge graph is constructed based on the large-scale of unlabeled documents respectively for the source and target domain. Two steps are contained to construct the knowledge graph. The first step is to extract the aspect-opinion relational triplets based on the syntactic knowledge structures of each unlabeled document. Specifically, for each unlabeled document, we use the Standard CoreNLP libraries [6] to recognize the part-of-speech of each token and the dependency relations among the tokens. The tokens which belong to *nouns*, *adjectives* or *adverbs* are treated as the seed nodes and are connected together by the dependency relations (i.e., the *"nsubj"*, *"advmod"*, *"amod"*, *"obj"* and *"conj"* dependency relations). Thus, the relationships between the aspect and opinion terms can be extracted based on the syntactic knowledge structure. For example shown in Fig. 3, the aspect-opinion relational triplets (e.g., *"soup* $\xrightarrow{nsubj}$ *delicious"* in Sample$_{(1)}$) are obtained, where the *"soup"* and *"delicious"* can be respectively seen as the aspect and opinion terms.

Second, the *ConceptNet* knowledge graph [20] is utilized to link the commonsense knowledge triplets based on the aspect-opinion relational triplets. Motivated by the success in the node embedding techniques [56], the nodes with similar neighborhoods will have the close feature embeddings. For example shown in Fig. 3, the aspect terms *"soup"* and *"pizza"* share most of the neighborhoods, which indicates they have similar semantic features. Utilizing the bridge of the shared relational knowledge structure, the sentiment features can be transferred from the few support instances to the query instances in Phase 2. Specifically, based on the seed triple *"soup* $\xrightarrow{nsubj}$ *delicious"*, one hop of the commonsense knowledge triplets (e.g., *"soup"* $\xrightarrow{relatedTo}$ *"meat"*, *"soup"* $\xrightarrow{isA}$ *"food"* and so on) can be extracted from *ConceptNet*. In this way, with the unlabeled documents for the source and target domain, the knowledge graph $G = (V, \Psi, R)$ with nodes $v_i \in V$ and triplets $(v_i, r_{i,j}, v_j) \in \Psi$ can be constructed, where $r_{i,j} \in R$ denotes the relation between the node v_i and v_j.

3.2.2 Knowledge Graph Pre-Training

As described in Section 3.2.1, the relational knowledge graph is respectively constructed from the source and target domains. Currently, the Relational Graph Convolutional Network (GCN) encoder [24], [25], [57], [58] is proven to have the ability of accumulating relational evidence in multiple inference steps from the local neighborhood around a given node. Following Ghosal et al. 2020 [9], for each node (i.e., term) v_i in G, we utilize a two-layer GCN encoder (which are stacked one another) to learn its feature representation $\boldsymbol{g}_i$, as follows,

$$\boldsymbol{g_i} = \boldsymbol{h}_i^{(2)} = f(\boldsymbol{h}_i^{(1)}, 2); \boldsymbol{h}_i^{(1)} = f(\boldsymbol{v}_i, 1) \tag{2}$$

$$f(\boldsymbol{x}_i, l) = \sigma\left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{|N_i^r|} W_r^{(l)} \boldsymbol{x}_j + W_0^{(l)} \boldsymbol{x}_i\right) \tag{3}$$

where N_i^r denotes the neighbouring nodes of node v_i under the relation $r \in R$; σ denotes the activation function such as ReLU; $\boldsymbol{v}_i$ is the randomly initialized representation vector; $W_r^{(1/2)}$ and $W_0^{(1/2)}$ denotes the learnable parameters.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

To learn the relational knowledge features of the aspect or opinion terms, two self-supervised learning tasks (i.e., relation classification and sentiment alignment) are conducted to pre-train the GCN encoder model [57] in our constructed knowledge graph. Both of these two self-supervised learning tasks respectively adopt different strategies to constrain the GCN encoder to learn the graph feature representation, which benefits the model to transfer and expand the domain-specific sentiment features with a few support instances in Phase 2. Specifically, following Schlichtkrull et al. 2018 [57] and Ghosal et al. 2021 [9], the self-supervised relation classification task is adopted to force or push the GCN encoder to learn the graph structure features by accumulating the relational evidence. The nodes with similar neighborhoods (i.e., the relational knowledge structure) will have the close feature embeddings [56], which indicates they have similar semantics. Utilizing the bridge of the shared relational knowledge structure, the sentiment features can be transferred from the few support instances to the query instances in Phase 2 (e.g., both the aspect terms "*soup*" and "*pizza*" described by the same opinion term "*delicious*" can be referred to have the same sentiment polarity (i.e., Positive) as shown in Fig. 3) .

Moreover, the sentiment alignment task is designed to help the GCN encoder to capture the sentiment alignment features among aspect-opinion pairs by exploiting their co-occurrence features within documents. For example, the aspect-opinion pairs "*terrible↔product*" and "*battery↔fast*" are often co-occurring in the same document and share the same sentiment polarity. The sentiments of the aspect-opinion pairs can be derived through the sentiments of their contextual aspect-opinion pairs, which facilitates the transfer and expansion of sentiment features.

Specifically, for the self-supervised relation classification task, the model takes as input the triplets Ψ' (named positive triplets) from Ψ in G and the equal number of the negative triplets. Note that the negative triplets are created by randomly modifying either one of the nodes or the relations in the positive triplets. Both the positive and negative samples are merged into a set T and their labels are respectively denoted as 0 and 1 (i.e., $y \in \{0, 1\}$). Therefore, given the triplets $(v_i, r_{i,j}, v_j) \in T$, a binary classification task is conducted to train the GCN encoder with the cross-entropy loss:

$$\pounds_{\mathrm{G}} = -\frac{1}{|T|}\sum_{(v_i,r_{i,j},v_j,y)\in T}(y\log s(v_i, r_{i,j}, v_j)+ (1-y)\log(1-s(v_i, r_{i,j}, v_j))) \tag{4}$$

$$s(v_i, r_{i,j}, v_j) = \sigma(\boldsymbol{g}_i^T R_r \boldsymbol{g}_j) \tag{5}$$

where $s(v_i, r_{i,j}, v_j)$ denotes the DistMult factorization [59] score function; Each relation $r \in R$ is associated with a diagonal matrix $R_r \in \mathbb{R}^{d\times d}$.

For the self-supervised sentiment alignment learning task, the model takes as input the aspect-opinion relational triplets (named positive triplets) in one review and the equal number of negative triplets which are created by randomly selecting the aspect-opinion relational triplets from other reviews. Then both the positive and the negative triplets are merged into a set P and their labels are respectively denoted as 0 and 1 (i.e., $y \in 0, 1$). Thus, given the triplets $(v_i, r_{i,j}, v_j) \in P$, the model conducts a binary classification task to train the GCN encoder with the cross-entropy loss:

$$\pounds_{\mathrm{align}} = -\sum_{k=1}^{N}(\frac{1}{|P_k|}\sum_{(v_i,r_{i,j},v_j,y)\in P_k}(y\log s(v_i, r_{i,j}, v_j) + (1-y)\log(1-s(v_i, r_{i,j}, v_j))))$$

where N denotes the number of the unlabeled reviews in source or target domain; P_k denotes the aspect-opinion relational triplets set of the kth unlabeled reviews.

Finally, the GCN encoder can be optimized with the cross-entropy loss (i.e., $\pounds_{\mathrm{G}}$ and $\pounds_{\mathrm{align}}$) by simultaneously conducting the relation classification task and sentiment alignment task for each unlabeled document review, aiming to guide or force the GCN encoder to capture the graph structure features.

3.3 Phase 2: Knowledge-Expansion based Few-Shot Learning

3.3.1 Sentence Encoder

Given a review instance, two kinds of encoders (i.e., *Graph Feature Encoder* and *Domain-Adapted BERT Encoder*) are designed to obtain the instance representation. Specifically, the graph feature encoder (i.e., GCN Encoder) can be obtained from Phase 1, which aims to capture the graph structure features. Moreover, following Zhou et al. 2020 [10], the domain-adapted BERT encoder is adopted to capture the domain-invariant and sentiment-aware text semantic features. Finally, the feature representation of the given instance can be obtained by fusing the graph structure features (which are encoded by the GCN encoder) and text semantic features (which are encoded by the domain-adapted BERT encoder) with a reconstruction loss.

Graph Feature Encoder. The three modules (i.e., *Aspect-Opinion Relational Triplets Extraction*, *Relational Knowledge Expansion* and *GCN Autoencoder*) are shared in both phases, which aims to obtain the graph feature representation of the given instances. Similar to the Phase 1, given an instance x, the aspect-opinion relational triplets can be obtained based on the syntactic knowledge structures by the Standard CoreNLP libraries [6]. With the aspect-opinion relational triplets, an expanded commonsense knowledge sub-graph G_x for the instance x can be obtained by linking the external knowledge graph *ConceptNet* [20]. Then, each node in graph G_x can be encoded as a d-dimensional vector $\boldsymbol{v}_{node}$ by the pre-trained GCN encoder. The instance x can be represented as $\boldsymbol{x}_g$ by averaging all the nodes' representations, as follows:

$$\boldsymbol{x}_g = \frac{1}{M}\sum_{i=1}^{M}\mathbf{v}^i_{G_x}, \tag{7}$$

where M denotes the number of nodes in the graph G_x and $\mathbf{v}^i_{G_x} \in \mathbb{R}^{d_g}$ denotes the representation of the ith node in the graph G_x for the instance x.

Domain-Adapted BERT Encoder. The pre-trained language model BERT has shown to be effective in many NLP tasks, but is task-agnostic and little understanding of opinion text [8], [10]. To adapt the BERT into the specific domains (including both source and target domain), we conduct

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

several pre-training tasks at both token level and sentence level to obtain the domain-invariant sentiment knowledge by masking and prediction. Following Zhou et al. 2020 [10], three token-level (i.e., *sentiment-aware word prediction*, *word sentiment prediction* and *emotion pretiction*) and one sentence-level prediction tasks (i.e., *emoticon prediction*) are utilized to fine-tune the BERT encoder. Therefore, based on the pre-trained BERT encoder[1], we can obtain the semantic feature representation $\boldsymbol{x}_w \in \mathbb{R}^{d_w}$ for the instance x, where d_w denotes the dimension of the vector.

Feature Fusion. Each instance x can be encoded as the relational knowledge feature representation $\boldsymbol{x}_g$ and the semantic feature representation $\boldsymbol{x}_w$ respectively by the graph feature encoder and domain-adapted BERT encoder. According to our observation, the feature representations respectively encoded by the graph feature encoder and the domain-adapted BERT encoder are in different embedding spaces. The feature fusion using a simple concatenation or averaging operation will lead to bias in the distance metric of the prototypical network [9], [16]. To reduce the feature space discrepancy from the GCN encoder and the domain-adapted encoder, a feature mapping layer with a reconstruction loss (mean-squared error) is adopted as follows:

$$\boldsymbol{x}_g^m = W_g \boldsymbol{x}_g + b_g \tag{8}$$

$$\boldsymbol{x}_{recon} = W_{recon} \boldsymbol{x}_g^m + b_{recon}, \tag{9}$$

where $W_g \in \mathbb{R}^{d\times d}$, $W_{recon} \in \mathbb{R}^{d\times d}$, b_g and b_{recon} are the trainable parameters. The reconstruction loss is obtained by using the cosine similarity function, aiming to keep the features before and after the mapping operation unchanged.

$$\pounds_{\text{recon}} = \frac{\mathbf{x}_{\text{g}} \cdot \mathbf{x}_{\text{recon}}}{\|\mathbf{x}_{\text{g}}\|_2 \cdot \|\mathbf{x}_{\text{recon}}\|_2} \tag{10}$$

Therefore, the feature representation $\boldsymbol{x}$ of the given instance x can be obtained as follows:

$$\boldsymbol{x} = [\boldsymbol{x}_g^m; \boldsymbol{x}_w] \tag{11}$$

3.3.2 Shared-Knowledge Aware Attention

As shown in Fig. 3, not all the external knowledge nodes are equally important to the query instance. To capture the shared relational knowledge structures between the support and query instances, the shared-knowledge aware attention is designed. Specifically, given the support instance x and query instances q, the corresponding two sub-graphs G_x and G_q can be obtained by linking the *ConceptNet* (i.e., the two steps: aspect-opinion relational triplets extraction and relational knowledge expansion in Fig. 4). Then, these two sub-graphs are respectively encoded by the pre-trained GCN encoder as $\boldsymbol{V}_{G_x} \in \mathbb{R}^{N_x\times d_g}$ and $\boldsymbol{V}_{G_q} \in \mathbb{R}^{N_q\times d_g}$, where N_x and N_q are respectively the number of nodes in the sub-graph G_x and G_q; d_g denotes the dimensional size of the graph feature representation. The graph feature representation $\boldsymbol{x}_g$ in Eq. (8) is replaced by Eq. (12) as follows:

1. The pre-trained BERT checkpoint files can be downloaded in *https://github.com/12190143/SentiX.*

$$\boldsymbol{x}_g = \sum_{i=1}^{N_x} \boldsymbol{\alpha}_i (W_{att} \boldsymbol{v}_{G_x}^i + b_{att}), \tag{12}$$

where $W_{att} \in \mathbb{R}^{d_g\times d_g}$ and $b_{att} \in \mathbb{R}^{d_g}$ are the learnable parameters; α_i represents the important degree of the ith node in the sub-graph G_x and is calculated as follows:

$$\alpha_i = \frac{exp(e_j)}{\sum_{k=1}^{N_x} exp(e_k)} \tag{13}$$

$$e_i = sum\{\sigma((\boldsymbol{v}_{G_x}^i)^{\mathrm{T}} \times (W_{att} \boldsymbol{V}_{G_q}^{\mathrm{T}} + b_{att}))\}, \tag{14}$$

where $\sigma(\cdot)$ is an activation function $tanh$ and $sum\{\cdot\}$ denotes the sum of all elements of the vector.

3.3.3 Prototypical Network

The prototypical network [16] is utilized to conduct the few-shot sentiment classification. With the sentence encoder described in Section 3.3.1, both the support instances in support set S and query instance q are respectively encoded into low-dimensional vectors $\boldsymbol{x}_i^j$ and $\boldsymbol{x}_q$. For each sentiment polarity categories $c_i \in C$ with k support instances (i.e., 2-way k-shot setting), we can obtain the prototype $\boldsymbol{p}_i$ of category c_i as follows:

$$\boldsymbol{p}_i = \frac{1}{k} \sum_{j=1}^{k} \boldsymbol{x}_i^j \tag{15}$$

Finally, the probability of query instance q belonging to sentiment polarity category $c_i \in C$ can be measured as follows:

$$p_\phi(c_i|q) = \frac{exp(-d(\boldsymbol{p}_i, \boldsymbol{x}_q))}{\sum_{j=1}^{C} exp(-d(\boldsymbol{p}_j, \boldsymbol{x}_q))} \tag{16}$$

where ϕ denotes all the trainable parameters in sentence encoder; $d(.,.)$ is the Euclidean distance function for the two given vectors.

3.4 Loss Layer

Finally, the loss function of the whole architecture can be defined as follows:

$$\pounds = \pounds_{\text{recon}} + \pounds_{\text{softmax}} \tag{17}$$

where $\pounds_{\text{softmax}}$ denotes the cross-entropy loss of prototypical network in phase 2.

4 Experiment

4.1 Dataset and Experiment Setting

We conduct experiments on the Amazon-reviews benchmark dataset for cross-domain sentiment classification [60] with a few support instances of the target domain. The dataset ranges across four domains: Books (B), DVDs (D), Electronics (E), and Kitchen appliances (K). The reviews in all domains are associated with a rating denoting their sentiment polarity. Following Du et al. 2020[8], reviews with rating up to 3 stars are considered as negative sentiment and 4 or 5 stars as positive sentiment. Each domain has 2000 labeled reviews and approximately 4000 unlabeled reviews.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
Hyperparameters Settings

Parameter Name	Value
Maximum Instance Length L	200
Hidden Layer Dimension d	768
Batch Size	4
d_g	100
Initial Learning Rate	0.02
Query Set Size $\lvert Q\rvert$	5
keep dropout rate	0.0

In the training stage, our model is trained by conducting the meta-task in each episode to optimize the models' parameters [19]. For the N-way K-shot scenario, the support set S and query set Q are constructed for each meta-task. Note that N equals 2 (i.e., the positive and negative sentiment categories) in the cross-domain sentiment classification scenario. Specifically, we randomly select K instances for each sentiment category $c_i \in C$ to construct the support set S. Meanwhile, the query set Q is constructed by randomly selecting $|Q| = 5$ instances respectively from the positive and negative sentiment category, where $S \cap Q = \emptyset$. Our proposed models are optimized by conducting $iter = 20000$ meta-tasks. Similar to the task of cross-domain sentiment classification, the model is trained on the source domain and tested on the target domain. All the hyperparameters are shown in Table 1.

As shown in Tables 2, 3, and 4, we compare the performance of our proposed model with that of two categories of related works: 1) the few-shot learning baselines (i.e., *GNN* [49], *MetaNet* [50], *SNAIL* [61], *Proto-CNN* [16], *Proto-CNN* with adversarial training (Proto-CNN†) [17], *Proto_HATT* [51], *Proto-BERT* [17], *Proto-BERT* with adversarial training (Proto-BERT†) [17], *BERT-PAIR* [17], *MLADA* [54], *PtNet* [55]) and 2) current cross-domain sentiment classification models (i.e., *DANN* [40], *PBLM* [35], *HATN* [34], *ACAN* [43], *IATN* [42], *HATN-BERT* [34], *CoCMD* [14], *KinGDOM* [9], *BERT-DAAT* [8]) and SENTIX [10]).

4.2 Result Analysis

In our experiments, we compare the performance of our proposed model with 11 few-shot learning based baselines (i.e., *GNN*, *MetaNet*, *SNAIL*, *Proto-CNN*, *Proto_HATT* [51], *Proto-BERT*, *Proto-CNN* with adversarial training (Proto-CNN†), *Proto-BERT* with adversarial training (Proto-BERT†), *BERT-PAIR* [17]), *MLADA* [54]), *PtNet* [55] and 10 current cross-domain sentiment classification models (i.e., *DANN* [40], *PBLM* [35], *HATN* [34], *ACAN* [43], *IATN* [42], *HATN-BERT* [34], *CoCMD* [14], *KinGDOM* [9], *BERT-DAAT* [8]) and *SENTIX* [10] and 3 supervised learning based baselines (i.e., *CNN*, *LSTM* [1] and *BERT* [62]) which are trained with 1000 target domain labeled data, validated with 200 labeled data and tested with 800 labeled data. Furthermore, the ablation experiments and visualization analysis are conducted to evaluate the effectiveness of different modules in our proposed model. We analyze the experimental results from four perspectives as follows.

4.2.1 Comparison With Few-Shot Learning Baselines

Currently, the few-shot learning technique obtains success on several NLP tasks (e.g., text classification and relation classification). To evaluate the effectiveness of our proposed model, eight popular few-shot learning baselines designed for relation classification or text classification are adapted into the few-shot cross-domain sentiment classification scenario. As shown in Tables 2 and 3, two experimental settings (i.e., 1-shot and 5-shot) are considered in our experiments. As we can observe, our proposed model achieves higher accuracies with a large margin in all of the cross-domain experimental settings. Specifically, a few (e.g., 1 or 5) support instances are given for the few-shot sentiment classification task and only cover a small amount of domain-specific features. Different from the few-shot learning baselines, our proposed model can effectively enhance the support information with the given few support samples from the target domain, which effectively improves the performance in cross-domain learning.

As shown in Table 2 and Table 3, the performance of the three baselines *GNN*, *MetaNet* and *SNAIL* is worse than our models with a large margin. According to our observation, these three baselines even perform worse when the number of support instances increases in some cross-domain tasks (e.g., the *Kitchen→Book* cross-domain task for the model *SNAIL* shown in Tables 2 and 3). We analyze that not all the provided support features (or instances) can benefit identifying the sentiment polarities of the query instances. The support instances may contain irrelevant features and even noises when identifying the sentiment polarity of the query instances [51]. Motivated by this, the shared-knowledge aware attention module is designed to weight the expanded commonsense knowledge nodes, aiming to alleviate the effect of the irrelevant and noisy knowledge. In the experiments shown in Table 2 and 3, our proposed model achieves better performance as the number of support instances increases and can make full use of the few provided support instances.

Moreover, compared with the prototypical network based baselines (i.e., the model *Proto-CNN*, *Proto_HATT*, *Proto-BERT*), our proposed model (which is also based on the prototypical network) performs better with a large margin. It can further evaluate the effectiveness of the external commonsense knowledge learning module (i.e., *Phase 1: Aspect-Opinion Correlation Aware Graph Feature Learning*) designed in our model. With the help of the constructed external knowledge graph, rich support information can be effectively expanded and further improve the performance of few-shot learning methods. Compared with the BERT-based few-shot learning baselines (i.e., *Proto-BERT*, Proto-BERT†, *BERT-PAIR*), our proposed model with domain-adapted BERT encoder achieves around 2%-4% higher accuracies. The pre-trained language model BERT has a powerful ability in language understanding and is enabled to improve many NLP tasks. However, the BERT is task-agnostic and has no domain awareness. Motivated by this, the language model BERT is fine-tuning in the specific domain texts, aiming to force the model to obtain the domain-awareness ability.

To the best of our knowledge, we are the first to focus on the cross-domain few-shot sentiment classification task. Currently, many efforts have been devoted to the unsupervised domain adaptation on sentiment classification task [8], [9], [40], [43]. Among them, adversarial training [53] has been proved to be efficient in finding domain-invariant

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

TABLE 2
Average Accuracies (%) Comparison in 1-Shot Scenario With Current Few-Shot Learning Based Adapted Baselines are Conducted in the Amazon-Reviews Benchmark Dataset

Models	D → B	E → B	K → B	B → D	E → D	K → D	B → E	D → E	K → E	B → K	D → K	E → K	Avg
GNN	68.3	67.0	63.2	65.7	64.2	67.5	64.5	69.4	70.7	66.5	65.7	76.4	67.4
MetaNet	70.5	65.9	67.8	61.6	69.1	72.8	68.1	67.7	77.3	69.5	70.5	79.7	70.0
SNAIL	69.5	64.0	66.9	70.0	62.3	61.2	61.1	63.4	73.1	64.7	65.1	72.1	66.1
Proto-CNN	58.4	56.8	57.0	60.4	59.7	57.7	56.4	57.7	65.5	57.3	56.6	65.0	59.41
Proto_HATT	59.2	55.7	56.3	62.1	57.4	58.2	57.8	58.5	66.6	57.2	58.1	65.8	57.2
Proto-BERT	79.6	76.8	72.8	80.3	75.6	76.7	71.8	72.9	84.0	75.1	80.2	85.4	77.6
BERT-PAIR	72.9	70.3	61.2	78.5	70.6	72.6	78.2	81.0	81.1	72.0	79.8	73.4	74.3
Proto-CNN†	57.3	55.1	56.2	60.2	57.8	57.5	54.7	56.9	63.5	53.7	52.3	61.2	59.0
Proto-BERT†	83.5	77.9	77.2	82.0	74.1	77.1	76.9	81.6	85.2	80.4	81.1	86.9	80.3
MLADA	54.6	52.3	51.5	55.3	53.1	52.4	54.2	54.4	55.9	52.5	53.1	56.5	53.8
PNet	64.2	61.7	61.8	65.2	61.4	62.5	60.4	61.0	67.1	62.7	62.6	69.5	63.4
AKFSM	**88.8**	**89.1**	**89.3**	**88.9**	**88.5**	**87.7**	**90.7**	**90.3**	**91.9**	**92.6**	**94.2**	**94.7**	**90.6**

The results are the average accuracies of 10000 meta tasks.

TABLE 3
Average Accuracies (%) Comparison in 5-Shot Scenario With Current Few-Shot Learning Based Adapted Baselines are Conducted in the Amazon-Reviews Benchmark Dataset

Models	D → B	E → B	K → B	B → D	E → D	K → D	B → E	D → E	K → E	B → K	D → K	E → K	Avg
GNN	70.5	66.5	66.0	71.5	64.5	68.2	65.6	69.5	74.6	67.2	68.6	75.8	69.0
MetaNet	72.0	64.8	68.2	74.1	67.9	71.1	68.2	69.6	78.1	66.6	70.7	79.3	70.9
SNAIL	70.1	64.6	61.7	68.4	65.7	64.5	62.2	63.1	72.0	65.3	63.1	74.1	66.2
Proto-CNN	68.2	62.7	65.2	69.2	65.5	64.9	63.2	65.9	76.5	64.4	64.1	76.6	67.2
Proto_HATT	68.1	63.5	62.8	69.6	64.9	64.3	64.8	64.8	75.2	66.6	64.4	77.3	65.0
Proto-BERT	86.7	83.4	84.4	87.0	82.7	83.2	82.8	85.7	89.6	91.2	87.3	91.6	86.3
BERT-PAIR	81.0	81.0	68.2	86.6	78.9	80.8	84.0	82.9	87.1	75.8	85.2	89.0	81.7
Proto-CNN†	64.6	63.2	64.6	69.0	60.7	65.6	62.9	61.9	74.8	62.5	60.3	69.9	67.2
Proto-BERT†	84.2	81.4	82.4	85.8	82.4	84.2	84.8	83.2	86.4	82.2	82.6	91.2	84.2
MLADA	63.6	61.4	60.1	64.6	62.6	60.7	61.3	62.6	64.5	63.1	63.5	67.7	63.0
PNet	75.9	70.9	71.7	76.7	70.9	72.5	70.4	71.7	78.8	73.9	72.6	80.2	73.9
AKFSM	**92.5**	**91.4**	**91.2**	**92.3**	**91.4**	**91.5**	**93.7**	**93.9**	**94.7**	**96.7**	**96.2**	**96.5**	**93.5**

The results are the average accuracies of 10000 meta tasks.

TABLE 4
Average Accuracies (%) Comparison With Current Cross-Domain Sentiment Classification Models on the Amazon-Reviews Benchmark Dataset

S → T	D → B	E → B	K → B	B → D	E → D	K → D	B → E	D → E	K → E	B → K	D → K	E → K	Avg
DANN	81.7	78.6	79.3	82.3	79.7	80.5	77.6	79.7	86.7	76.1	77.4	84.0	80.3
PBLM	82.5	71.4	74.2	84.2	75.0	79.8	77.6	79.6	87.1	82.5	83.2	87.8	80.4
HATN	86.3	81.0	83.3	86.1	84.0	84.5	85.7	85.6	87.0	85.2	86.2	87.9	85.2
ACAN	82.4	79.8	80.8	83.5	81.8	82.1	81.2	82.8	86.6	83.1	78.6	83.4	82.2
IATN	87.0	81.8	84.7	86.8	84.1	84.1	86.5	86.9	87.6	85.9	85.8	88.7	85.8
HATN-BERT	89.8	87.1	87.9	89.4	88.8	87.8	87.2	87.0	90.3	89.4	87.6	92.0	88.7
CoCMD	81.8	76.9	77.2	83.1	78.3	79.6	83.0	83.4	87.2	85.3	85.5	87.3	82.4
KinGDOM	82.7	78.4	80.0	85.0	80.3	82.3	83.9	83.9	88.6	86.6	87.1	89.4	84.0
BERT-DAAT	90.9	88.9	88.0	89.7	90.1	88.8	89.6	89.3	91.7	90.8	90.5	93.2	90.1
SENTIX	91.2	90.4	89.6	91.3	91.2	89.9	93.3	93.6	93.6	96.2	96.0	96.2	92.7
AKFSM	**92.5**	**91.4**	**91.2**	**92.3**	**91.4**	**91.5**	**93.7**	**93.9**	**94.7**	**96.7**	**96.2**	**96.5**	**93.5**

Five support instances of target domain are given in our proposed model.

features and achieved remarkable performance in the cross-domain setting. Motivated by this, current few-shot learning baselines (e.g., Proto-CNN† and Proto-BERT† [17]) adopt the adversarial training strategy to solve the cross-domain few-shot classification problem. Compared with the adapted few-shot learning methods, our proposed models achieve higher accuracies with a large margin in all experimental settings, as shown in Tables 2 and 3. The few-shot learning methods with adversarial training strategy only focus on the domain-invariant features but neglect the domain-specific features which are the strong indicators for sentiment classification. Note that the adversarial training

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

strategy even degrades the performance in some cross-domain tasks (e.g., the cross-domain task *DVD→BOOK* in the 5-shot setting by comparing the model *Proto-BERT* and Proto-BERT†). We analyze that the domain-invariant features are scarce in the given few support instances, which limits the performance of sentiment analysis in the target domain.

Recently, two related works focusing on the cross-domain few-shot text classification (i.e., MLADA [54] and PtNet [55] in Tables 2 and 3) are proposed. They are also adapted into the cross-domain sentiment classification task. As shown in Tables 2 and 3, comparing with the performance of MLADA and PtNet, our proposed model achieves higher accuracies with a large margin. We observe that they mainly focus on capturing the domain-invariant features but ignore the domain-specific features. Moreover, with a few (e.g., 1 or 5) support instances of the target domain, few domain-specific features are contained, which limits the performance of the sentiment classification in the target domain. Instead, with the commonsense knowledge graph, rich domain-specific sentiment features can be expanded in our proposed model and improve the performance of sentiment classification in the target domain.

4.2.2 Comparison With Related Cross-Domain Sentiment Classification Models

Meanwhile, we compare the performance of our proposed model with current cross-domain sentiment classification models. As shown in Table 4, our proposed model achieves higher accuracies in all cross-domain sentiment classification tasks. As we can observe, most of the cross-domain sentiment classification models (e.g., *DANN*[40] , *PBLM*[35], *ACAN*[43] *IATN*[42] and so on) mainly focus on extracting the domain-invariant features by the way of unsupervised learning, but ignoring the domain-specific features. As the discrepancy between the source and target domains increases, the performance of these models will decrease substantially. They have a large gap in the performance among different cross-domain tasks. For example, the model *KinGDOM* almost have around 11% accuracy gap among the cross-domain tasks (e.g., the two cross-domain tasks E $\rightarrow$ B and E $\rightarrow$ K, as shown in Table 4). In contrast, our proposed model only has 5.5% accuracy gap among all the cross-domain tasks. To some extent, it can evaluate that our proposed model can effectively capture the domain-specific features of the target domain and narrow the discrepancy between the source and target domains. Currently, to capture the domain-specific features, several works (e.g., the *CoCMD*[14]) learn the domain-specific features by providing a few (i.e., 50) number of labeled samples of the target domain. However, they are also based on the deep neural networks and are prone to suffer from the overfitting problem. As shown in Table 4, our proposed model (with 5 support instances) even achieves better performance with a large margin than the *CoCMD* (with 50 support instances). It can evaluate that our proposed model can effectively solve the overfitting problem and learn the rich domain-specific features with only a few support instances. Moreover, similar to our proposed model, the model *KinGDOM*[9] also utilizes the external commonsense knowledge graph in the cross-domain adaptation. It adopts the adversarial training strategy to capture the domain-invariant features. Nevertheless, *KinGDOM* also ignores the domain-specific features which are also the strong indicators for the sentiment analysis for the target domain. Moreover, the relations between the aspect and opinion can not be modeled, which leads to the sentiment transfer error problem. As we can observe, our proposed model obtains higher accuracies with a large margin than the model *KinGDOM* in all cross-domain tasks, which can prove that our model can effectively capture the domain-specific features and the aspect-opinion correlation features in the external commonsense knowledge learning.

TABLE 5
Average Accuracies (%) Comparison With Supervised Learning Baselines

Target Domain	B	D	K	E
CNN	63.1	69.2	73.5	70.9
LSTM	79.6	76.2	81.5	77.7
BERT	87.0	88.3	91.0	89.9
AKFSM(AVG)	**91.7**	**91.7**	**94.1**	**96.5**

The results of our model are the average accuracies of the three cross-domain tasks.

Meanwhile, several supervised learning based baselines are compared with our proposed model. As shown in Table 5, both the unsupervised learning based methods (e.g., *BERT-DAAT*[8] and *SENTIX*[10]) and our proposed model even perform better than the supervised learning based baselines with many shots (i.e., 1000). We analyze that the performance of the supervised learning based baselines highly depends on large-scale labeled data of target domain. The provided 1000 labeled samples even can not satisfy the optimization of the deep neural network based methods with the thousands of parameters.

4.2.3 Ablation Study

Several ablation experiments are conducted, as shown in Table 6. Specifically, comparing with the performance of AKFSM$^{\clubsuit}$, our proposed model *AKFSM* achieves higher accuracies with a large margin in all experimental settings. Though the pre-trained language models have achieved remarkable performance in cross-domain NLP tasks [10], they ignore the domain-invariant sentiment-specific knowledge (e.g., the opinion words "*bad*" and the emoticon). The domain-adapted BERT encoder is enabled to learn and understand the semantics of the domain-invariant sentiment features, which benefits the cross-domain few-shot learning. In addition, to evaluate the effectiveness of the expanded relational knowledge in our proposed model, we conduct the ablation experiment for the module *Graph Feature Enocder*. Comparing with the performance of AKFSM$^{\diamond}$ and *AKFSM*, our model with the Graph Feature Encoder achieves higher performance with a large margin. We analyze that the relational knowledge can effectively enrich the domain-specific information based on the provided few support instances and benefit the performance of cross-domain sentiment analysis tasks.

Moreover, the related work *KinGDOM* [9] also adopts the external commonsense knowledge graph to enrich domain

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

TABLE 6
Ablation Experiments for the Modules of our Proposed Model in the 5-Shot Setting

S → T	AKFSM$^{\clubsuit}$	AKFSM$^{\diamondsuit}$	AKFSM$^{\square}$	AKFSM†	AKFSM‡	AKFSM$^{\spadesuit}$	AKFSM
D → B	86.0	89.2	89.6	91.0	91.3	91.9	**92.5**
E → B	81.8	89.0	88.4	90.1	91.0	90.5	**91.4**
K → B	83.8	89.1	86.4	90.7	90.4	90.8	**91.2**
B → D	85.0	89.4	89.0	92.0	89.5	90.9	**92.3**
E → D	82.2	88.8	90.4	91.0	89.5	90.6	**91.4**
K → D	83.8	85.8	88.4	89.9	89.3	90.2	**91.5**
B → E	83.8	89.4	89.6	92.5	92.5	92.8	**93.7**
D → E	82.6	89.2	90.2	92.1	92.5	92.1	**93.9**
K → E	89.8	90.3	92.0	93.5	93.8	93.9	**94.7**
B → K	85.6	89.8	91.2	94.8	95.2	95.5	**96.7**
D → K	83.4	90.8	91.4	95.8	95.7	95.3	**96.2**
E → K	87.6	85.0	92.4	95.5	95.8	94.9	**96.5**
Avg	85.4	88.8	89.9	92.4	92.2	92.5	**93.5**

AKFSM$^{\clubsuit}$ *denotes our proposed model which replaces the* Domain-adapted BERT Encoder *with BERT Encoder without post-training;* AKFSM$^{\diamondsuit}$ *denotes our proposed model without the* Graph Feature Encoder; AKFSM$^{\square}$ *denotes our proposed model which replaces the aspect-opinion correlation aware sentiment knowledge graph with the knowledge graph in KinGDOM [9];* AKFSM† *denotes our proposed model without graph feature reconstruction strategy.* AKFSM‡ *denotes our proposed model without self-supervised sentiment alignment learning task for the GCN encoder pretraining in Phase 1.* AKFSM$^{\spadesuit}$ *denotes our proposed model without the shared-knowledge aware attention.*

TABLE 7
Average Accuracies (%) Comparison of our Proposed Model With 1-hop Knowledge Linking Strategy and 2-hop Knowledge Linking Strategy in Phase 1

S → T		D → B	E → B	K → B	B → D	E → D	K → D	B → E	D → E	K → E	B → K	D → K	E → K	Avg
1-shot	1-hop	88.8	89.1	89.3	88.9	88.5	87.7	90.7	90.3	91.9	92.6	94.2	94.7	90.6
	2-hop	**90.5**	**89.3**	**91.5**	**90.5**	**89.0**	**88.4**	**91.3**	**91.4**	**93.2**	**93.9**	**94.6**	**94.9**	**91.5**
5-shot	1-hop	92.5	91.4	91.2	92.3	91.4	91.5	93.7	93.9	94.7	96.7	96.2	96.5	93.5
	2-hop	**93.8**	**92.0**	**92.0**	**93.0**	**92.0**	**91.8**	**93.9**	**94.0**	**94.8**	**96.8**	**96.5**	**96.7**	**93.9**

features for the cross-domain sentiment analysis task. Different from our proposed model, the relations between aspect and opinion terms cannot be modeled, which leads to the sentiment transfer error problem, as shown in Fig. 1. *KinGDOM* adopts the adversarial learning strategy to capture the domain-invariant features but filters out the domain-specific features which are also the strong indicators for the sentiment classification in the target domain. Specifically, comparing with the model AKFSM$^{\square}$, our proposed model AKFSM with the *Aspect-Opinion Correlation aware Graph Feature Learning* module performs better with a large margin, which can evaluate that our model can effectively solve the sentiment transfer error problem and capture the domain-specific features. Furthermore, we also conduct the ablation experiments for the representation fusion strategy (i.e., the feature mapping layer with a reconstruction loss $\pounds_{recon}$). Compared with the performance of the model AKFSM† in Table 6, our model with the reconstruction loss achieves better performance in all cross-domain experimental settings. The performance of the prototypical network highly depends on the spacial distribution of instance embeddings [51]. We consider that the discrepancy of the embedding spaces from the graph feature encoder and domain-adapted BERT encoder leads to bias in the distance metric, thereby degrading the performance of prototypical networks. The comparable experimental results can evaluate the effectiveness of the designed feature fusion strategy with reconstruction loss.

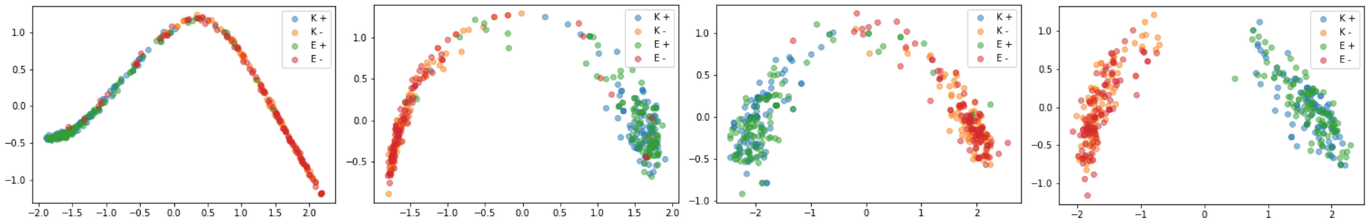
In addition, the self-supervised sentiment alignment task is designed in our model to force the GCN encoder to capture the sentiment alignment features among aspect-opinion pairs by exploiting their co-occurrence features within documents. To evaluate its effectiveness, an ablation experiment is conducted (i.e., the model AKFSM‡ and AKFSM in Table 6). The self-supervised sentiment alignment task in Phase 1 can effectively improve the performance of cross-domain sentiment analysis. We analyze that the sentiments of the aspect-opinion pairs can be derived through the sentiments of their contextual aspect-opinion pairs. The sentiment information (especially for the domain-specific features) can be captured by aligning the sentiment among the contextual aspect-opinion pairs.

Finally, we conduct the ablation experiments for the shared-knowledge aware attention. According to our observation, not all the support external knowledge benefits the sentiment analysis of the query instances. Motivated by this, the shared-knowledge aware attention is designed in our

TABLE 8
The Comparison of Resource Cost Between the Model With 1-hop Knowledge Triplets Strategy and 2-hop Knowledge Triplets Strategy

Settings	Num. of Triplets	Training Time	Memory Cost
1-hop	1 448 735	188.89 hour	5046Mib
2-hop	2 941 072	387.12 hour	15219Mib

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.



(a) *AKFSM*$^{\clubsuit}$ (b) *AKFSM*$^{\diamondsuit}$ (c) AKFSM$^{\spadesuit}$ (d) AKFSM

Fig. 5. Visualization analysis in cross-domain sentiment classification; $K+$ and $E+$ denote the positive category respectively from *Kitchen* domain and *Electronic* domain; $K-$ and $E-$ denote the negative category respectively from *Kitchen* domain and *Electronic* domain.

paper to select the important and relevant knowledge nodes in the graph. As shown in Table 6, comparing with AKFSM$^{\spadesuit}$ and AKFSM, our model with shared-knowledge aware attention obtains better performance, which evaluates the effectiveness of the shared-knowledge aware attention.

4.2.4 Analysis for N-Hop Knowledge Linking Strategy

We conduct the comparative experiments between our proposed model with 1-hop knowledge linking strategy and that with 2-hop knowledge linking strategy adopted in the Knowledge Graph Construction of Phase 1. As shown in Table 7, we can observe that the model with 2-hop knowledge linking strategy outperforms the model with 1-hop linking knowledge strategy by only less than 1%. Furthermore, we also conduct the resource cost comparison, as shown in Table 8. Under the same experimental settings (e.g., the batch size is set as 10 in the pre-training stage of Phase 1), the model with 2-hop knowledge linking strategy costs twice as much pre-training times as the model with 1-hop knowledge linking strategy. What's more, the model with 2-hop knowledge linking strategy needs nearly three times as much memory usage as the model with 1-hop knowledge linking strategy. Compared with the model with 1-hop knowledge linking strategy, the model with the 2-hop knowledge linking strategy has only a slight improvement in performance but requires more than twice or three times the resource cost. The selection of the number of hops in linking knowledge triplets is a trade-off between the performance and resource cost. The experimental results evaluate that the 1-hop linked knowledge triplets from the ConceptNet can already effectively describe and understand the terms of the reviews.

4.2.5 Visualization

To better understand the effectiveness of our proposed model, we randomly select 100 support instances from positive and negative categories and encode them into the hidden embeddings in the task of cross-domain (i.e., from *Kitchen* domain to *Electronic* domain) sentiment classification. Then, we map them into $2D$ points using Principal Component Analysis (PCA). As shown in Fig. 5, the instances expressing the same sentiment polarity are clustered together in the same distribution space, which demonstrates that the model performs better in domain adaptation task. Specifically, the effectiveness of domain-adapted BERT encoder can be evaluated by comparing with Figs. 5a and 5d. We can observe that the semantic understanding of the domain-specific (i.e., the target domain) features are significant for the cross-domain learning. Moreover, Fig. 5b shows the instance embedding distribution of our proposed model without the *Graph Feature Encoder*. Comparing with Figs. 5b and 5d, it can evaluate that our model with the aspect-opinion correlation aware graph feature learning module can effectively distinguish the positive and negative sentiment polarities in the same feature space. The relations between the aspect and opinion terms are beneficial to the cross-domain learning and effectively solve the sentiment transfer errors. Finally, we conduct the visualization analysis for our proposed model with the shared-knowledge aware attention. Comparing with Figs. 5c and 5d, we can find that the model with a shared-knowledge attention module can better distinguish the positive and negative sentiment polarities in the feature space, which can evaluate the effectiveness of the attention strategy in our model.

5 Conclusion

In this paper, we propose an aspect-opinion correlation aware and knowledge-expansion cross-domain sentiment classification model in the few-shot scenario. To solve the domain-specific features ignoring problem in current unsupervised domain adaptation methods, a few-shot cross-domain sentiment classification model is designed and can effectively capture the domain-specific features with only a few support instances. Furthermore, to solve the sentiment transfer error problem, we design an aspect-opinion correction aware graph learning module to capture the relational features between the aspect and opinion terms. The extensive experimental results and visualization analysis show that our proposed model obtains better performance in the cross-domain sentiment classification with only a few support data.

References

[1] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscipl. Rev. Data Mining Knowl. Discov.*, vol. 8, no. 4, 2018, Art. no. e1253.
[2] T. Shoryu, L. Wang, and R. Ma, "A deep neural network approach using convolutional network and long short term memory for text sentiment classification," in *Proc. IEEE 24th Int. Conf. Comput. Supported Cooperative Work Des.*, 2021, pp. 763–768.
[3] M. Yang, W. Yin, Q. Qu, W. Tu, Y. Shen, and X. Chen, "Neural attentive network for cross-domain aspect-level sentiment classification," *IEEE Trans. Affective Comput.*, vol. 12, no. 3, pp. 761–775, Jul.–Sep. 2019.
[4] H. Tang, Y. Mi, F. Xue, and Y. Cao, "Graph domain adversarial transfer network for cross-domain sentiment classification," *IEEE Access*, vol. 9, pp. 33 051–33 060, 2021.
[5] S. Zhang, X. Bai, L. Jiang, and H. Peng, "Dual adversarial network based on bert for cross-domain sentiment classification," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.*, 2021, pp. 557–569.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

[6] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics Syst. Demonstrations*, 2014, pp. 55–60.
[7] Z. Li, X. Li, Y. Wei, L. Bing, Y. Zhang, and Q. Yang, "Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process.*, 2019, pp. 4590–4600.
[8] C. Du, H. Sun, J. Wang, Q. Qi, and J. Liao, "Adversarial and domain-aware bert for cross-domain sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 4019–4028.
[9] D. Ghosal, D. Hazarika, A. Roy, N. Majumder, R. Mihalcea, and S. Poria, "Kingdom: Knowledge-guided domain adaptation for sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3198–3210.
[10] J. Zhou, J. Tian, R. Wang, Y. Wu, W. Xiao, and L. He, "SentiX: A sentiment-aware pre-trained model for cross-domain sentiment analysis," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 568–579.
[11] X. Zhang, J. Xu, Y. Cai, X. Tan, and C. Zhu, "Detecting dependency-related sentiment features for aspect-level sentiment classification," *IEEE Trans. Affective Comput.*, to be published, doi: 10.1109/TAFFC.2021.3063259.
[12] S. Wu, H. Fei, Y. Ren, D. Ji, and J. Li, "Learn from syntax: Improving pair-wise aspect and opinion terms extractionwith rich syntactic knowledge," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, 2021, pp. 3957–3963.
[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[14] M. Peng, Q. Zhang, Y.-G. Jiang, and X.-J. Huang, "Cross-domain sentiment classification with target domain specific information," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 2505–2513.
[15] H. Ren, Y. Cai, X. Chen, G. Wang, and Q. Li, "A two-phase prototypical network model for incremental few-shot relation classification," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 1618–1629.
[16] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. 31st Int. Conf. Neural Informat. Process. Syst.*, 2017, pp. 4080–4090.
[17] T. Gao et al., "FewRel 2.0: Towards more challenging few-shot relation classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process.*, 2019, pp. 6250–6255.
[18] A. Bražinskas, M. Lapata, and I. Titov, "Few-shot learning for opinion summarization," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2020, pp. 4119–4135.
[19] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, and J.-B. Huang, "A closer look at few-shot classification," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–17.
[20] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An open multilingual graph of general knowledge," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4444–4451.
[21] J. Xu, D. Chen, X. Qiu, and X.-J. Huang, "Cached long short-term memory neural networks for document-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1660–1669.
[22] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, 2016, pp. 1480–1489.
[23] Y. Yin, Y. Song, and M. Zhang, "Document-level multi-aspect sentiment classification as machine comprehension," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2044–2054.
[24] H. Chen, Z. Zhai, F. Feng, R. Li, and X. Wang, "Enhanced multichannel graph convolutional network for aspect sentiment triplet extraction," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 2974–2985.
[25] M. Zhao, J. Yang, J. Zhang, and S. Wang, "Aggregated graph convolutional networks for aspect-based sentiment classification," *Inf. Sci.*, vol. 600, pp. 73–93, 2022.
[26] A. Dai, X. Hu, J. Nie, and J. Chen, "Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis," *Int. J. Data Sci. Analytics*, vol. 14, no. 1, pp. 17–26, 2022.
[27] Z. Zhao, M. Tang, W. Tang, C. Wang, and X. Chen, "Graph convolutional network with multiple weight mechanisms for aspect-based sentiment analysis," *Neurocomputing*, vol. 500, pp. 124–134, 2022.
[28] W. Li, S. Yin, and T. Pu, "Lexical attention and aspect-oriented graph convolutional networks for aspect-based sentiment analysis," *J. Intell. Fuzzy Syst.*, vol. 42, pp. 1–12, 2022.
[29] D. Anand and B. S. Mampilli, "A novel evolutionary approach for learning syntactic features for cross domain opinion target extraction," *Appl. Soft Comput.*, vol. 102, 2021, Art. no. 107086.
[30] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 120–128.
[31] D. Bollegala, D. Weir, and J. Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1719–1731, Aug. 2012.
[32] J. Jiang and C. Zhai, "Instance weighting for domain adaptation in nlp," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 264–271.
[33] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.
[34] Z. Li, Y. Wei, Y. Zhang, and Q. Yang, "Hierarchical attention transfer network for cross-domain sentiment classification," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 5773–5780.
[35] Y. Ziser and R. Reichart, "Pivot based language modeling for improved neural domain adaptation," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, 2018, pp. 1241–1251.
[36] L. Li, W. Ye, M. Long, Y. Tang, J. Xu, and J. Wang, "Simultaneous learning of pivots and representations for cross-domain sentiment classification," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 8220–8227.
[37] T. Manshu and W. Bing, "Adding prior knowledge in hierarchical attention neural network for cross domain sentiment classification," *IEEE Access*, vol. 7, pp. 32 578–32 588, 2019.
[38] Y. Fu and Y. Liu, "Cross-domain sentiment classification based on key pivot and non-pivot extraction," *Knowl.-Based Syst.*, vol. 228, 2021, Art. no. 107280.
[39] A. Geethapriya and S. Valli, "An enhanced approach to map domain-specific words in cross-domain sentiment analysis," *Informat. Syst. Front.*, vol. 23, pp. 1–15, 2021.
[40] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.
[41] Z. Li, Y. Zhang, Y. Wei, Y. Wu, and Q. Yang, "End-to-end adversarial memory network for cross-domain sentiment classification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2237–2243.
[42] K. Zhang, H. Zhang, Q. Liu, H. Zhao, H. Zhu, and E. Chen, "Interactive attention transfer network for cross-domain sentiment classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5773–5780.
[43] X. Qu, Z. Zou, Y. Cheng, Y. Yang, and P. Zhou, "Adversarial category alignment network for cross-domain sentiment classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2019, pp. 2496–2508.
[44] J. D. M.-W. C. Kenton and L. K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
[45] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowl. Based Syst.*, vol. 235, 2022, Art. no. 107643.
[46] H. Wu, Z. Zhang, S. Shi, Q. Wu, and H. Song, "Phrase dependency relational graph attention network for aspect-based sentiment analysis," *Knowl. Based Syst.*, vol. 236, 2022, Art. no. 107736.
[47] S. Liang, W. Wei, X.-L. Mao, F. Wang, and Z. He, "BiSyn-GAT+: Bi-syntax aware graph attention network for aspect-based sentiment analysis," in *Proc. Findings Assoc. Comput. Linguistics*, 2022, pp. 1835–1848.
[48] Z. Zhang, Z. Zhou, and Y. Wang, "SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, 2022, pp. 4916–4925.
[49] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–13.
[50] T. Munkhdalai and H. Yu, "Meta networks," *Proc. Mach. Learn. Res.*, vol. 70, pp. 2554–2563, 2017.
[51] T. Gao, X. Han, Z. Liu, and M. Sun, "Hybrid attention-based prototypical networks for noisy few-shot relation classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 6407–6414.
[52] R. Geng, B. Li, Y. Li, J. Sun, and X. Zhu, "Dynamic memory induction networks for few-shot text classification," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 1087–1094.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.

[53] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *Statistical*, vol. 1050, 2015, Art. no. 20.

[54] C. Han, Z. Fan, D. Zhang, M. Qiu, M. Gao, and A. Zhou, "Meta-learning adversarial domain adaptation network for few-shot text classification," in *Proc. Findings Assoc. Comput. Linguistics: ACL-IJCNLP*, 2021, pp. 1664–1673.

[55] C. Zhang and D. Song, "A simple baseline for cross-domain few-shot text classification," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.*, 2021, pp. 700–708.

[56] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 9, pp. 1616–1637, Sep. 2018.

[57] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proc. Eur. Semantic Web Conf.*, 2018, pp. 593–607.

[58] H. T. Phan, N. T. Nguyen, and D. Hwang, "Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis," *Informat. Sci.*, vol. 589, pp. 416–439, 2022.

[59] B. Yang, S. W.-T. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–12.

[60] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 440–447.

[61] N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, "A simple neural attentive meta-learner," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–17.

[62] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2019, pp. 4171–4186.

Haopeng Ren received the BS degree from the School of Information and Computing Science, Zhongkai University of Agriculture and Engineering, Guangdong, China. He is currently working toward the PhD degree with the School of Software Engineering, South China University of Technology, Guangdong, China. His research interest focuses on information extraction, knowledge graph, and natural language processing.

Yi Cai (Member, IEEE) received the PhD degree in computer science from the Chinese University of Hong Kong. He is currently a professor in South China University of Technology (SCUT). His research interests include recommendation system, personalized search, semantic web, and data mining. His research works are published on many conferences and journals, such as *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, *Neural Networks*, *Knowledge-based Systems*, *EAAI and Neurocomputing*, as well as AAAI, COLING, CIKM, AAMAS, DASFAA and other international conferences about perspective mining, cognitive modeling, information retrieval and semantic analysis.

Yushi Zeng received the BS degree from the School of Software Engineering, Jishou University of Agriculture and Engineering, Hunan, China. She is currently working toward the master degree with the School of Software Engineering, South China University of Technology, Guangdong, China. Her research interest focuses on information extraction, knowledge graph, sentiment analysis, and natural language processing.

Jinghui Ye is currently working toward the MPhill degree with the Hong Kong University of Science and Technology (Guangzhou). Moreover, he is taking an internship in Tencent AI Lab. His current research interest is visual-language problems, including summarization, VQA, and sign language translation.

Ho-fung Leung (Senior Member, IEEE) received the BSc and MPhil degrees in computer science from the Chinese University of Hong Kong, and the PhD degree from the University of London, with DIC (Diploma of Imperial College) in Computing from Imperial College London. He is a professor in the Department of Computer Science and Engineering with the Chinese University of Hong Kong. He has authored more than 250 publications, including 5 research monographs, and 5 edited volumes. He was the chairperson of ACM (Hong Kong Chapter) in 1998. He is a Chartered Fellow of the BCS, a Fellow of the HKIE, a full member the HKCS. He is a Chartered Engineer registered by the Engineering Council.

Qing Li (Senior Member, IEEE) received the BEng degree in computer science from Hunan University, Changsha, and the MSc and PhD degrees in computer science, from the University of Southern California, Los Angeles. He is currently a chair professor with the Department of Computing, The Hong Kong Polytechnic University. His research interests include multi-modal data management, conceptual data modeling, social media, Web services, and e-learning systems. He has authored or coauthored more than 400 publications in these areas. He is actively involved in the research community. He is a Fellow of IEE/IET, U.K., and a Distinguished Member of CCF, China.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.

Authorized licensed use limited to: The University of British Columbia Library. Downloaded on January 27,2023 at 07:05:22 UTC from IEEE Xplore. Restrictions apply.