Advances in Artificial Intelligence and Machine Learning; Research 4 (3) 2468-2489 Received 27-06-2024; Accepted 13-08-2024; Published 20-08-2024

Improving Cross-Domain Aspect-Based Sentiment Analysis using Bert-BiLSTM Model and Dual Attention Mechanism

Yadi Xu yadixu@student.usm.my
School of Computer Science, Universiti Sains Malaysia
Penang, Malaysia

Noor Farizah Ibrahim nfarizah@usm.my
School of Computer Science, Universiti Sains Malaysia
Penang, Malaysia

Corresponding Author: Noor Farizah Ibrahim

Abstract

Data across different domains can be influenced by variations in language styles and expressions, making it challenging to migrate specialized words, particularly when focusing on aspectual words. This complexity poses difficulties in conducting cross-domain aspect-based sentiment analysis. The article begins by introducing BERT for generating word vectors as representations of training texts, enhancing text semantics in the word vector representation stage. To capture more nuanced interaction information and context-related details, the paper proposes the Bert-BiLSTM model with a dual attention mechanism(BB-DAM), which divides the original input sequence into three parts: above, aspectual words, and below. A dual attention mechanism was used to assess the interaction of aspect words with the three aspects (above, below, and neighboring words) in the three discourse segments. This mechanism allows for the comprehensive extraction of interaction information. By comparing with other modeling approaches, the experimental results show that the BB-DAM model produces good results in fine-grained cross-domain sentiment analysis.

Keywords: Aspect-based sentiment analysis, Cross-domain, BERT-BiLSTM, Dual Interaction Mechanism.

1. INTRODUCTION

With the major advances in information technology in China, the process of Internet usage is growing rapidly. According to the report [1], the country has achieved a remarkable 73.0% penetration rate of Internet access among the general population, with over one billion Internet users. The semantic information contained within this data possesses significant application value across various fields. Thus, sentiment analysis has emerged as a valuable tool [2–4]. Sentiment analysis involves the analysis of text-based data to infer the attitude or emotional polarity expressed by the text creator. In general, sentiment analysis can be classified into two main types based on the level of granularity of the target data object. Coarse-grained sentiment analysis is further divided into two types [5]:

Citation: Yadi Xu and Noor Farizah Ibrahim. Improving Cross-Domain Aspect-Based Sentiment Analysis using Bert-BiLSTM Model and Dual Attention Mechanism. Advances in Artificial Intelligence and Machine Learning. 2024;4(3):144.

text-level sentiment analysis and sentence-level sentiment analysis. Where text-level sentiment analysis assesses the overall sentiment polarity of the entire text, sentence-level sentiment analysis focuses on assessing the sentiment polarity of individual sentences. Fine-grained sentiment analysis, on the other hand, is analyzed at the aspectual level, with the aim of assigning specific sentiment polarity to each mentioned object of evaluation (called an aspectual word) in a sentence [6].

At present, scholars at home and abroad mainly use implicit Markov models based on sequence tagging, maximum entropy models, conditional random fields (CRF), etc., and combine various rhetorical devices and grammar rules to extract features manually and combine them with traditional machine learning models. For example, Kiritchenko et al [7], represent each sentence as a feature vector combining n-means (n-gram) features, lexical features, lexical labels, etc., and use an SVM classifier for sentiment classification. Another common approach to accomplish sentiment classification is to construct sentiment dictionaries [8]. Recurrent neural networks (RNN) are now widely used in the field of sentiment analysis because of their excellent performance on time series. RNN can not only consider the sequence of sequences, but also better mine the structural and logical information in the text, and can efficiently handle input data of different lengths by converting the input data into a fixed-length feature representation. Many scholars [9, 10], combined CRF with RNN to obtain a powerful feature representation by learning text features without manually designing a feature extractor but overcoming the limitations of language and grammar rules through techniques such as likelihood function, backpropagation mechanism, and character-level embedding. However, the literature [11], shows that the traditional RNN structure is prone to phenomena such as “gradient disappearance” and “gradient explosion”, leading to the poor effectiveness of RNN in dealing with long-term dependencies. To address this phenomenon, some scholars introduced RNN into the forgetting gate mechanism, while combining long-term memory and short-term memory to make RNN models improved. Li [12], constructed a new method using two LSTMs and achieved the association of feature words and emotion words through multi-task learning. More importantly, they introduced some sentences with strong emotions to help the extraction of other long-time memories. Ma et al [13], extended LSTM by computing sentiment common sense knowledge using a sequence encoder and integrated it into a deep neural sequence model to efficiently process multiple aspects of words present in a sentence. Based on the idea of a bidirectional gating mechanism, Luo [14], used Bert-BiLSTM to construct sequence-tagged syntactic dependency trees to obtain better aspectual word extraction. Pham et al [15], combined feedforward neural networks, word embedding techniques, and combined vector models to design a multilayer framework that combines word-level, sentence-level, aspectual word-level, and higher-level feature representations are fused to predict the sentiment of aspectual words. Meanwhile, many scholars have tried to introduce techniques such as image processing and computer vision into sentiment analysis.

Sentiment classification methods usually require extensive training, yet in real life, emerging domains often lack sufficient data. The differences between domains make the application of sentiment classification models in different domains limited. Therefore, cross-domain sentiment analysis is more challenging than within-domain sentiment analysis. A few researchers have considered how to find the dataset closest to the target domain from a dataset perspective and use that dataset as a training set for classifiers to compensate for the adverse effects of differences in data distribution across domains. For example, Ponomareva et al [16], evaluated the similarity of inter-domain datasets based on domain similarity and complexity and combined these two aspects to propose a linear regression model. However, due to the lack of high-quality databases, most of the existing

studies have focused on discovering commonalities between domains and achieving migration by analyzing these commonalities. In 2021, Khan et al [17], incorporated multiple source domains to learn sentiment classification models in their study and utilized a domain-independent generic sentiment dictionary based on cosine similarity to compute polarity scores, which enhanced the performance of sentiment classification while improving the classification effect; in the same year, Zhu et al [18], combined multi-source migration learning and domain adaptation to accurately extract domain-shared features in a common feature space by aligning source and target domain distributions to accurately extract domain-shared features; Dai et al [19], proposed two migration learning frameworks (WS-UDA and 2ST-UDA) based on multi-source domain adaptation methods and derived better target hypotheses by combining source hypotheses. And by 2022, the research fever has further increased, Yang et al. [20], proposed a Granger-causal goal (MDA-GC) based multi-source domain adaptation approach, which uses a kind of sentiment-guided capsule network for each source domain to build an expert model to capture the domain-invariant knowledge, and then designs the attention mechanism to assign importance weights to the experts, in which each expert specializes in a different source domain, which effectively improves the classification accuracy; while Zhao et al [21], proposed a framework with multi-source domain adaptation and joint learning for a multi-source cross-domain sentiment classification task and validated it on two datasets; in the same year, Cui et al [22], proposed a multi-source unsupervised domain adaptation method from the perspective of how to select appropriate source domains, which combines self-training with an attention mechanism together, showing a good competitive performance.

Transfer learning is critical in cross-domain sentiment analysis. Research conducted by [23], demonstrated that if two domains share the same feature space and probability distribution, inductive transfer learning can be employed to extract aspect words that are common to both domains. This facilitates the discovery of commonalities between the domains. At present, methods utilizing direct inductive transfer learning have shown promising applications in cross-domain sentiment analysis. For instance, Kong [24], proposed a multi-label classification algorithm based on direct push transfer learning, effectively addressing the classification problem of unlabeled data in the target domain. Marcacini [25], focused on aspect word extraction and utilized direct push transfer learning to integrate knowledge from different domains. They constructed a heterogeneous network to map diverse characteristics of tagged aspect words, employing principles such as convergence. In another study, Shu et al [26], introduced a CRF (Conditional Random Field) model based on a lifelong machine learning mechanism. This approach preserves prior knowledge acquired from previous domains and enhances the learning capability for future sentiment prediction.

In recent years, the scope of domains involved in online comments has become more and more extensive, and it is time-consuming and laborious to manually annotate the data of each domain in terms of data processing, and there may also be difficulties of insufficient comment data in a certain domain. In addition, considering that source and target domains often differ in terms of data distribution and feature space, it is easy to cause the same word to express different sentiments in different domains, while different domains contain their own domain-specific features, i.e., classifiers trained in source domains are often far from the expected results if they are directly used to process target domain classification tasks. Therefore, cross-domain sentiment classification is a research topic with important theoretical value and practical significance. In this paper, we take two perspectives of deep learning and attention mechanism as the entry point to carry out the research on the application of the BERT training model in fine-grained cross-domain sentiment analysis. We propose a hybrid approach of Bert-BiLSTM model and dual attention mechanism applied to fine-

grained cross-domain sentiment analysis, and compare and experimentally validate Bert-BiLSTM with two mainstream methods, BiLSTM and BERT-TextCNN, and use the experimental results to evaluate the new model proposed in this paper.

2. THEORETICAL FOUNDATIONS OF CROSS-DOMAIN FINE-GRAINED SENTIMENT ANALYSIS

2.1 Natural Language Processing (NLP)

NLP is an interdisciplinary field that combines computer science and artificial intelligence to explore the ways in which computers can comprehend, analyze, and generate human language. Words are the basic unit of meaning and the cornerstone of natural language. It is essential to represent words in a suitable manner and convert them into a computer-friendly data format. This serves as a crucial preliminary step for various tasks, and it is achieved through text representation. Research [27], has shown that a well-designed text representation can lead to improved mapping within the text space, subsequently enhancing the efficiency and effectiveness of subsequent algorithms. By employing effective text representation techniques, the underlying information and semantics of words can be effectively captured, enabling more accurate and meaningful analysis and processing of natural language data.

2.1.1 Word embedding

Word Embedding is a prevalent technique employed in natural language processing to convert discrete words into continuous vector representations [28, 29]. The fundamental concept behind word embedding is to encode the semantic and syntactic relationships between words by representing them as real-numbered vectors. This allows computers to gain a deeper understanding of text data and facilitates more effective processing. By utilizing word embedding, words with similar meanings or contexts tend to have vectors, enabling algorithms to capture semantic relationships and make more nuanced associations. Word embedding is important in various NLP tasks, including language modeling, information retrieval, sentiment analysis, and machine translation [30].

Word Embedding is the process of mapping words into a vector space. In traditional text processing, words are usually represented using monothermal encoding, and words are represented as a high-dimensional sparse vector with only a one-dimensional value of 1 and the rest having a value of 0. However, One-Hot Encoding fails to capture semantic and syntactic information and neglects the relationships and similarities between words [31–33].

In contrast, word embedding offers a more effective semantic representation by mapping words. This allows similar words to be closer together in the vector space. Each word is represented as a fixed-length vector of real numbers, where each dimension represents a semantic or syntactic feature. Through model training, word embeddings can capture various relationships between words, including proximity, contextual associations, and lexical properties [34, 35].

By leveraging word embeddings, NLP models can benefit from the ability to encode and understand the nuanced relationships between words, enabling more sophisticated language understanding and processing tasks.

2.1.2 Bert word vectors

Since 2018, Transformer-based pre-training models have been proposed successively and used for different downstream tasks. Bert models can capture deeper semantic information. Bert-based text classification models are composed of two parts [28, 29]. The pre-training is self-supervised, using a large unlabeled text corpus to complete the training, which can learn the text semantic features and deep text vector representation; the pre-tuning starts from the pre-training Bert model, and its fitting and convergence need to be completed according to the specific classification task [36–38].

2.2 Transfer Learning

Transfer learning is an approach utilized to overcome the scarcity of high-quality labeled data that matches a specific task in real-world scenarios. In the face of this challenge, there is a growing demand for cross-domain solutions. Transfer learning offers a quick and effective means to address this issue. When there are disparities in the data sample space or probability distribution, they are treated as distinct domains. The objective of transfer learning is to extract knowledge from the source domain or source task and transfer it to the target task. These different types of transfer learning methods are employed based on the specific differences observed in the migration context.

Migration learning is used to solve the problem of limited labeled samples and to effectively utilize the rich information in the source domain. In this process, it is crucial to select appropriate source data samples for training. A common research method is to assign weights to the source domain samples with the aim of making the sample distribution of the target domain as close as possible to the sample distribution of the source domain. This approach helps to construct highly accurate and reliable learning models, for example, it has been widely used in untargeted labeling, unsupervised domains, and so on [39–41].

However, traditional methods that focus on minimizing the divergence between probability distributions may not always be effective due to the disparities in data distribution between the source and target domains. In some cases, these methods can even lead to incorrect classification results and negative transfer effects. To overcome these challenges, new transfer learning methods have been proposed. For instance, transfer learning based on feature selection achieves knowledge transfer by identifying common features shared by the source and target domains. However, constructing a mapping relationship that maintains the consistency of source and target domain data in the new feature space remains a challenging task [42, 43].

Addressing this challenge requires ongoing research and innovation to develop robust transfer learning methods that can effectively capture and utilize the shared information between domains while adapting to the differences in data distributions. Model-based transfer learning, on the other hand, focuses on finding shared model parameters in the model space of source and target data to achieve knowledge reuse and transfer. In natural language processing, pre-training and fine-tuning methods

commonly used in sentiment analysis are model-based transfer learning methods. In addition, relational transfer learning focuses on how to efficiently analyze unrelated but homogeneous and heterogeneous data and achieve similar transfer through the association between them. a second classification approach to describe such relational transfer learning was proposed by Pan et al. (FIGURE 1). It should be noted that in practical applications, there may be some overlap between the two migration learning methods mentioned above. Since transfer learning is essentially a learning paradigm, different types of transfer learning methods can be used according to specific tasks without strictly following the above classifications.

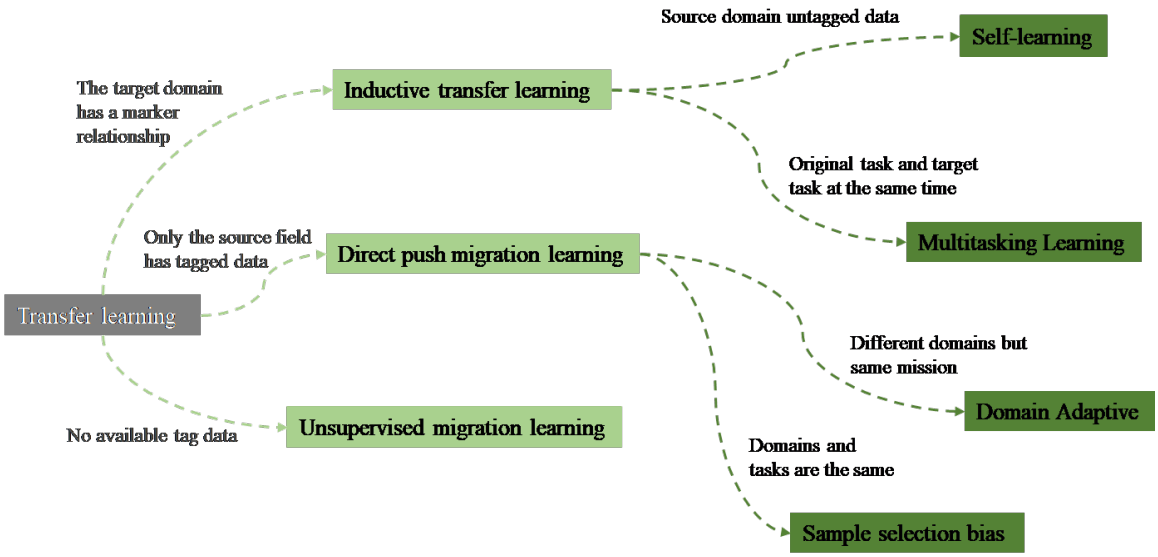

Figure 1: The transfer learning classification approach proposed by Pan

Transfer learning plays a crucial role in sentiment analysis and offers several advantages:

2.3 Recurrent Neural Networks

RNN is particularly suitable for processing sequential data, such as text data in natural language processing, time series data, etc. The unique feature of RNN is the introduction of recurrent connections that allow the network to maintain state or memory as it processes each sequence element and passes that state to the next element. This memory capability allows RNN to long-distance dependencies in sequences.

RNN have achieved remarkable results in language modeling, machine translation, text generation, and sentiment analysis. For fine-grained sentiment analysis tasks, RNN can effectively model text sequences and learn the sentiment information in the sequences. By processing each word or character in a sequence step by step, RNN can capture the contextual relationships between words to better understand and analyze sentiment. Although RNN excels in handling sequence data, it has some limitations such as difficulty in handling long sequences. To overcome these problems, researchers have proposed some improved RNN variants. The fusion approach used in this paper is feature hierarchical fusion, where the original word vector is first transformed into a

Table 1: The role of transfer learning in sentiment analysis

Advantages	Details
Leveraging existing knowledge	Transfer learning enables the utilization of knowledge acquired from one domain or task to enhance performance in a different domain or task. By leveraging pre-trained models or learned representations, sentiment analysis models can benefit from prior knowledge, reducing the need for extensive.
Addressing data scarcity	Labeled data may be limited or costly to obtain. Transfer learning allows the model to leverage labeled data from a related domain with abundant resources, enhancing models in the target domain with limited labeled data.
Improving generalization	Sentiment analysis models trained on a specific domain often struggle to generalize well to different domains due to variations in language styles, expressions, and sentiments. Transfer learning aids in improving the generalization capability of models by learning common features or representations that are relevant across multiple domains. This enables sentiment analysis models to capture domain-independent sentiment patterns.
Reducing training time and resources	By leveraging pre-trained models or learned representations, transfer learning can significantly reduce the training time resources required to develop effective sentiment analysis models. Instead of training from scratch, models can be fine-tuned on the target domain with a smaller dataset, leading to more efficient model development.
Handling domain shift	In sentiment analysis, the distribution of sentiments may vary across different domains. Transfer learning helps address this issue by aligning the sentiment distributions, enabling the model to adapt to the sentiment characteristics of the target domain more effectively.

high-dimensional feature representation using a neural network, and then the fusion is performed for the extracted high-dimensional features, and the overall process is shown in FIGURE 2.

3. METHOD: CROSS-DOMAIN ASPECT-BASED SENTIMENT ANALYSIS MODEL DESIGN

3.1 Bert Word Embedding

BERT is a bidirectional language model that uses the self-attentive mechanism in the Transformer structure and encoders to generate context-sensitive word vectors. Compared to traditional unidirectional language models, BERT has several significant features such as bidirectional modeling, Transformer structure, and large-scale pre-training.

BERT enables the model to consider both left and right information in the context by using the Transformer's self-attentive mechanism, and the model structure in FIGURE 3. BERT generates word vectors that incorporate the entire contextual information and more comprehensively under-

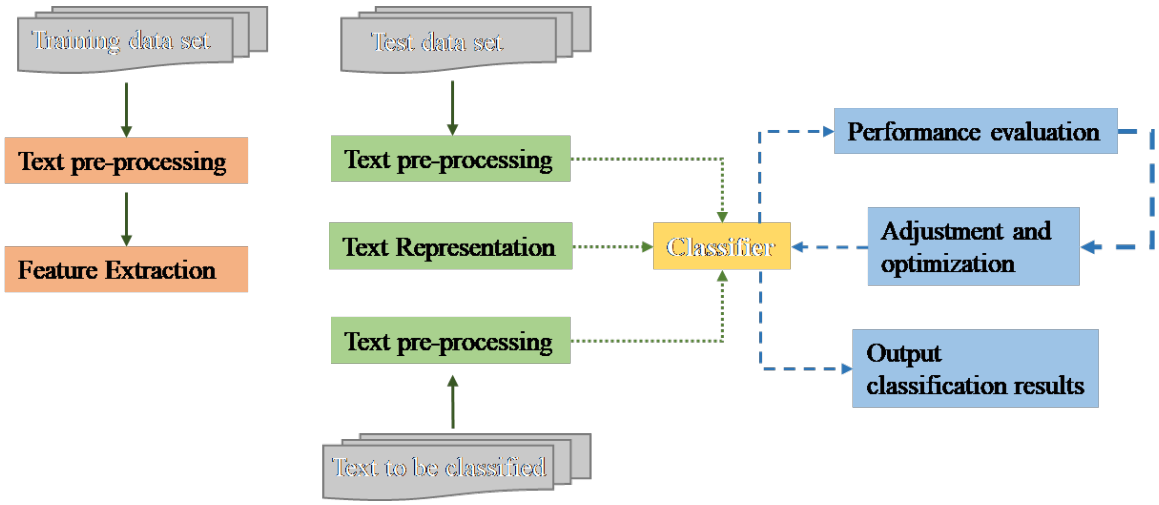

Figure 2: Text classification flow chart

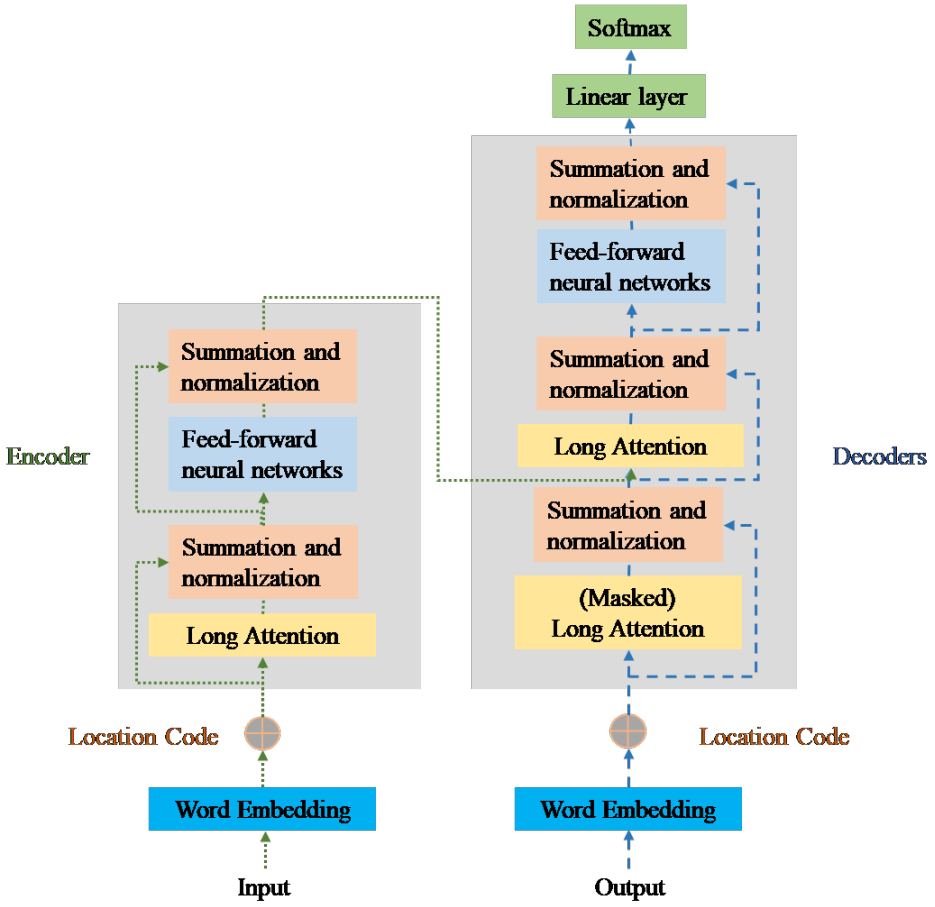

Figure 3: Transformer model structure

stand the context of each word in the whole sentence. BERT adopts Transformers as its underlying structure. The transformer is based on the self-attentive mechanism with high parallel computing power and long context modeling capability. By stacking multiple layers of Transformer, BERT is able to learn sentence representations and features at a deeper level. BERT is pre-trained on a large-scale corpus to learn generic language representations by unsupervised learning. Such a pre-training process enables BERT to capture a wide range of linguistic knowledge and contexts, thus providing robust feature representations for various downstream natural language processing tasks. After pre-training, BERT can be adapted to specific tasks by fine-tuning on task-specific labeled data. Through fine-tuning, BERT can transfer its general language representation capabilities to specific tasks and achieve better performance.

The introduction of BERT has had a significant impact on the field of natural language processing. It has results, including question and answer, text classification, and named entity recognition. Its bi-directional modeling and powerful feature extraction capabilities make BERT one of the most advanced pre-trained language models available.

The BERT model utilizes the encoder part of the Transformer architecture and consists of multiple stacked encoder layers. FIGURE 4 illustrates the structure of the BERT model. The input E_i to the model is the sum of three embedding vectors: Token Embeddings, Position Embeddings, and Segment Embeddings.

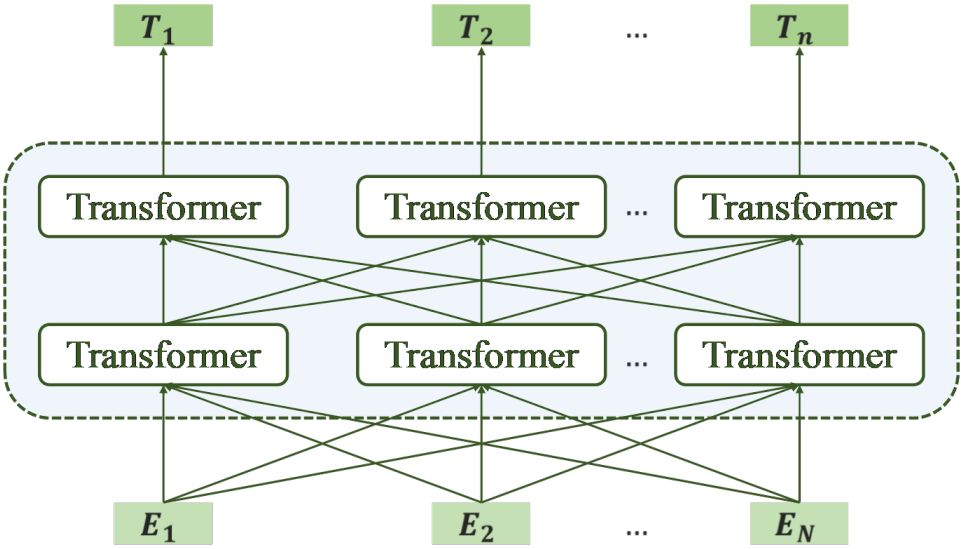

Figure 4: BERT model structure

During the pre-training phase, BERT employs two unsupervised tasks to learn generic language representations: Masked Language Model (MLM) and Next Sentence Prediction (NSP).In the Masked Language Model (MLM) task of BERT, the objective is for the model to learn to predict specific words in the input sequence. To accomplish this, approximately 15% of the words in the input sequence are randomly selected. Out of these selected words, around 80% are replaced with a special mask token, denoted as [Mask]. The model's task is to predict the original masked words based on the contextual information provided by the surrounding tokens. This process is akin to a completion or fill-in-the-blank task, where the model learns to understand and generate appropriate

words within the given context. By training on this MLM task, BERT can capture the contextual relationships between words and develop comprehensive language representations. In this way, BERT allows the model to have bidirectional semantic understanding without the need for labels, thus better understanding the context of the language. The NSP task aims to allow the model to determine whether two sentences are consecutive sentences in the original language. For each pair of input samples, BERT takes one of the sentences as the A sequence and the other as the B sequence and swaps the two sentences with a certain probability. Then, BERT learns to predict whether the B sequence is the next sentence of the A sequence. The NSP task helps BERT learn sentence-level semantic relations and reasoning ability to better handle tasks that require understanding sentence relatedness, such as question and answer systems.

With the training of these two tasks, BERT can learn bidirectional context-aware word vectors and be able to better understand and capture the semantic relationships between sentences and words. This pre-training approach gives BERT powerful language understanding and representation capabilities and provides useful features for downstream natural language processing tasks.

3.2 Bert-BiLSTM Model

LSTM (Long Short-Term Memory) is a specialized variant of recurrent neural networks (RNN) that effectively addresses the issue of long-term dependencies. It overcomes the vanishing gradient problem, which can hinder traditional RNN from capturing long-term dependencies in sequential data.

LSTM achieves this by incorporating three key components: forgetting gate. These gates control the flow of information within the LSTM unit, allowing it to selectively retain or discard information at different time steps. The structure of an LSTM unit is illustrated in FIGURE 5.

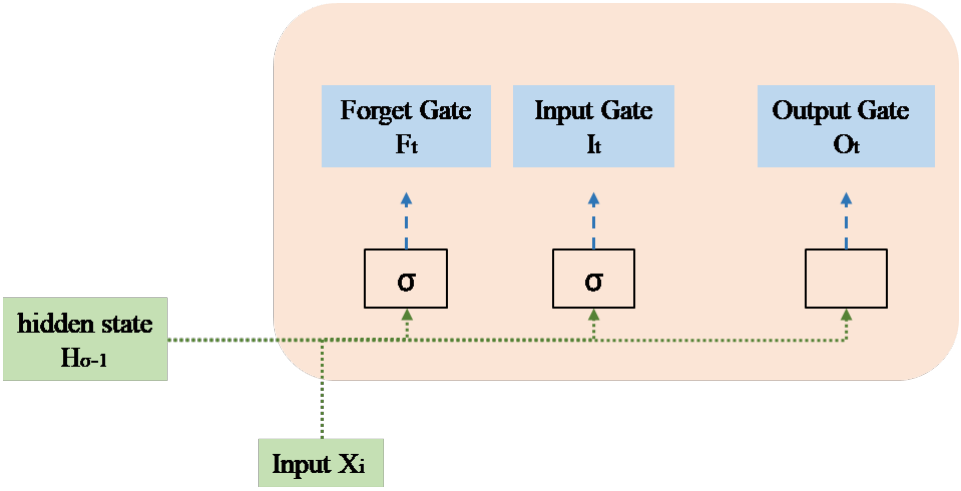

Figure 5: LSTM structure

The bi-directional recurrent network comprises two LSTM layers: one for forward and the other for reverse. This setup allows the network to consider both past and future information in the sequence.

While a unidirectional LSTM predicts the current output based on the preceding information, the bidirectional LSTM incorporates information from both the past and the future. Similar to LSTM, LSTM has gated states that enable it to capture longer-distance dependencies. The bidirectional LSTM model can express each character or word in the context of a sentence, taking into account the dependency relationships between them.

We employ the Bert-BiLSTM model to capture the contextual and semantic information of each word. The Bert-BiLSTM model integrates the power of BERT, a language representation model, with the bidirectional LSTM. By combining these two components, the model can capture semantics within each word. The structure of the Bert-BiLSTM model is depicted in FIGURE 6.

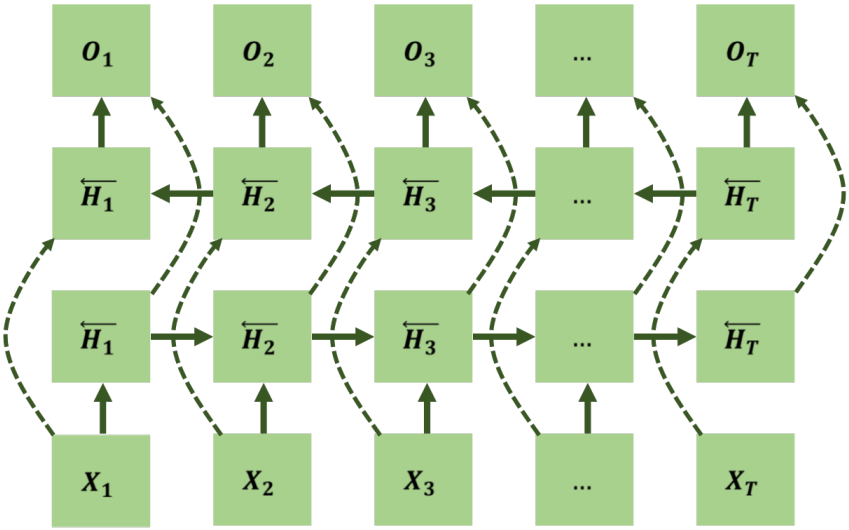

Figure 6: Bert-BiLSTM structure

3.3 Feature Extractor

(1) Text representation layer

Firstly, the input sequence is divided into words using a word separation step. Then BERT is used as the distributed representation of the text, and the generated low-dimensional dynamic word vectors can better represent the semantic information of the text. In this section, a pre-trained BERT model is used to generate corresponding word vector representations for each word element in the above, aspectual, and below parts, respectively. The upper text is assumed to be $x^l = [x_1^l, x_2^l, ..., x_L^l]$ and the aspect words are $x^a = [x_1^a, x_2^a, ..., x_A^a]$, and the following is $x^r = [x_1^r, x_2^r, ..., x_R^r]$. where L, A, and R represent the number of lexical elements contained in the above, aspectual words, and below, respectively. The BERT model will pass each word through the embedding layer, which generates three types of embedding vectors, namely word element embedding, sentence embedding, and location embedding, as shown in FIGURE 7.

The word element embedding layer has the same basic function as other word vectors, i.e., embedding representation for word elements. Besides, BERT specifically adds a sentence embedding layer and position embedding layer for enhanced representation. The representation after training by

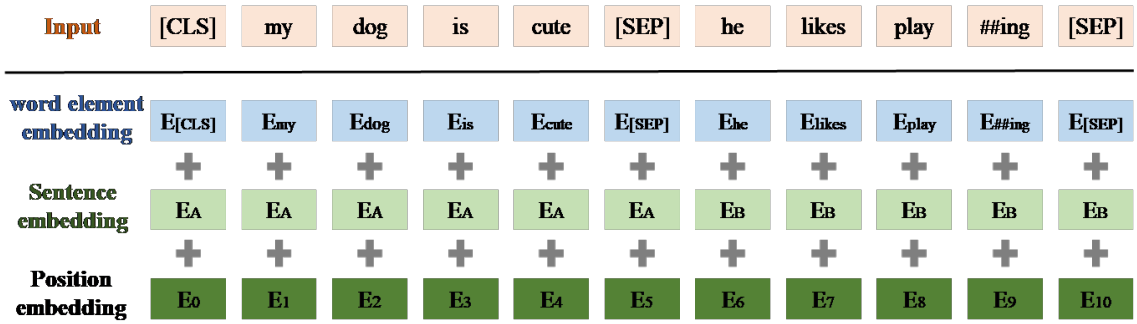

Figure 7: BERT embedding layer

BERT model is the dynamic representation of word elements, even if the same words are in the same input sequence, as long as they are in different positions, the corresponding word vectors generated will be completely different. [CLS] is a special token corresponding to the final hidden state that incorporates all the information and represents the total representation of this input sequence. The [SEP] token is used to distinguish multiple input sentences from each other. The corresponding elements of the three types of embedding vectors are summed and then passed to the coding layer of BERT as the input representation. Thus, after the BERT model is trained, the word vector representations corresponding to above, aspectual, and below $E^l = [e_1^l, e_2^l, ..., e_L^l] \in \Re^{l \times d}$, $E^a = [e_1^a, e_2^a, ..., e_L^a] \in \Re^{A \times d}$, $E^{lr} = [e_1^r, e_2^r, ..., e_L^r] \in \Re^{r \times d}$, where d represents the word vector dimension.

(2) Semantic feature extraction layer

For natural language, different words have different degrees of influence on affective polarity judgments depending on the position they are in. Most of the existing works consider the use of location coding or location weighting to enrich the embedding representation information of words. However, such approaches have certain shortcomings and ignore the important role played by dependencies in grammar rules for sentiment determination. In this paper, this problem is effectively improved by introducing Semantic-Relative Distance (SRD) to dynamically weight the words in context. The formula for SRD can be expressed as shown in Equation 1:

$$SRD_i = |i - p_a| - \left|\frac{m}{2}\right| \tag{1}$$

Where i denotes the computed word; P_a denotes the aspect word; and m denotes the sequence length of the aspect word.

At the same time, to attenuate the impact of non-sentiment keywords with large SRD, this paper uses dynamic weighting for them, and the weight w_i is calculated as shown in Equation 2:

$$w_i = \left(1 - \frac{SRD_i - \alpha}{N}\right) \cdot I \tag{2}$$

where I denotes a vector of all 1s with the same dimension as the input vector; α denotes the threshold; and N denotes the total length of the input vector corresponding to the original sequence.

The obtained weights are used to implement weighting operations on the above embedding vector E^l and the below embedding vector E^r, respectively, to obtain the new above vector representation

V^l and the below vector representation V^r. At this point, V^l and V^r have been highlighted for sentiment keywords, reducing the probability of misclassification of non-sentiment keywords for sentiment categories. Feature extraction using Bert-BiLSTM networks is performed for V^l, E^a and V^r to generate feature vectors that incorporate contextual information and are richer in semantic representation, respectively H^l, H^a and H^r. The weights are calculated as shown in Equation 3:

$$H^l = \left[\overrightarrow{h_l}, \overleftarrow{h_l}\right] = \left[\overrightarrow{LSTM}\left(V_l\right), \overleftarrow{LSTM}\left(V_l\right)\right] \tag{3}$$

For the hidden states H^l, H^a and H^r obtained through the Bert-BiLSTM network, the attention weights are calculated by the dual interactive attention mechanism to establish the correlation between the aspect words and the context as well as the aspect words.

First attention mechanism: Firstly, a representative matrix M is constructed, $M^l \in R^{L\times A}$, $M^r \in R^{R\times A}$ are set to represent the relevance of the above and below pairs of aspectual words, respectively, and activated with the ReLU function. The two computational equations are shown below:

$$M^l = ReLU(w_1 H^l H^{aT}) \tag{4}$$

$$M^r = ReLU(w_2 H^r H^{aT}) \tag{5}$$

Where W_1, W_2 represent the trainable parameter matrices. M^l_{ij} represents the correlation of the i lexical element above with the j lexical element of the aspect word, and M^r_{ij} represents the correlation of the i lexical element below with the j lexical element of the aspect word; then the correlation weights of the j lexical element of the aspect word for the i lexical element above and the i lexical element below are computed as follows. α^l_{ij}, α^r_{ij} Calculated formulas for the two are shown below:

$$\alpha^l_{ij} = \frac{\exp(M^l_{ij})}{\sum_{k=1}^{A}\exp(M^l_{ik})} \tag{6}$$

$$\alpha^r_{ij} = \frac{\exp(M^r_{ij})}{\sum_{k=1}^{A}\exp(M^r_{ik})} \tag{7}$$

Combining the above formulas to obtain the relevant representations V^l_i, V^r_i under the first weighted attention mechanism, the computational equations are shown below:

$$V^l_i = \sum_{k=1}^{A} \alpha^l_{ik} * H^l_{ik} * H^{aT}_{ik} \tag{8}$$

$$V^r_i = \sum_{k=1}^{A} \alpha^r_{ik} * H^r_{ik} * H^{aT}_{ik} \tag{9}$$

V^l_i represents the relevance representation of the word i above under the influence of the aspect word, V^r_i and the same is true.

The correlation weights α^{al}_{ij}, α^{ar}_{ij} of the i lexical element above to the j lexical element of the aspect word, and the i lexical element below to the j lexical element of the aspect word, are similarly obtained by following the above steps; and in turn, V^{al}_i, V^{ar}_i are obtained.

The second attention mechanism carries out further computation on the basis of V_i^l, V_i^r, V_i^{al} and V_i^{ar} obtained by the first attention mechanism to find out the attention weights of above and below for each lexical element within and the attention weights of aspect words for each lexical element therein, respectively, as shown in Equation 10 to Equation 12:

$$\alpha_i^l = \frac{\exp(V_i^l)}{\sum_{k=1}^{L} \exp(V_k^l)} \tag{10}$$

$$\alpha_i^r = \frac{\exp(V_i^r)}{\sum_{k=1}^{R} \exp(V_k^r)} \tag{11}$$

$$\alpha_j^a = \frac{\exp(V_j^{al})}{\sum_{k=1}^{A} \exp(V_k^{al})} + \frac{\exp(V_j^{ar})}{\sum_{k=1}^{A} \exp(V_k^{ar})} \tag{12}$$

Above is the computation process of dual-interactive attentional weights. Finally, X^l, X^a and X^r are stitched together to obtain the final feature representation $X = [X^l; X^a; X^r]$. The resulting feature representations of the source-domain samples and target-domain samples are obtained.

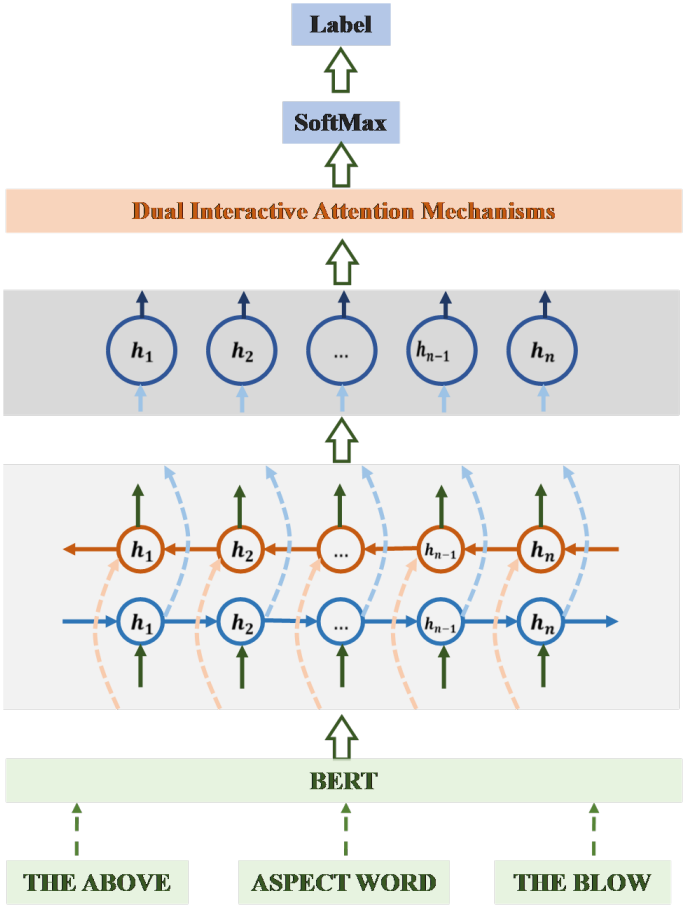

Figure 8: The Proposed General framework of the Bert-BiLSTM with Dual Attention Mechanism model

The structural diagram of the BB-DAM model is shown in FIGURE 8. This model uses the character vector generated by Bert as the character embedding layer in the upstream part, models Bert-BiLSTM as the feature extractor in the downstream part, and uses Dropout to reduce the risk of overfitting, and finally inputs the Softmax function to predict text classification.

The product and operation of the attention weights and the hidden state vectors obtained by BiLSTM are performed to obtain the weighted above representation V_l, aspect word representation V_a and below representation V_r. The specific calculations are shown below, respectively:

$$V_l = \sum_{i=1}^{L} \alpha_i^l \cdot H_i^l \tag{13}$$

$$V_a = \sum_{i=1}^{A} \alpha_i^a \cdot H_i^a \tag{14}$$

$$V_r = \sum_{i=1}^{R} \alpha_i^r \cdot H_i^r \tag{15}$$

The three feature representations are spliced to obtain the final feature vector V for classification, as shown in Equation 16.

$$V = [V_l; V_a; V_r] \tag{16}$$

Finally, the feature vector V is passed into the classification network to complete the sentiment classification, i.e., a fully connected layer is connected to map the feature vector V to the target space and the predicted probability distribution is computed by the softmax function as shown in Equation 17.

$$p(y) = soft\max(W \cdot V + b) \tag{17}$$

Where W and b represent the weight matrix and bias term, respectively.

The loss function is shown in equation 18.

$$Loss = -\sum_{i=1}^{C} y_i \log p(y_i) + \lambda ||\theta||^2 \tag{18}$$

Where C denotes the number of categories of sentiment, θ denotes the deflation coefficient of the L_2 regular term, and θ is the parameter matrix.

Above is the computation process of dual-interactive attentional weights. Finally, X^l, X^a and X^r are stitched together to obtain the final feature representation $X = [X^l; X^a; X^r]$. The resulting feature representations of the source-domain samples and target-domain samples are obtained.

The structural diagram of the BB-DAM model is shown in Fig. 8. This model uses the character vector generated by Bert as the character embedding layer in the upstream part, models Bert-BiLSTM as the feature extractor in the downstream part uses Dropout to reduce the risk of overfitting, and finally inputs the Softmax function to predict text classification.

4. EXPERIMENTAL DATA ANALYSIS

4.1 Experimental Dataset and Task Construction

To identify the effectiveness of the migration learning model, this chapter utilizes publicly available datasets from four distinct domains: Restaurant (R), Laptop (L), Device (D), and Service (S). The experimental datasets comprise different sources of data related to each domain. TABLE 2 presents the datasets.

Table 2: Statistical information on datasets

Fields	Source	Number of training samples	Number of test samples	Total number of samples
R	SemEval2014 [44] 2015 [45] 2016 [46]	2276	1306	3682
L	SemEval2014 [44]	1358	392	1750
D	Hu et al. [47]	905	451	1356
S	Toprak et al. [48]	1294	704	1998

By combining data from different domains two by two, 12 cross-domain fine-grained sentiment classification tasks can be constructed using R→L, R→D, R→S, L→R, L→D, L→S, D→R, D→L, D→S, S→R, S→L, S→D the 12 groups (Transfer Pair) are represented. Since the D-domain dataset contains review information of five electronic digital devices, the similarity with the L-domain dataset is relatively high. Therefore, this chapter follows the setting of Li et al [49], to delete L→D and D→L migration pairs, and finally, a total of 10 sets of migration pairs are formed. In each migration pair, the source domain is represented by the arrow tail, while the target domain is indicated by the arrow pointing towards it.

4.2 Experimental Setup

The experiments in this chapter use Micro-F1 as the index, and only when the aspect words are extracted correctly and the corresponding sentiment polarity is judged correctly, i.e., the sequence annotation is the same as the real sample, it can represent the case of correct prediction. the calculation procedure of Micro-F1 is shown in Equation 19 to Equation 21.

$$\Pr ecision_{micro} = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FP_i} \tag{19}$$

$$Recall_{micro} = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FP_i} \tag{20}$$

$$Micro - F1 = \frac{2 * \Pr ecision_{micro} * Recall_{micro}}{\Pr ecision_{micro} + Recall_{micro}} \tag{21}$$

4.3 Comparison of the Experimental Setup

To evaluate the effectiveness of the proposed model for fine-grained cross-domain sentiment analysis, several widely used models for cross-domain sentiment analysis are selected for comparison as shown in TABLE 3. The selected models are as follows:

Table 3: Cross-domain aspect-based sentiment analysis performance of different models on benchmark datasets (%)

Migration	BB-DAM	AHF	Bert	AD-SAL
R→L	40.71	34.89	35.07	34.13
R→D	37.94	37.33	32.08	35.44
R→S	30.90	33.23	22.92	28.01
L→R	38.43	43.49	30.98	43.04
L→S	35.29	33.05	31.63	28.01
D→R	43.51	44.57	40.15	41.01
D→S	35.05	34.96	31.27	26.62
S→R	51.28	46.55	54.93	41.03
S→L	40.93	29.01	36.96	27.04
S→D	40.05	39.61	38.17	33.56
Average	39.41	37.67	35.42	33.71

- Bert-BiLSTM-DAM(BB-DAM): Word vectors are trained by Bert and input as a word embedding layer into the Bert-BiLSTM layer for feature extraction.
- AHF [50]: An adaptive hybrid framework that integrates pseudo-label-based training and adversarial training into a unified network.
- Bert [51]: Pre-trained models for direct tuning in downstream tasks.
- AD-SAL [49]: The Model for dynamically learning word-to-word alignments through adversarial training Adversarial training.

BB-DAM is the best and AD-SAL is the worst because the sample of the fine-grained sentiment analysis dataset is characterized by more non-aspect words than aspect words, which would be poor if all labeled symbols were treated equally, thus illustrating the importance of confidence threshold fine-tuning according to the number of iterations and label set. The average Micro-F1 value of AD-SAL is about 17% lower than that of BB-DAM, which indicates that the Bert model gets performance improvement because it makes full use of the labeled data.BB-DAM can effectively extract the interaction between aspectual words and context and the semantic information of word granularity, and then complete the task of fine-grained sentiment analysis. The importance of the sentiment keywords is emphasized by dividing the original sequence into the above, the aspectual words, and the below, and then coping with the problem of multiple meanings of a word by using the dynamic word vectors with the dynamic weighting based on the relative distances of the semantics, and the S-double interactive attentional mechanism enables the semantic information to be sufficiently extracted, which helps judge the polarity of the sentiments.

BB-DAM is the best and BiLSTM is the worst because the sample of the fine-grained sentiment analysis dataset is characterized by more non-aspect words than aspect words, which would be poor if all labeled symbols were treated equally, thus illustrating the importance of confidence threshold fine-tuning according to the number of iterations and label set. The average Micro-F1 value of BiLSTM is about 15% lower than that of BB-DAM, which indicates that the Bert model gets performance improvement because it makes full use of the labeled data.

4.4 Sample Analysis

To further illustrate that the BB-DAM model has good transferability, this paper selects some data in the R domain from the L→R migration pair task as a sample to verify the sequence labeling capability of the Bert-BiLSTM model, as shown in TABLE 4. To facilitate reading and understanding, the aspect words and their corresponding sentiments are marked in red in the table, "None" indicates that the prediction marker value is empty, and the "✓" symbol and "✗" symbol are used to indicate whether the prediction is correct or not. From TABLE 4, we can see that BB-DAM is the best prediction and AD-SAL has the worst prediction performance. The reason for the incorrect prediction of "waiters" by AD-SAL may be that the model's memory of domain-specific features is confused due to adversarial training, resulting in "negative migration". In the case of sentences containing multiple aspect words and the aspect words are relatively rare domain-specific words, the domain differences are too large for all three to implement correct annotation.

Table 4: Sample analysis in the R domain

Migration	BB-DAM	AHF	AD-SAL
The [*sushi*]*pos* is cut in blocks bigger than my cell phone.	[*sushi*]*pos* **(✓)**	[*sushi*]*pos* **(✓)**	None **(✗)**
The [*service*]*pos* was exceptional - sometime there was a feeling that we were served by the army of friendly [*waiters*]*pos*	[*service*]*pos* **(✓)**	[*service*]*neg* **(✓)**	[*service*]*pos* **(✓)**
	[*waiters*]*pos* **(✓)**	[*waiters*]*pos* **(✗)**	None **(✗)**
My friend enjoyed the [*grilled Alaskan King Salmon*] *pos* with delectable [*creamed Washington russet potatoes*]*pos* and crisp [*green beans*]*pos*	None **(✗)**	None **(✗)**	None **(✗)**
	None **(✗)**	None **(✗)**	None **(✗)**
	None **(✗)**	None **(✗)**	None **(✗)**

5. CONCLUSION

Our study employs a hybrid approach of the Bert-BiLSTM model and dual attention mechanism to enhance cross-domain aspect-based sentiment analysis. In this paper, we first introduce the BB-DAM model in the word vector representation stage to generate word vectors as the word representation of the training text for text semantic enhancement, which can effectively extract the interaction relationship between aspect words and context and the semantic information of word

granularity, and then complete the task of fine-grained sentiment analysis. The importance of sentiment keywords is emphasized by dividing the original sequence into three parts: above, aspectual and below, and then using dynamic word vectors with dynamic weighting based on semantic relative distance to cope with the problem of word polysemy. Finally, we compare and experimentally validate BB-DAM with three methods, namely, AHF, AD-SAL, and Bert, and the experimental results show that the Bert-BiLSTM model and the dual-attention mechanism proposed in this paper have good results in fine-grained cross-domain sentiment analysis and can be vigorously promoted in the future.

References

[1] China Internet Network Information Center (CNNIC). The 49th Statistical Report on the Current Status of Internet Development in China [R]; 2022.

[2] Wankhade M, Rao AC, Kulkarni C. A Survey on Sentiment Analysis Methods, Applications, and Challenges. Artif Intell Rev. 2022;55:5731-5780.

[3] D'Aniello G, Gaeta M, La Rocca I. Knowmis-Absa: An Overview and a Reference Model for Applications of Sentiment Analysis and Aspect-Based Sentiment Analysis. Artif Intell Rev. 2022;55:5543-5574.

[4] Zhang W, Li X, Deng Y, Bing L, Lam W. Towards Generative Aspect-Based Sentiment Analysis. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; 2021:504-510.

[5] Guo X, Yu W, Wang X. An Overview on Fine-Grained Text Sentiment Analysis: Survey and Challenges. J Phys Conf S. 2021;1757:012038.

[6] Xiao Y, Li C, Thürer M, Liu Y, Qu T. User Preference Mining Based on Fine-Grained Sentiment Analysis. J Retailing Con Serv. 2022;68:103013.

[7] Kiritchenko S, Zhu X, Cherry C, Mohammad S. Nrc-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. Proceedings of the 8th International Workshop on Semantic Evaluation; 2014:437-442.

[8] Wagner J, Arora P, Cortes S, et al. Dcu: Aspect-Based Polarity Classification for Semeval Task 4. In: Proceedings of the 8th International Workshop on Semantic Evaluation; 2014:223-229.

[9] Wang W, Pan SJ, Dahlmeier D, Xiao X. Recursive Neural Conditional Random Fields for Aspect-Based Sentiment Analysis. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018:616-626.

[10] Yin Y, Wei F, Dong L, et al. Unsupervised Word and Dependency Path Embeddings for Aspect Term Extraction. Proceedings of the 25th International Joint Conference on Artificial Intelligence; 2016:2979-2985.

[11] Bengio Y, Simard P, Frasconi P. Learning Long-Term Dependencies With Gradient Descent Is Difficult. IEEE Trans Neural Netw. 1994;5:157-166.

[12] Li X, Lam W. Deep Multi-Task Learning for Aspect Term Extraction With Memory Interaction. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; 2017: 2886-2892.

[13] Ma Y, Peng H, Khan T, Cambria E, Hussain A. Sentic Lstm: A Hybrid Network for Targeted Aspect-Based Sentiment Analysis. Cognit Comput. 2018;10:639-650.

[14] Luo H, Li T, Liu B, Wang B, Unger H. Improving Aspect Term Extraction With Bidirectional Dependency Tree Representation. IEEE ACM Trans Aud Speech Lang Process. 2019;27:1201-1212.

[15] Pham DH, Le AC. Learning Multiple Layers of Knowledge Representation for Aspect Based Sentiment Analysis. Data Knowl Eng. 2018;114:26-39.

[16] Ponomareva N, Thelwall M. Biographies or Blenders: Which Resource Is Best for Cross-Domain Sentiment Analysis?. International Conference on Intelligent Text Processing and Computational Linguistics; 2012:488-499.

[17] Khan FH, Qamar U, Bashir S. Enhanced Cross-Domain Sentiment Classification Utilizing a Multi-Source Transfer Learning Approach. Soft Comput. 2019;23:5431-5442.

[18] Zhu Y, Zhuang F, Wang D. Aligning Domain-Specific Distribution and Classifier for Cross-Domain Classification From Multiple Sources. Proceedings of the 2021 Conference on Association for the Advancement of Artificial Intelligence (AAAI). 2021;33:5989-5996

[19] Dai Y, Liu J, Ren X, et al. Adversarial training based multi-source unsupervised domain adaptation for sentiment analysis[C]//Proceedings of the 2021 Conference on Association for the Advancement of Artificial Intelligence (AAAI). 2021;34:7618-7625.

[20] Yang M, Shen Y, Chen X, Li C. Multi-Source Domain Adaptation for Sentiment Classification With Granger Causal Inference. Proceedings of the 2022 Conference on Special Interest Group on Information Retrieval (SIGIR). 2020:1913-1916.

[21] Zhao C, Wang S, Li D. Multi-Source Domain Adaptation With Joint Learning for Cross-Domain Sentiment Classification. Knowl Based Syst. 2020;191:105254.

[22] Cui X, Bollegala D. Multi-Source Attention for Unsupervised Domain Adaptation. Proceedings of the 2022 Conference of Asia-Pacific Chapter of the Association for Computational Linguistics and the 2022 International Joint Conference on Natural Language Processing (AACL/IJCNLP); 2022:873-883. .

[23] Rana TA, Cheah YN. A Two-Fold Rule-Based Model for Aspect Extraction. Expert Syst Appl. 2017;89:273-285.

[24] Kong X, Ng MK, Zhou ZH. Transductive Multilabel Learning via Label Set Propagation. IEEE Trans Knowl Data Eng. 2011;25:704-719.

[25] Marcacini RM, Rossi RG, Matsuno IP, Rezende SO. Cross-Domain Aspect Extraction for Sentiment Analysis: A Transductive Learning Approach. Decis Support Syst. 2018;114:70-80.

[26] Shu L, Xu H, Liu B. Lifelong Learning Crf for Supervised Aspect Extraction. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017:148-154.

[27] Ali I, Melton A. Graph-Based Semantic Learning, Representation and Growth From Text: A Systematic Review 13th International Conference on Semantic Computing. Vol. 2019. IEEE PUBLICATIONS. INTERNATIONAL COUNCIL OF SHOPPING CENTERS; 2019:118-123.

[28] Wang Y, Wang W, Chen Q, Huang K, Nguyen A, et al. Fusing External Knowledge Resources for Natural Language Understanding Techniques: A Survey. Inf Fusion. 2023;92:190-204.

[29] Qiu S, Liu Q, Zhou S, Huang W. Adversarial Attack and Defense Technologies in Natural Language Processing: A Survey. Neurocomputing. 2022;492:278-307.

[30] Donnelly LF, Grzeszczuk R, Guimaraes CV. Use of Natural Language Processing (Nlp) in Evaluation of Radiology Reports: An Update on Applications and Technology Advances. Semin Ultrasound CT MR. 2022;43:176-181.

[31] He T, Boudewyn MA, Kiat JE, Sagae K, Luck SJ. Neural Correlates of Word Representation Vectors in Natural Language Processing Models: Evidence From Representational Similarity Analysis of Event□Related Brain Potentials. Psychophysiology. 2022;59:e13976.

[32] Pruneski JA, Pareek A, Nwachukwu BU, Martin RK, Kelly BT, et al. Natural Language Processing: Using Artificial Intelligence to Understand Human Language in Orthopedics. Knee Surg Sports Traumatol Arthrosc. 2023;31:1203-1211.

[33] Tabassum N, Menon S, Jastrzębska A. Time-Series Classification With Safe: Simple and Fast Segmented Word Embedding-Based Neural Time Series Classifier. Inf Process Manag. 2022;59:103044.

[34] Raaijmakers S. Deep Learning for Natural Language Processing. Simon Schuster; 2022.Widmann T. Widmann T, Wich M. Creating and Comparing Dictionary, Word Embedding, and Transformer-Based Models to Measure Discrete Emotions in German Political Text. Pol Anal. 2023;31:626-641.

[35] Feghali J, Jimenez AE, Schilling AT, Azad TD et al. Overview of Algorithms for Natural Language Processing and Time Series Analyses. Machine Learning in Clinical Neuroscience: Foundations and Applications. Springer International Publishing; 2022:221-242.

[36] Han X, Zhang Z, Ding N, Gu Y, Liu X, et al. Pre-trained Models: Past, Present and Future. AI Open. 2021;2:225-250.

[37] Wan Y, Zhao W, Zhang H, Sui Y, Xu G, et al. What Do They Capture? A Structural Analysis of Pretrained Language Models for Source Code. In: Proceedings of the 44th International Conference on Software Engineering; 2022:2377-2388.

[38] Mozafari M, Farahbakhsh R, Crespi N. A Bert-Based Transfer Learning Approach for Hate Speech Detection in Online Social Media. In: Complex Networks and Their Applications VIII. Proceedings of the Eighth International Conference on Complex Networks and Their Applications Complex Networks 2019.Springer International Publishing; 2020:928-940.

[39] Zaib M, Sheng QZ, Emma Zhang WE. A Short Survey of Pretrained Language Models for Conversational AI – A New Age in Nlp. In: Proceedings of the Australasian Computer Science Week Multiconference; 2020:1-4.

[40] Gong Z, Zhou K, Zhao X, Sha J, Wang S, et al. Continual Pre-training of Language Models for Math Problem Understanding With Syntax-Aware Memory Network. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics 2022;1:5923-5933.

[41] Kouw WM, Loog M. A Review of Domain Adaptation Without Target Labels. IEEE Trans Pattern Anal Mach Intell. 2019;43:766-785.

[42] Ganin Y, Lempitsky V. Unsupervised Domain Adaptation by Backpropagation. In: International Conference on Machine Learning, PMLR; 2015:1180-1189.

[43] Tian L, Tang Y, Hu L, Ren Z, Zhang W. Domain Adaptation by Class Centroid Matching and Local Manifold Self-Learning. IEEE Trans Image Process. 2020;29:9703-9718.

[44] Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, et al. Semeval-2014 Task 4: Aspect Based Sentiment Analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014:27-35.

[45] Pontiki M, Galanis D, Papageorgiou H, Manandhar S, Androutsopoulos I. Semeval-2015 Task 12: Aspect Based Sentiment Analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); 2015:486-495.

[46] Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, AL-Smadi M, et al. Semeval-2016 Task 5: Aspect Based Sentiment Analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016); 2016:19-30.

[47] Hu M, Liu B. Mining and Summarizing Customer Reviews. In: Proceedings of the 10th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining; 2004: 168-177.

[48] Toprak C, Jakob N, Gurevych I. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. 2010:575-584.

[49] Li Z, Li X, Wei Y, Bing L, Zhang Y, et al. Transferable End-To-End Aspect-Based Sentiment Analysis With Selective Adversarial Learning. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 2019:4589-4599.

[50] Zhou Y, Zhu F, Song P, Han J, Guo T, et al. An Adaptive Hybrid Framework for Cross-Domain Aspect-Based Sentiment Analysis. AAAI. 2021;35:14630-14637.

[51] Kenton JD, Toutanova LK. Bert: Pretraining of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of The NAACL-HLT. 2019:4171-4186.