T-LBERT with Domain Adaptation for Cross-Domain Sentiment Classification

Hongye Cao
School of Software, Northwestern Polytechnical University, China
2020204278@mail.nwpu.edu.cn

Qianru Wei
School of Software, Northwestern Polytechnical University, China
weiqianru@nwpu.edu.cn

Jiangbin Zheng
School of Software, Northwestern Polytechnical University, China
zhengjb@nwpu.edu.cn

Abstract: *Cross-domain sentiment classification transfers the knowledge from the source domain to the target domain lacking supervised information for sentiment classification. Existing cross-domain sentiment classification methods establish connections by extracting domain-invariant features manually. However, these methods have poor adaptability to bridge connections across different domains and ignore important sentiment information. Hence, we propose a Topic Lite Bidirectional Encoder Representations from Transformers (T-LBERT) model with domain adaption to improve the adaptability of cross-domain sentiment classification. It combines the learning content of the source domain and the topic information of the target domain to improve the domain adaptability of the model. Due to the unbalanced distribution of information in the combined data, we apply a two-layer attention adaptive mechanism for classification. A shallow attention layer is applied to weigh the important features of the combined data. Inspired by active learning, we propose a deep domain adaption layer, which actively adjusts model parameters to balance the difference and representativeness between domains. Experimental results on Amazon review datasets demonstrate that the T-LBERT model considerably outperforms other state-of-the-art methods. T-LBERT shows stable classification performance on multiple metrics.*

Keywords: *Cross-domain, sentiment classification, topic model, attention, domain adaption.*

Received November 1, 2021; accepted April 4, 2022
https://doi.org/10.34028/iajit/20/1/15

1. Introduction

Sentiment classification is an important research topic in the field of natural language processing [4, 31]. In current e-shopping websites, sentiment of product evaluation has become increasingly important. Sentiment analysis can help businesses and consumers make right decisions [11, 14, 24]. The development of large-scale language models has resulted in the prevalent use of language models in sentiment classification [20]. Bidirectional Encoder Representations from Transformers (BERT) [6] and A Lite BERT (ALBERT) [12] model can capture the sentiment information through supervised learning on the specific domain dataset.

However, many domains have insufficient labeled data. The language model cannot achieve supervised learning in the target domain. Manually labeling the target domain data requires considerable labor costs. Therefore, cross-domain sentiment classification is an important research hotspot. Using the knowledge of the source domain to achieve sentiment classification inthe target domain can reduce the cost of manual labeling and effectively improve the utilization of source domain data.

A traditional solution to cross-domain sentiment classification is instance adaption [1, 30, 33]. This method realizes data resampling which combines the target domain data to filter the source domain label data. The language model is trained on the resampled dataset to reduce the difference between source and target domains. However, instance adaptation requires a lot of manual intervention [7]. The classification effect of instance adaptation declines when the source domain data and the target domain data are quite different.

Existing cross-domain sentiment classification method is feature adaptation which builds data feature maps or finds pivot features to establish connections between domains [9, 17, 28]. It realizes the sentiment classification through the feature correspondence between domains. With the development of deep learning, deep learning methods have been widely used to extract the domain-invariant features between the target and source domains [8, 19, 27]. However, feature adaptation method is highly dependent on feature selection among data and ignores important sentiment information. This method has poor adaptability across different domains.

Although the above methods achieve good classification results, they have limitations. They have poor transfer adaptability and require a lot of manual intervention. The active learning [13, 22] method can learn the association and access probability loss from the feature representation generated by the deep neural network through the adaptive mechanism. The adaptive optimization of parameters can improve the adaptability of the model. Hence, we propose a Topic

Lite Bidirectional Encoder Representations from Transformers (T-LBERT) model. It combines the supervised learning content of the A Lite Bidirectional Encoder Representations from Transformers (ALBERT) model in the source domain with the unsupervised topic information extracted by the topic model in the target domain. Applying the combined data into a two-layer attention adaptive network for classification. A shallow attention [25] layer is applied to weigh the important features of the combined data. Inspired by active learning, a deep domain adaption layer is proposed to actively adjust model parameters to balance the difference and representativeness between domains.

The main contributions of this article are as follows:

- We propose a T-LBERT model, which fuses the learning content of the source domain with the topic information of the target domain. We fully mine the sentiment information of the source and target domains.
- A shallow attention layer is applied to enhance important features. A deep domain adaptation layer is proposed to effectively balance the difference and representativeness between domains to improve adaptability.
- We compare the proposed model with the current advanced cross-domain sentiment classification models on the Amazon review datasets. The experiment results show that T-LBERT considerably outperforms the state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 introduces related work on cross-domain sentiment classification. Section 3 provides the overview of proposed framework T-LBERT and analyzes each component. Section 4 introduces the experimental datasets and settings, and introduces the evaluation metrics. Then, topic model selection, comparison experiments, and ablation experiments are introduced. Section 5 summarizes the conclusion of the paper.

2. Related Works

Cross-domain sentiment classification has been a challenging research topic in the field of natural language processing. The problems of cross-domain sentiment classification focus on applying the labeled data of the source domain to predict the sentiment labels of the unlabeled data in the target domain. Cross-domain sentiment classification can be divided into three types of solutions: instance adaption, feature adaption, and model adaption.

The main idea of instance adaption method is to start from the source domain data and combines the target domain data instance to resample the training data [10, 30]. This method can balance the data differences caused by different domains. It can balance the sample distribution. The classification model is trained based on the newly sampled data to achieve sentiment classification. However, resampling the target domain instances is difficult when the source domain and target domain differ greatly.

Compared with instance adaption method, feature adaption method aims to construct the domain-invariant features between domains. It conducts joint training on the source and target domains [5, 23]. Feature adaptation explores the association between domains. The proposed Structural Correspondence Learning (SCL) algorithm [3] is based on alternating structure optimization multitask learning algorithm. It models the correlation between hub features and other features. Pan *et al.* [18] proposed a Spectral Feature Alignment (SFA). They used graph-based clustering to achieve cross-domain sentiment classification by finding unified and independent features between domains and constructing collaborative relationships between features. Jia *et al*. [9] proposed a cross-domain sentiment classification method based on association rules for word alignment. Wang *et al*. [26] learned the sentiment coefficient from the data features and used the sentiment coefficient to classify the data in different domains. Given the need to construct the feature correspondence between the source and target domains, feature optimization algorithm is relatively slow. It also ignores important sentiment information. Feature adaptation method has poor adaptability. It is not suitable for general model applications.

The model adaption method starts from the training of the model and fine-tunes the model through the training of data in different domains [32]. Model adaption methods are divided into single model adaption and joint training model adaption. For the single model adaptation, the important features can be highlighted through the attention mechanism. Myagmar *et al*. [16] fine-tuned the cross-domain sentiment classification model BERT. They effectively solved the task of cross-domain sentiment classification. Du *et al*. [7] proposed a transmission network based on Wasserstein to share the domain invariant information of the source and target domains. They obtained a wealth of knowledge through BERT and acquired the textual semantic information of the text. The attentional recurrent neural network is used to capture features automatically. The important domain invariant features are captured through confrontation training. Zhao *et al*. [33] proposed a cross-domain sentiment classification method realized by parameter transfer and attention sharing mechanism. Joint training model adaption is fine-tuned through joint training of multiple language models. Different language models use their advantages to achieve deep capture of data features. Meng *et al*. [15] proposed an attention network based on feature sequence for cross-domain sentiment classification. They used attention mechanism to pay attention to important semantic features.

The above review shows that model adaption method can effectively improve the model performance of cross-domain sentiment classification for cross-domain sentiment classification tasks. We apply model adaption method to cross-domain classification. We train the source domain data based on the ALBERT model and fuse it with the topic model, which extracts topic feature information from the target domain data. We apply a shallow attention layer to weigh the features of the fusion data. Then, we apply the deep domain adaption layer to predict the sentiment label.

3. Proposed Model

We propose the T-LBERT model, which combines the learning content of the source domain with the topic information of the target domain. The combined data are applied to a two-layer attention adaptive mechanism for classification. The architecture of this model is shown in Figure 1.

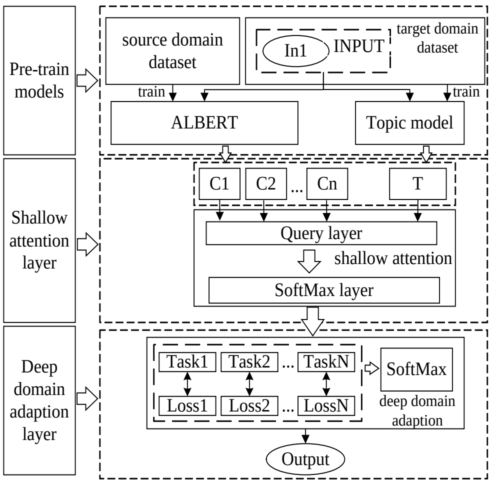


Figure 1. Architecture of the T-LBERT model.

ALBERT model learns knowledge from the source domain labeled data in a supervised way. The topic model extracts the topic features of the target domain in an unsupervised way and integrates the learning content of the source domain. This joint learning method makes full use of domain data. Then, the important features of the combined data are weighed through the shallow attention layer to capture key sentiment information. The deep domain adaptive layer is proposed to actively adjusts parameters via active learning. This attention adaptive mechanism can effectively balance the difference and representativeness between domains to improve adaptability.

3.1. Pre-Train Models

We first process data and pretrain models. We apply the source domain labeled dataset to train the ALBERT model. ALBERT is used to process the data and obtain the learning content vector.

ALBERT uses [SEP] as a separator to separate the input sentence after segmentation. The special character [CLS] is used for downstream classification tasks. After segmenting and encoding the input data, we input the encoded data into a two-way transmission network and obtain the vector.

The vector C in the last layer of ALBERT is used to represent the sentence, as shown in the following formula.

$$C = ALBERT(In_1) \tag{1}$$

Where In_1 is the input sentence, and C is the output vector of ALBERT model.

Then, we use the dataset of the target domain without labels to perform topic model training in an unsupervised way. We obtain a topic model for topic information extraction. We input the text into the topic model and obtain the topic vector. First, we divide the sentence into the following forms:

$$In_1 = [s_1, ..., s_j, ..., s_m] \tag{2}$$

In this formula, s_j represents the word segmentation items of the data.

We place the divided data items into the topic model and calculate the topic information of the data items, as shown in the following formula.

$$T = Topic(s_1, ..., s_j, ..., s_m) \tag{3}$$

$$T = \{t_1, ..., t_i, ...t_n\} \tag{4}$$

Where t_i is the output topic vector of the topic model.

Finally, the topic information t_i of the target domain is combined with the output C of ALBERT and put the combined data into the shallow attention layer.

3.2. Shallow Attention Layer

Due to the unbalanced distribution of sentiment information in the combined data, we propose a two-layer attention adaptive mechanism for classification. A shallow attention layer is applied to capture important features of domain data. It can enhance the weight of important features in a sentence. The following Figure 2 shows the process of the shallow attention layer. We use a small amount of target domain labeled data to fine-tune the weight of the attention layer.

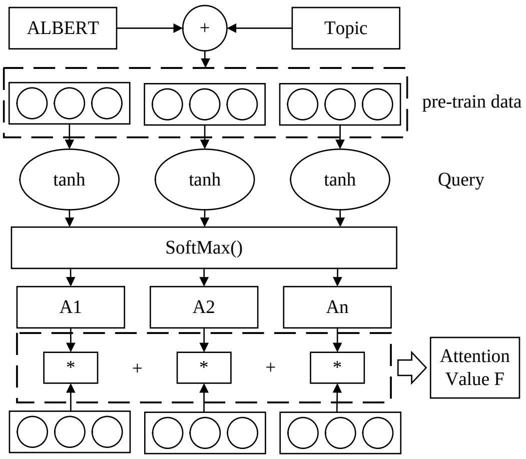


Figure 2. The process of shallow attention layer.

We first connect the topic information and the output vector of ALBERT.

$$h_i = C \oplus t_i \tag{5}$$

Where $\oplus$ is a connection operator and t_i is the i-th output topic vector of a single input text in the topic model. Then, we input h_i into the *tanh*-function of attention to obtain the attention score A_i.

$$A_i = \tanh(w_i h_i + b_i) \tag{6}$$

Where w_i and b_i represent the attention parameter. Then, the normalization process is performed to obtain the attention weight matrix α_i.

$$\alpha_i = \frac{\exp(A_i)}{\sum_i \exp(A_i)}, \sum_i a_i = 1 \tag{7}$$

Where A_i is the attention score. We use the matrix α_i to get the final weighted feature data representation F.

$$F = \{x_1, ..., x_i, ... x_n\} \tag{8}$$

$$x_i = a_i * h_i \tag{9}$$

Where α_i is the attention weight matrix and h_i is the combined data item. We input the data item F into the deep domain adaption layer.

3.3. Deep Domain Adaption Layer

We apply the above shallow attention to weigh important features. Then, we propose a deep domain adaption layer to effectively balance the difference and representativeness between domains by actively optimizing model parameters. The Algorithm 1 is the deep domain adaption layer classification algorithm. We integrate the input data In_i into the data item x_i (lines 2-5). Then, we place the data item x_i into the function of Domain-adapt (line 6). According to the probability calculated by this function, the label item with the highest probability is set as the sentiment classification label for the input data. This label is added to the sentiment label set L (lines 7-10). Finally, we return the sentiment label set L. Next, we explain how to calculate the domain adaption parameter λ in the function of Domain-adapt.

We apply the method of active learning to deep domain adaptation. The weight factor of deep domain adaption is adjusted adaptively along with the labeling sample process. Inspired by multi-task learning weight optimization, we adopt an adaptive objective function to learn weights. Its objective function can be expressed as the following formula:

$$T_{total} = \sum_i \lambda_i T_i \tag{10}$$

In this objective function, T_i is the output of the subtask, and λ_i is the task weight parameter. The choice of parameter λ_i is crucial.

In this study, adaptive dynamic adjustment is used to adjust the weight value. For an input data, we set the classification task of the combined data item of each topic information as a subtask. We define $f^{\lambda}(x)$ as the output of x in the proposed model, and the probability vector is defined as follows:

$$P(y \mid f^{\lambda}(x)) = soft\max(f^{\lambda}(x)) \tag{11}$$

Where x represents an output of the shallow attention layer for an input data. The likelihood of its multiple subtasks is defined as follows:

$$P(y_1, ..., y_n \mid f^{\lambda}(x)) = P(y_1 \mid f^{\lambda_1}(x_1)) ... P(y_n \mid f^{\lambda_n}(x_n)) \tag{12}$$

The output is $y_1, \dots, y_n$. We use the same variance uncertainty as the basis of the weighted loss in the learning problem. The same variance uncertainty is the uncertainty index for task discrimination.

A loss function based on Gaussian likelihood maximization with homoscedastic uncertainty is derived as an optimization function for minimization [13]. The function is as follows:

$$T = -\log P(y_1, ..., y_n \mid f^{\lambda}(x)) \propto \frac{1}{2\beta_1^2} \| y_1 - f^{\lambda_1}(x) \|^2 + ... + \frac{1}{2\beta_n^2} \| y_n - f^{\lambda_n}(x) \|^2 + \log \beta_1 ... \beta_n \tag{13}$$

$$T = \frac{1}{2\beta_1^2} l_1(\lambda_1) + ... + \frac{1}{2\beta_i^2} l_i(\lambda_i) + ... + \frac{1}{2\beta_n^2} l_n(\lambda_n) + \log \beta_1 ... \beta_i ... \beta_n \tag{14}$$

Where $l_1(\lambda_1)$ is the loss function of the first output variable. The same goes for the $l_i(\lambda_i)$. We can obtain the sum of β_i according to the training data. We learn the weight parameters β_i for the loss functions $l_i(\lambda_i)$. The weight of loss $l_i(\lambda_i)$ decreases when β_i increases. We use a small number of target domain labeled data to perform active learning of adaptive parameters and obtain adaptive parameters.

Finally, we fuse the adaptive weight parameters with the data processed by the shallow attention layer. The prediction of the data sentiment label is performed. The SoftMax function of the predicted label is shown in the following formula.

$$P = soft\max(\sum_{i=1}^{n} x_i \lambda_i) \tag{15}$$

In formula (15), x_i is a combined data item processed by shallow attention layer of the weighted feature data representation F, and λ_i is the parameter obtained by adaptive adjustment. The adaptive mechanism can effectively balance the weights of different topic feature information. The label of the highest probability is marked as the target data label by calculating the label probability.

Algorithm 1: deep domain adaption layer classification

Input: target domain dataset T_U,
source domain dataset S_L,
domain adaption parameter λ.
Output: sentiment classification label set L
Input: target domain dataset T_U,
source domain dataset S_L,
domain adaption parameter λ.
Output: sentiment classification label set L
1: repeat
2: input data $\{In_i\}_{i=1}^{m}$ from T_U
3: c_i = ALBERT (In_i, S_L)
4: t_i = Topic (In_i, T_U)
5: x_i = Attention-con (c_i, t_i)
6: P = Domain-adapt (x_i, λ)
7: if P [1] > P [2] then
8: $L \leftarrow L \cup \{L_P[1]\}$
9: else

4. Experiment Design and Analysis

4.1. Experiment Datasets and Settings

We select the Amazon review dataset [21], which has data in four domains: books (B), DVD (D), electronics (E), and kitchen (K). The description of the dataset is shown in Table 1. We make full use of the reviews in the four domains of these datasets. We create 12 cross-domain sentiment classification tasks to test the performance of our approach: D→B, E→B, K→B, B→D, E→D, K→D, B→E, D→E, K→E, B→K, D→K, E→K, where the domain at the left of the arrow represents the source domain, and the domain at the right of the arrow represents the target domain.

We choose 6000 labeled data of the source domain as the training dataset of the ALBERT model. The unlabeled data of the target domain is used for the training of the topic model. 100 labeled data of target domain are randomly selected for model fine-tuning. All remaining labeled data in the target domain will be used for testing.

The experimental parameters of T-LBERT model are set as follows. The word vector dimension is set to 300. The basic ALBERT model comprises 12 layers and 60 M parameter. The gradient descent training method and the adaptive learning rate method Adam are used in the training process. The initial learning rate is set to 0.001. The dropout rate is set to 0.5. We limit the maximum input length to 512. All models in this experiment are implemented in the TensorFlow framework. The model experiment is completed in the following environment: Intel(R) Core (TM) i7-7700 CPU @ 3.60GHz, RAM 16.0 GB, and NVIDIA GeForce GTX 1080 8.0 GB.

Table 1. Dataset description.

Domain	Positive	Negative	Unlabeled
Electronics	3000	3000	17009
Kitchen	3000	3000	13856
Books	3000	3000	9750
DVD	3000	3000	11843

4.2. Experiment Evaluation Index

We use accuracy, precision, recall, and F1-score as evaluation indicators in the cross-domain sentiment classification applications. The calculation formulas are as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

$$precision = \frac{TP}{TP+FP} \tag{17}$$

$$recall = \frac{TP}{TP+FN} \tag{18}$$

$$F-score = 2 * \frac{precision * recall}{precision + recall} \tag{19}$$

In the above formulas, the True Positive (TP) represents the number of correctly identified matched data. The False Positive (FP) refers to the number of wrongly identified data; in particular, the model identifies the unmatched function data as matched. The False Negative (FN) refers to the number of wrongly identified unmatched data. The True Negative (TN) is the negative samples that are also predicted as negative samples.

Accuracy represents the classification accuracy of the model. This indicator is the most important for model classification. Precision measures the percentage of the matched function pairs that are correctly labeled. Recall represents the ability to identify matched function pairs correctly. F1-score refers to the harmonic mean of precision and recall.

4.3. Analysis of Experiment Results

4.3.1. Selection of Topic Model

Different topic models have different ways to extract topic information. To select an optimal topic model, we choose the Biterm Topic Model (BTM) [29] and the Latent Dirichlet Allocation (LDA) topic model [2] for comparative experiments. The LDA topic model is the most widely used topic model, and the BTM topic model is a topic information extraction model for short texts. Both models can achieve accurate capture of text topics. We combine the two models with ALBERT to perform sentiment classification and apply into 12 cross-domain tasks separately.

As shown in the Figure 3, the experimental results indicate that among the 12 cross-domain task experiments, the classification accuracy of the LDA-ALBERT model is higher than the classification accuracy of BTM-ALBERT model. This finding shows that LDA model is more suitable for model combination than BTM model. Hence, we choose the LDA model as the topic model for T-LBERT model.

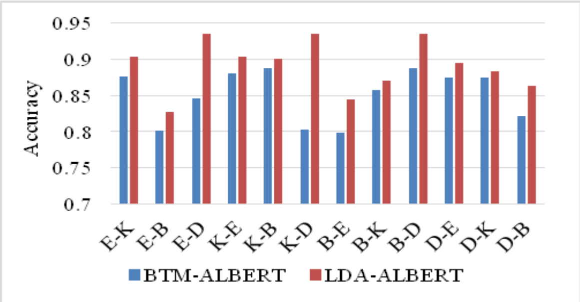


Figure 3. Topic model comparison results.

4.3.2. Comparative Experiment

After constructing the T-LBERT model, we set up the comparative experiment to compare the model classification effect. We choose BERT [6], ALBERT [12], and LDA [2], SCL [3] and SFA [18] models for comparisons. The accuracy classification results in 12 cross-domain tasks are shown in Table 2. The proposed T-LBERT model has the highest accuracy values in the 12 cross-domain tasks. The accuracy of each task of the calculation accuracy rate reaches more than 80%. The average accuracy of T-LBERT model reaches 88.78%. It is higher than the average accuracy of the other five models.

Table 2. Comparative experiment results of accuracy.

Task	BERT (%)	ALBERT (%)	LDA (%)	SCL (%)	SFA (%)	T-LBERT (%)
E-K	73.02	76.00	77.14	82.06	84.15	**90.25**
E-B	47.79	58.75	45.21	71.42	72.42	**82.75**
E-D	82.94	88.75	78.90	74.25	75.50	**93.50**
K-E	64.35	79.25	79.12	81.74	82.92	**89.25**
K-B	59.86	58.75	65.88	71.28	73.60	**89.10**
K-D	84.62	88.75	84.13	75.25	77.20	**91.50**
B-E	68.62	66.00	70.12	76.42	76.20	**84.50**
B-K	70.55	66.75	62.15	77.08	78.22	**87.00**
B-D	64.53	88.75	72.82	79.70	82.74	**93.50**
D-E	81.48	77.75	68.14	75.30	77.34	**89.50**
D-K	79.12	72.50	72.10	77.84	78.40	**88.25**
D-B	58.72	72.25	66.19	79.16	80.14	**86.25**
AVE	69.63	74.52	70.16	76.79	78.24	**88.78**

The average accuracy of the basic language model BERT, ALBERT and LDA models are less than 75%. The average accuracy of SCL and SFA models only reached 76.79% and 78.24%. These feature adaption methods have weak domain adaptability due to the selection of pivot features. The feature adaption methods between non-similar domains have lower classification accuracy. For example, on the task of E-B and K-B, the accuracy of the SCL and SFA models is lower than 75%.

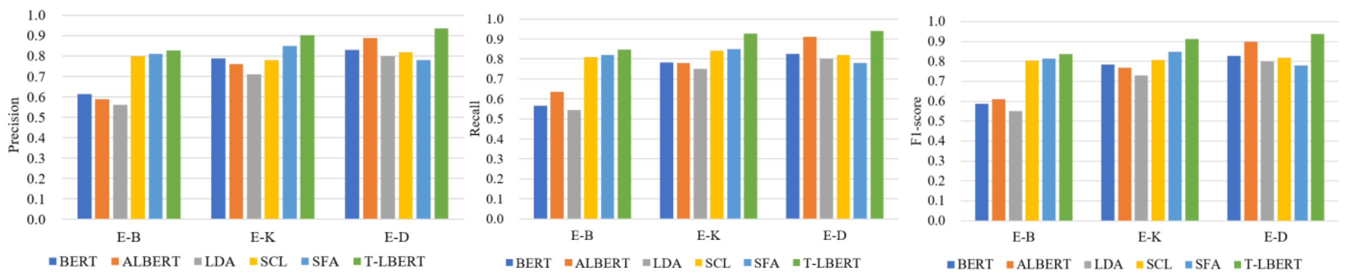
On the contrary, the proposed T-LBERT model can effectively capture the sentiment information between domains. On tasks E-D, the accuracy rate can reach 93.50%. It indicates that our model can accurately capture sentiment information from similar domain data. For domains with tremendous differences, such as DVD as the source domain and kitchen as the target domain, the accuracy rate can reach 88.25%. T-LBERT model can also accurately classify tasks with huge differences between domains. The reason is that we introduce the topic features of the target domain into the prior knowledge of the source domain. We make full use of the information in the target domain. Domain adaptation method can balance the uncertainty and difference of data in different domains. This mechanism makes the model much adaptable to datasets in different domains.

Then, we use three other indicators of precision, recall, and F1-score to evaluate the classification effect of models comprehensively. As shown in Figures 4, 5, 6, and 7 the classification value of T-LBERT model is higher than 0.8 in the 12 cross-domain tasks in the precision calculation. The precision value of T-LBERT model is higher than the precision value of the other five models. This finding shows that the T-LBERT model can achieve accurate classification of sentiment data. In the recall calculation, the recall calculated by T-LBERT model is higher than 0.8. It is higher than that by the other five models. Finally, the values of T-LBERT model in the 12 cross-domain tasks are all higher than the values of the other five models in F1-score calculation. The calculation of three evaluation indicators can explain that the classification effect of T-LBERT model is better than that of the other five models in 12 cross-domain sentiment classification tasks. The proposed model is highly adaptable in multiple domains and has good model migration ability.

Moreover, we select four instances from B-E and E-B tasks to analyze the sentiment feature recognition in Table 3. Words marked in bold in each sentence are identified as the most important sentiment feature. The sentiment information of “hope” and “great” in positive sentences can be accurately captured. The sentiment information of “boring” and “bad” in negative sentences can also be accurately identified. We can find that T-LBERT can effectively identify the most important sentiment information in sentences. And T-LBERT is not disturbed by noisy data such as “horrendous”, “interesting” and “small”. This further demonstrates the robustness and adaptability of the proposed model.

Table 3. Visualization of important feature selection in T-LBERT of the B-E and E-B tasks.

Book domain	Electronic domain
Label: positive Despite the horrendous abuse, this story gives me **hope** in mankind again.	Label: positive the disk told me how to fix that. I like it, I think it **great**, light and small.
Label: negative I found it interesting but somehow **boring** as the above story developed little and the focus was on the characters.	Label: negative The quality of the recording was very **bad**. Little squares showed up all the time. I would definitely not buy it again.

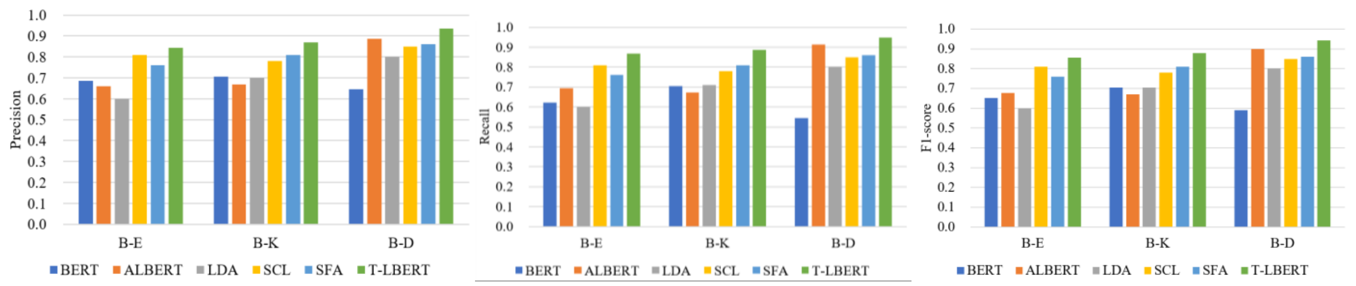


a) Precision. b) Recall. c) F1-score.

Figure 4. Source domain of electronics.

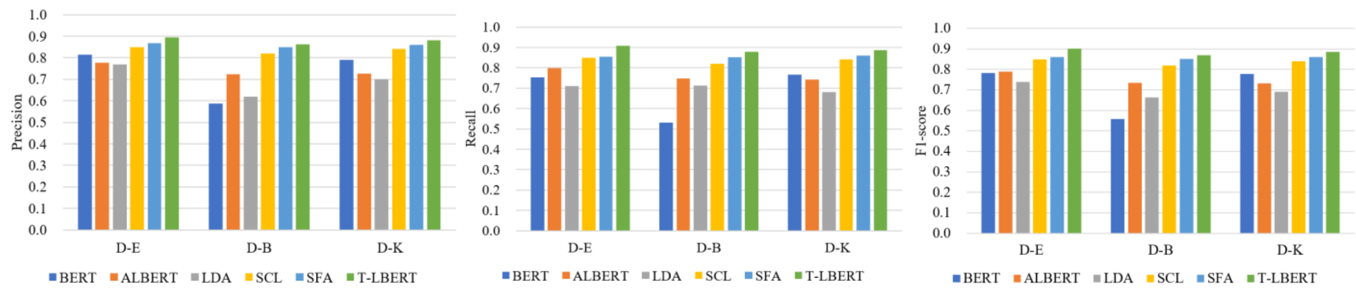
a) Precision. b) Recall. c) F1-score.

Figure 5. Source domain of kitchen.



a) Precision. b) Recall. c) F1-score.

Figure 6. Source domain of book.



a) Precision. b) Recall. c) F1-score.

Figure 7. Source domain of DVD.

4.3.3. Ablation Experiment

Our proposed model achieves cross-domain sentiment classification through a two-layer attention adaptive mechanism. We set up the ablation experiment to verify the effectiveness of the shallow and deep layers. We separately set the models without attention adaptive layers, without deep domain-adaption layer, without shallow attention layer and the proposed model. We apply these four models to classify 12 cross-domain classification tasks and evaluate models by accuracy. The results of the ablation experiment are shown in Table 4. Among the models, the proposed T-LBERT model has the highest accuracy on 12 tasks. The average accuracy rate of T-LBERT model is higher than that of the other three models, obtaining 15.74%, 6.55%, and 3.58%. This finding shows that the shallow attention and deep domain layers can effectively improve the accuracy of cross-domain classification.

For the model with only a shallow attention layer, the classification accuracy of the model can reach more than 75%. It is higher than the model without attention adaptive layers. This shows that the shallow attention mechanism weights the important features in the data and improves the accuracy of the model.

Moreover, the accuracy of deep domain-adaption model is higher than that of the shallow layer attention model on 12 tasks. This finding shows that our deep domain-adaption mechanism can effectively improve domain adaptability. When the target domain is the kitchen, the model can achieve the accuracy of over 85% for tasks in three different source domains. When the source domain is the electronics, the accuracy of the model is over 80% for tasks in three different target domains. This shows that the deep adaptive mechanism effectively improves the multiple domains adaptability of the model.

Table 4. Ablation experiment results of accuracy (w/o represents without).

Task	T-LBERT (w/o attention adaptive layers)	T-LBERT (w/o deep domain adaption)	T-LBERT (w/o shallow attention)	T-LBERT (proposed)
E-K	76.10%	84.12%	85.10%	**90.25%**
E-B	68.13%	78.91%	81.35%	**82.75%**
E-D	78.49%	86.54%	89.01%	**93.50%**
K-E	77.95%	79.45%	85.42%	**89.25%**
K-B	60.11%	81.86%	85.15%	**89.10%**
K-D	79.09%	85.42%	87.08%	**91.50%**
B-E	66.01%	76.42%	80.12%	**84.50%**
B-K	65.45%	80.55%	86.62%	**87.00%**
B-D	79.49%	88.53%	89.05%	**93.50%**
D-E	78.23%	82.76%	84.25%	**89.50%**
D-K	74.29%	83.07%	85.14%	**88.25%**
D-B	73.10%	79.12%	84.10%	**86.25%**
AVE	73.04%	82.23%	85.20%	**88.78%**

5. Conclusions

A T-LBERT model based on domain adaptation is presented in this study for cross-domain sentiment classification, which, in previous works, suffers from poor adaptability across different domains. The proposed model combines topic features from the target domain with the learning content of the ALBERT model. A shallow attention layer is applied to weigh important features. A deep domain adaption layer is applied to effectively balance the difference and representativeness between domains by actively optimizing model parameters. Through the ablation experiment, it is proved that this two-layer attention adaptive mechanism improves the adaptability of the model in multiple domains. We set up experiments with advanced cross-domain sentiment classification models. The experimental results show that the proposed model considerably outperforms the state-of-the-art models. The stable classification performance on cross-domain tasks shows that the model has strong domain adaptability and portability. In future work, we will consider using much complex topic models and apply this method to other fields, such as machine translation and entity recognition.

Acknowledgements

The work is supported by the National Social Science Foundation of China under Grant No. 21CTQ024.

References

[1] Beigi O. and Moattar M., "Automatic Construction of Domain-Specific Sentiment Lexicon for Unsupervised Domain Adaptation and Sentiment Classification," *Knowledge-Based Systems*, vol. 213, pp. 106423-106426, 2021.

[2] Blei D., Ng Y., and Jordan M., "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[3] Blitzer J., Dredze M., and Pereira F., "Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification," *in Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, pp. 440-447, 2007.

[4] Cambria E., Das D., Bandyopadhyay S., and Feraco A., *Affective Computing and Sentiment Analysis*, *A Practical Guide to Sentiment Analysis*, Springer, 2017.

[5] Chen T., Liao Y., Chuang C., Hsu W., Fu J., and Sun M., "Show, Adapt and Tell: Adversarial Training of Cross-Domain Image Captioner," *in Proceedings of the IEEE International Conference on Computer Vision*, Venice, pp. 521-530, 2017.

[6] Devlin J., Chang M., Lee K., and Toutanova K., "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[7] Du Y., He M., Wang L., and Zhang H., "Wasserstein Based Transfer Network for Cross-Domain Sentiment Classification," *Knowledge-Based Systems*, vol. 204, pp. 106162-106178, 2020.

[8] Ganin Y., Ustinova E., Ajakan H., Germain P., Larochelle H., Laviolette F., Marchand M., and Lempitsky V., "Domain-Adversarial Training of Neural Networks," *The Journal of Machine Learning Research*, vol. 17, no.1, pp. 2096-2030, 2016.

[9] Jia X., Jin Y., Li N., Su X., Cardiff B., and Bhanu B., "Words Alignment Based on Association Rules for Cross-Domain Sentiment Classification," *Frontiers of Information Technology and Electronic Engineering*, vol. 19, no. 2, pp. 260-272, 2018.

[10] Jiang J. and Zhai C., "Instance Weighting for Domain Adaptation in NLP," *in Proceedings of the 45th Annual Meeting of The Association of Computational Linguistics*, Prague, pp. 264-271,

2007.

[11] Khondaker A., Khandaker A., and Uddin J., "Computer Vision-based Early Fire Detection Using Enhanced Chromatic Segmentation and Optical Flow Analysis Technique," *The International Arab Journal of Information Technology*, vol. 17, no. 6, pp. 947-953, 2020.

[12] Lan Z., Chen M., Goodman S., Gimpel K., Sharma P. and Soricut R., "Albert: A lite bert for Self-Supervised Learning of Language Representations," *arXiv preprint arXiv:1909.11942*, 2019.

[13] Li Y., Xiao F., Chen Z., and Li B., "Adaptive Active Learning for Semi-Supervised Learning," *Ruan Jian Xue Bao/Journal of Software*, vol. 31, no. 12, pp. 3808-3822, 2020.

[14] Liu B., "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1-167, 2012.

[15] Meng J., Dong Y., Long Y., and Zhao D., "An Attention Network Based on Feature Sequences for Cross-Domain Sentiment Classification," *Intelligent Data Analysis*, vol. 25, no. 3, pp. 627-640, 2021.

[16] Myagmar B., Li J., and Kimura S., "Cross-Domain Sentiment Classification with Bidirectional Contextualized Transformer Language Models," *IEEE Access*, vol. 7, pp. 163219-163230, 2019.

[17] Ostendorff M., Bourgonje P., Berger M., Moreno-Schneider J., Rehm G., and Gipp B., "Enriching Bert With Knowledge Graph Embeddings for Document Classification," *arXiv preprint arXiv:1909.08402*, 2019.

[18] Pan S., Ni X., Sun J., Yang Q., and Chen Z., "Cross-Domain Sentiment Classification Via Spectral Feature Alignment," *in Proceedings of the 19th International Conference on World Wide Web*, Raleigh, pp. 751-760, 2010.

[19] Qu X., Zou Z., Cheng Y., Yang Y., and Zhou P., "Adversarial Category Alignment Network for Cross-Domain Sentiment Classification," *in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, pp. 2496-2508, 2019.

[20] Ramshankar N. and Prathap P., "A Novel Recommendation System Enabled by Adaptive Fuzzy Aided Sentiment Classification for E-Commerce Sector Using Black Hole-Based Grey Wolf Optimization," *Sādhanā*, vol. 46, no. 3, pp. 1-24, 2021.

[21] Sener O., Song H., Saxena A., and Savarese S., "Learning Transferrable Representations for Unsupervised Domain Adaptation," *Advances in Neural Information Processing Systems*, vol. 29, pp. 2110-2118, 2016.

[22] Settles B., "Active Learning Literature Survey," Technical Report, University of Wisconsin, 2009.

[23] Sharma R., Bhattacharyya P., Dandapat S., and Bhatt H., "Identifying Transferable Information Across Domains for Cross-Domain Sentiment Classification," *in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, pp. 968-978, 2018.

[24] Tang D., Qin B., and Liu T., "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, pp. 1422-1432, 2015.

[25] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser Ł., and Polosukhin I., "Attention is All You Need," *in Advances in Neural Information Processing Systems*, Long Beach, pp. 5998-6008, 2017.

[26] Wang L., Niu J., Song H., and Atiquzzaman M., "Sentirelated: A Cross-Domain Sentiment Classification Algorithm for Short Texts through Sentiment Related Index," *Journal of Network and Computer Applications*, vol. 101, pp. 111-119, 2018.

[27] Wei X., Lin H., Yu Y., and Yang L., "Low-Resource Cross-Domain Product Review Sentiment Classification Based on A CNN with an Auxiliary Large-Scale Corpus," *Algorithms*, vol. 10, no.3, pp. 81, 2017.

[28] Wu F., Huang Y., and Yan J., "Active Sentiment Domain Adaptation," *in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, pp. 1701-1711, 2017.

[29] Yan X., Guo J., Lan Y., and Cheng X., "A Biterm Topic Model for Short Texts," *in Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, pp. 1445-1456, 2013.

[30] Yu J., Elkaref M., and Bohnet B., "Domain Adaptation for Dependency Parsing Via Self-Training," *in Proceedings of the 14th International Conference on Parsing Technologies*, Bilbao, pp. 1-10, 2015.

[31] Zadrozny B., "Learning and Evaluating Classifiers Under Sample Selection Bias," *in Proceedings of the 20st International Conference On Machine Learning*, Banff Alberta, pp. 114, 2004.

[32] Zhao C., Wang S., and Li D., "Multi-Source Domain Adaptation with Joint Learning for Cross-Domain Sentiment Classification," *Knowledge-Based Systems*, vol. 191, pp. 105254-1053.13, 2020.

[33] Zhao C., Wang S., Li D., Liu X., Yang X., and Liu J., "Cross-domain Sentiment Classification Via Parameter Transferring and Attention

Sharing Mechanism," *Information Sciences*, vol. 578, pp. 281-296, 2021.

Hongye Cao is currently pursuing the Master degree at School of Software, Northwestern Polytechnical University, China. His research interests are machine learning and natural language processing.

Qianru Wei received the Ph.D. degree in Computer Science and Technology from Xidian University in 2016. She is currently an assistant professor at the School of Software, Northwestern Polytechnical University, China. Her research interests include machine learning and natural language processing.

Jiangbin Zheng received the Ph.D. degree in Computer Science and Technology from Northwestern Polytechnical University in 2002. He is currently a professor at the School of Software, Northwestern Polytechnical University, China. His research interests include machine learning and natural language processing.