IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 7 April 2023, accepted 18 May 2023, date of publication 24 May 2023, date of current version 14 June 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3279396

RESEARCH ARTICLE

Aspect-Context Level Information Extraction via Transformer Based Interactive Attention Mechanism for Sentiment Classification

SANIA NAYAB, MUHAMMAD KASHIF HANIF, RAMZAN TALIB, AND MUHAMMAD UMER SARWAR
Department of Computer Science, Government College University, Faisalabad 38000, Pakistan
Corresponding author: Muhammad Kashif Hanif (mkashifhanif@gcuf.edu.pk)

ABSTRACT Aspect-context sentiment classification aims to classify the sentiments about an aspect that corresponds to its context. Typically, machine learning models considers the aspect and context separately. They do not execute the aspect and context in parallel. To model the contexts and aspects separately, most of the methods with attention mechanisms typically employ the Long Short Term Memory network approach. Attention mechanisms, on the other hand, take this into account and compute the parallel sequencing of the aspects-context. The interactive attention mechanism extracts features of a specific aspect regarding its context in the sequence, which means aspects are considered when generating context sequence representations. However, when determining the relationship between words in a sentence, the interactive attention mechanism does not consider semantic dependency information. Moreover, the attention mechanisms did not capture the polysemous words. Normally conventional embedding models, such as GloVe word vectors, have been used. In this study, transformers are embedded into the attention mechanism approaches to overcome the semantic relationship problem. For this reason, the BERT pre-train language model is used to capture the relationship among the words in a sentence. The interactive attention mechanism is then applied to the model's distribution of that word. The final sequence-to-sequence representation in terms of context and aspect is used into general machine learning classifiers for aspect-level sentiment classification. The proposed model was evaluated on the two datasets, i.e., Restaurant and Laptop review. The proposed approach has state-of-the-art results with all attention mechanisms and attained significantly better performance than the existing ones.

INDEX TERMS Aspect-context level sentiment classification, transformer based interactive attention mechanism, BERT, BERT interactive attention representation, aspect-context feature extracted data, machine learning classifiers.

I. INTRODUCTION

Opinions and reviews by gazillion of people are expressed on products, offers, services, etc., [1] via online platforms such as social networks, blogs, wikis, and discussion chatbots. With the arising of social networks, specifically social media, public reviews are constantly generated about something (like a particular political situation, a public person, a product, or a movie). These reviews are generated worldwide and are easily accessible via the web. The public follows the opinions on social media, blogs, tweets, YouTube, and many companies evaluate product reviews, preferences, or client satisfaction via these sources. So naturally, the extraction of these types of information, i.e., expressed opinions, has to be significant for marketing, businesses, professionals, and researchers [2], [3].

Society has long been involved in public opinion. Sentiment Analysis (SA) [4], [5], also called Opinion Mining, is a great challenge [6], [7], [8] and it is the most active area for Research in Natural Language Processing (NLP) and

The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu.

Deep Learning (DL)/Machine Learning (ML) [9]. It collects personal information from text such as evaluations, sentiments, appraisals, opinions, semantics relationships, and attitudes [2], [10]. Classifying text into negative, positive, and, sometimes, neutral is a major interest for research purposes. However, sometimes we do multi-classification according to the given scenarios, and such classification tasks are known as polarity classification problems. This research focuses on the sentiment polarity discovery related to its aspects and context by getting the contextualized embedding.

The aspect-level sentiment classification targets the reviewer's sentiment learning of a product or a service based on a said review phrase. It's more comprehensive subtask of the sentiment analysis [9]. Precisely, aspect-level sentiment classification is supplementary fine-grained than document-level sentiment classification [4], [5]. In many real-world applications, it is more valuable because of its aspect-based sentiment polarity, specifically focused on a specific aspect. Usually, an opinion contains many aspects with relevant sentiments present in complex contexts [11]. 40% of the prediction errors in sentiment's polarity are given by Jiang et al. [12], and these are originated when the opinion targets (aspects) are not reflected in their relevant context. That is why the main challenge in aspect-level sentiment classification is how to get specific aspect-level sentiment polarity concerning its context when there is a mixture of multiple aspects and context-level sentiments.

Attention-based Long Short Term Memory (LSTM) network has been stated good results, and these methods' results depend on the attention mechanism with the specified aspects in its context [13], [14], [15], [16]. However, attention-based LSTM fails to capture a word close to its particular aspect with relvant context. Such as typical attention-based LSTM network, i.e., Interactive Attention Network (IAN), it employs hidden states for aspects and context separately [15]. Attention mechanism networks with LSTM are time-consuming as LSTM takes more time to train a lengthy text data set. Aspect Context Interactive Attention (ACIA) network proposed by [11] leverages all word pairs between context and aspects. This approach enables the network to better understand final-representations by getting aspects and context relations in fine-grained word-level sentiment polarity. To state the major research challenge of aspect-level sentiment polarity by only the network architecture containing the attention mechanism is proposed by Vaswani et al. [17] and Wu et al. [11]; they use the same approach along with interactive attention mechanisms.

It is applicable when forming the aspect phrases, i.e., an aspect with multiple words, to get shared class of multiple aspect reviews phrases. However, with attention mechanism and typical word embedding, modeling representations regarding its aspects-contexts skip the correspondence among words in the similar sentence. The existing models have not considered the semantic relationship among the words in a phrase to get the context with its aspects. It also used the existing word embedding models, which do not capture polysemous expressions, i.e., the disambiguation of words with multiple meanings [11], [17], [18]. So these models also do not solve the problem of polysemy in context. Therefore, in response to these problems, attention mechanisms merged with Bidirectional Encoder Representation from Transformers (BERT) pre-trained language models to capture the relationship among the words to improve the ability to have aspect-level sentiment classification. Moreover, aspect-context extracted data frame never used in general ML classifiers for evaluation [19], [20], [21]. The three classifiers, namely Support Vector Machine (SVM), Naive Bayes (NB) and Random Forest (RF), were deployed for extended evaluation on extracted aspect-context features [21], [22].

This research proposed a novel pre-trained interactive-attention-based modeling method for extracting aspect-context-level information for positive, neutral and negative sentiment classification purposes. Comprehensively, our foremost contributions are encapsulated as follows:

- The BERT embedding is utilized here to get the contextualized words and word relationships from sentences. Then, BERT-MultiACIA can capture the contexts by considering specific words from a sentence linked with its relevant aspect. Here, this task is performed parallel so that training time can be reduced.
- The proposed BERT-based interactive attention mechanism can give the final representation for sentiment classification, which contains core aspect phrases from the sentences. These aspect phrases were retrieved by considering the relevant context in the previous feature extraction step.
- Our model results reveal that BERT-MultiACIA experimentation on the Restaurant data set is slightly improved and the Laptop data set experiment is also giving competitive results as compared to the state-of-art model. Moreover, SVM performed significantly better among ML classifiers on both data sets with aspect-context feature representations than the RF and NB.

This article proposed an aspects-context level information extraction using the BERT-ACIA mechanism, which is further used for sentiment classification via ML classifiers. We initiate by defining the relevant background information in Section II along with all sub-topics of the proposed methodology. Section III gives us the overall picture of a few related, existing models for aspect-context level sentiment classification. In addition, Section IV is explaining overview of the whole methods of our study in detail. We show the results of B-ACIA along with ML classification outcomes and interpretation of comparison with other state-of-the-art models in Section V. Finally, we have concluded our study in Section VI with some limitations and future research ideas.

II. BACKGROUND

A. SENTIMENT ANALYSIS

Sentiment analysis is another term used to look at users' expressions and analyze their emotional tendency from many subjective reviews over the Internet [3], [23], [24]. It also refers to analyzing public opinions, attitudes, and appraisals towards any entity or product or any challenging discussion on social network platforms [5], [25]. Emotions aka sentiments are purely dependent on the individual's experience regarding that specific product or event etc. Moreover, at some point, sentiments can also be defined as some attachment to that thing/event or reflects a memory, which is associated when we are giving some opinion about that thing [26], [27], [28]. The same scenario considers when opinions can be expressed. For example, the following expression could be judged in many ways.

This advanced laptop is bad!

The obvious meaning from the review is that the user dislikes the advanced laptop. Moreover, sentiment analysis also looks into the depth to check its aspect and context by considering the other parameters such as the time (for instance, the previous model of the laptop is working well till this advanced laptop came into the market. or at the review moment, the user is not in a mood to comment about this laptop.) at which the review has been expressed, the surroundings in which he is currently and expressing his opinions, anything regarding its features or something about its working. However, as mentioned earlier, the statement does not express the sentiment analysis according to some specific context and aspects regarding that advanced laptop, and it shows only the negative polarity in general [29]. Here the polarity means in which context the review has been added against that laptop. It purely based on the other parameters such as the word 'bad' used in the negative context so this review has negative polarity aka opinion.

B. ASPECT-CONTEXT LEVEL SENTIMENT ANALYSIS

Despite considering the opinion mining into a single classification problem, aspect-level sentiment analysis focuses on getting the reviewers' statements' tendency which consists of a series of tasks, and this mechanism is referred to as fine-grained presentation of the expression [11], [30], [31]. For instance, figure 1 gives the detail of the given expression according to its specific terms to catch the aspects and context. Underline colors shows the polarities (opinions) according to the context (the other parameters which are affecting the main theme of the review), such as a blue line for neutral, a red line for negative and a green line for positive polarity. Overall, it shows a positive review of the laptop. However, evaluating its depth by exploring its terms and their relevant polarities provides the Aspect-Context level sentiment analysis from the targeted sentence.

C. ATTENTION MECHANISM

Attention mechanism emerged as an enhancement over NLP's encoder-decoder-based translations system neural machine. It is a core part of a neural architecture that allows focusing related features of the input dataset asynchronously, and in NLP, it gathers a sequence of textual data elements [32] traditionally. Interactive attention captures how the considered word relates to the other words in the said sequence. This mechanism can work in both cases, i.e., on the raw input or to its higher level final representation [33].

D. BERT

Bidirectional Encoder Representations from the transformer (BERT) [34] is a state-of-the-art language representation model. There are two stages in the framework of BERT (i.e., pre-training and fine-tuning). As the name implies, the model is trained on different pre-training tasks in the first pre-training stage. The data used to train the model was not labeled. In the second fine-tuning stage, the model initialization is done with the pre-trained parameters. The fine-tuning of all the parameters is conducted by utilizing the labeled data from downstream tasks. Every task has distinct fine-tuned models, although they were initialized with similar parameters. The architecture of the BERT language model is unified in different working problems.

It is a bidirectional transformer encoder that has multiple layers. The number of transformer blocks (layers) is denoted by the letter L, the hidden size is represented by H, and the self-attention head count is denoted by the letter C. The two models are $BERT_{BASE}$ and $BERT_{LARGE}$. The $BERT_{BASE}$ has 12 layers (L), a hidden size (H) of 768, an attention head (C) size is 12, and the total parameters are 110 million. The $BERT_{LARGE}$ has 24 layers or transformer blocks; the hidden size is 1024, 16 self-attention heads, and 340 million parameters.

III. REVIEW OF LITERATURE

During the last years, neural network-based models have been used for aspect levels sentiment classification, such as Recurrent Neural Network (RNN), recursive neural network, gated neural network, and Convolutional Neural Network (CNN) [35], [36], [37]. Hybrid neural networks, including the Bi-Directional LSTM (Bi-LSTM) and CNN, are proposed by [38] for event detection for multiple languages. A Gated Convolutional network is presented with aspect embedding to get the Aspect Based Sentiment Analysis (ABSA). Here, ABSA is divided into two further subtasks. One is Aspect Category Sentiment Analysis (ACSA), and the other is Aspect Term Sentiment Analysis (ATSA). This model comprises two approaches: CNN and a gating mechanism [39].

With the help of attention-based network models, calculated weights can be assigned to separate words by directing the importance of the significant parts of the sentence. So these networks can easily be used to enhance the efficiency of aspect-level sentiment classification [15], [40]. The sentence-level sentiment analysis using word embedding (WEMB) proposed by Hayashi et al. [41] is a machine learning-based approach to sentiment analysis considering word importance in a sentence [16]. LSTM and the attention mechanism concentrated on the words of sentiment polarity

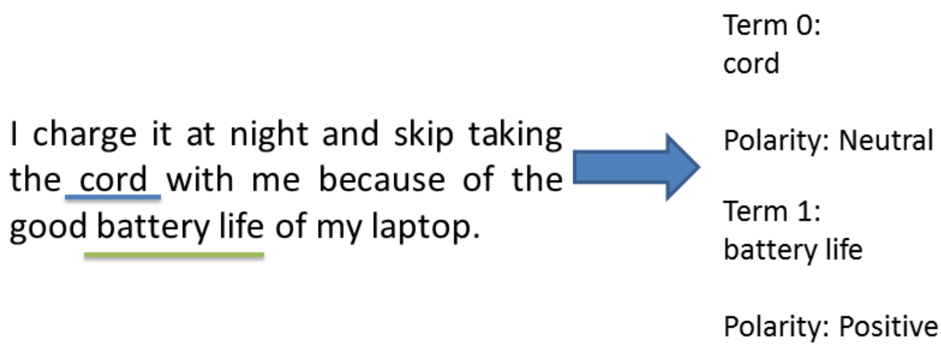


FIGURE 1. A Laptop review example. Aspect-context based polarities provides deep insight into the sentiment analysis. We have underlined the words according to their polarity such as Blue underline shows the neutral polarity, Green color shows positive polarity and Red color shows the negative polarity.

similar to their existing context. However, getting aspect-level sentiment classification based only on some phrase embedding stitching is inefficient.

Interactive attention mechanisms are proposed by Ma et al. and Huang et al. [15], [42] used to have the aspect-level sentiment classification with regards to its context. These interactive mechanisms using LSTM can generate sentiment polarities by generating aspect-based context information and context-based aspect representations. Multi-layer Aspect-Context Interactive Attention (MultiACIA) model was proposed to classify the sentiments with both aspect-context [11]. It relies on the interactive attention mechanism to produce the sequence representations of views in both context and aspects categories. It showed whether, compared with LSTM and LSTM with other attention mechanisms, this method is more appropriate for the aspect level sentiment classifications than the LSTM baseline method. But, the order information of aspects and context sequencing is ignored here. Without capturing the relationship among the words, i.e., semantic dependency in a sentence, this context-aspect-based approach is insufficient to improve the aspect-level sentiment classification accuracy. The pre-trained language models can effectively combine with the Interactive attention mechanisms to improve the accuracy of aspect-level sentiment classification.

To transfer the knowledge, He et al. [43] and Chen et al. [44] used pre-training and capsule networks, respectively. Other pre-trained language models signified by the BERT have been shown in the mentioned studies [23], [34], [45]. On many NLP tasks, these models are constantly getting state-of-the-art results. On aspect-level sentiment classification, various models along with the BERT obtained strong results [14], [23], [46], [47], [48]. The following table 1 provides a comprehensive summary of the related state-of-the-art models and we have followed the summarize literature review strategy of the authors of [49].

The relationship extraction model via the BERT [56], a model to have the word-relation tasks for the sentiment classification. This study focused on the effectiveness of the BERT model and interactive attention mechanisms to have the relationship information among the words in a sentence. Hence, it is natural to extend the BERT for word relation tasks. With the help of self-attention transformers in the BERT architecture, it is easy to perform a words-relation task for the aspect-level sentiment classification.

IV. METHODOLOGY

This section presents the BERT transformer based Aspect-Context Interactive Attention (B-ACIA) mechanism to get users' opinions towards the aspect-context level. We proposed an architecture based on the combination of BERT and deep learning approaches (i.e., attention mechanisms). The features were extracted using deep learning and then machine learning based classifiers were also used to classify the sentiments for further comparison. Figure 2 demonstrates the main steps of the proposed methodology workflow.

A. B-ACIA

We used BERT language model to get the word embedding. Each word was preprocessed and fed into $BERT_{BASE}$ to get the real-valued vector

$$b_w \in R^{d_w \times |V|} \tag{1}$$

where $|V|$ is the size of vocabulary and d_w is the embedding dimension and a matrix of pairs of keys and values with respective relationships. The classification problems that includes the aspect level of sentiments, the conventional approach is to handle particular aspects as queries. The values and the keys are mapped from the context. The architecture of ACIA [11] is based on calculating the interactive attention between contexts and aspects. It works by assigning weights to words in the context and aspect at the same time. In addition to this, the contextualized embedding are added and also the model has an integration structure so there is no need of two distinct attention mechanisms as it is in IAN (interactive attention network) [15]. The main architecture B-ACIA is following the BERT embedding outputs in the first phase of methodology. So this research proposed the hybrid model BERT-ACIA or B-ACIA and its extraction used for further contributions.

TABLE 1. Overview of transformer models with attention mechanism approaches used for aspect-context sentiment analysis.

Proposed Model	Advantages	Disadvantages	Datasets)	Implications	Ref
BERT based architecture for linear classification	1) Superiority of BERT on capturing aspect-based sentiment	1) Parameters of BERT component are fixed 2) issue of context independence	Originate SemEval datasets: LAPTOP REST (restaurant)	Need the task specific fine-tuning of BERT	[50]
BERT as representation focused model (RepBERT)	high accuracy for semantic level text retrieval with fixed length embeddings	conflict in the distribution of test and training data, inconsistency	MS MARCO passage ranking	plan to test model's generalization on different datasets with long text retrievals	[51]
Semantic-Relative Distance (SRD)-Local Context Focus (LCF) mechanism	the LCF-BERT model refreshes state-of-the-art performance	just focused local information	the laptop and restaurant datasets of SemEval-2014 and the ACL twitter dataset	SRD calculation can be further improved by considering extra auxiliary information the transfer-ability of CDM and CDW designs	[52]
Target-Specific Transformation Networks (TNet)	achieves a new state-of-the-art performance	didn't considered the relationship among the words in a sentence	three benchmark datasets: LAPTOP, TWITTER and REST are from SemEval ABSA challenge	semantic level relationship needs to be focused	[53]
interact intra- and inter- segment information from long text using BERT multi-task learning approach (LordBERT)	Reduces computational complexity; improved accuracy by 0.4% and 0.5% on both datasets	due to small segments, it loses semantic integrity	20NewsGroup, IMDB and ohsumed	to design adaptive segment ordering according to the different tasks	[54]
empirical study to evaluate embedding encoders and pre-trained BERT	improved results on different datasets and competitive on others found gap for further investigations	evaluated on specific tasks	different datasets	N/A	[55]

BERT used the positional embeddings in its structure, which considers the words before the verb (verb means the main word used to express the sentiment/review) or after the verb. In this way, we can have the polarity (negative, positive or neutral) by considering the context or aspect of the review.

Figure 3 shows the proposed architecture of B-ACIA.

$$S = [b_{w_1}, b_{w_2}, \cdots, b_{w_n}] \tag{2}$$

In the equation 2, S represents the Sentence.

$$A_i = [b_{w_j}, \cdots, b_{(w_{j+m-1})}] \tag{3}$$

In the equation 3, A represents Aspects.

In order to calculate the attention, the Sentence S and Aspect A is mapped by utilizing the fully connected layer. It will produce the pairs of keys and values. Those pairs can be represented as $< key_S, value_S >$ and $< key_A, value_A >$.

The parameters for mapping are such as W and b are weights and biases and ReLU is the activation function. After calculating the keys and values, the dot product of key_S and key_A is computed. It will produce pair-wise matching matrix product which can be seen in the figure 3.

All the entries in product depicts the relevance of words between context and aspect. these values are converted into probabilities by using the softmax function. The probabilities

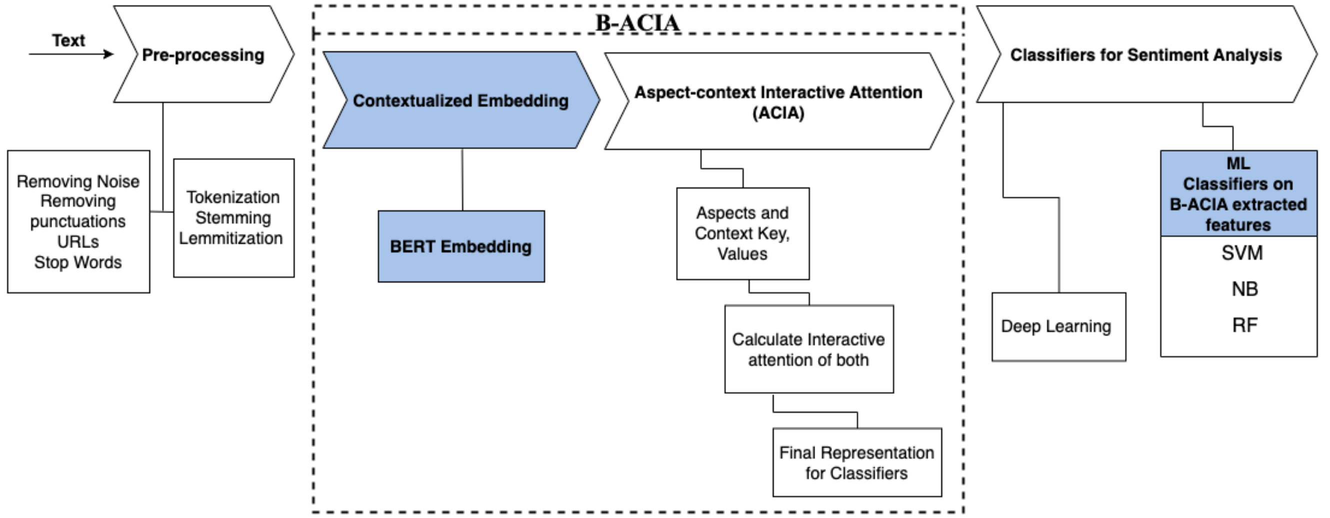


FIGURE 2. Overview of the proposed methodology.

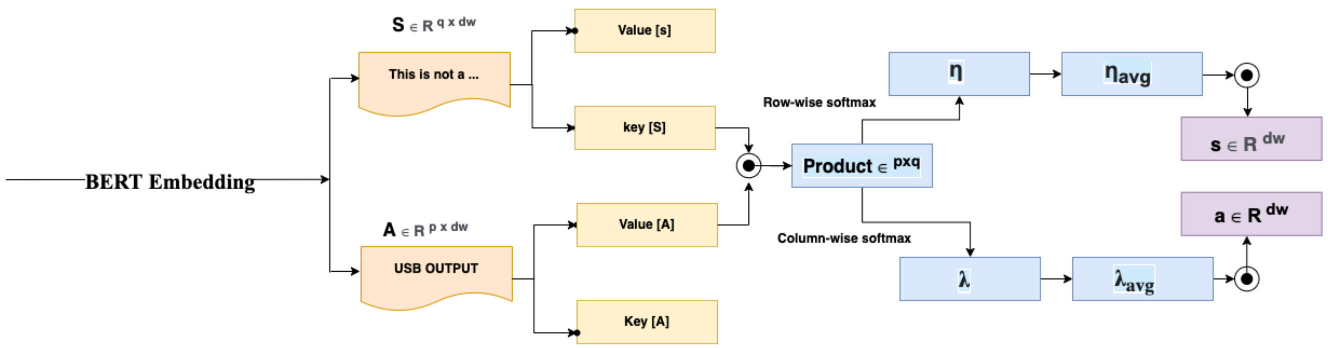


FIGURE 3. Proposed B-ACIA architecture.

range from zero to one. The large value of probabilities indicates the stronger relation. The softmax function can be represented in the 4 and 5:

$$\gamma_{ij} = \frac{\exp\left(\text{product}_{ij}\right)}{\sum_{i=1}^{m} \exp\left(\text{ product }_{ij}\right)} \tag{4}$$

$$\eta_{ij} = \frac{\exp\left(\text{product}_{ij}\right)}{\sum_{j=1}^{n} \exp\left(\text{product}_{ij}\right)} \tag{5}$$

Here γ denotes the column-wise softmax of the correlations of contexts and aspects. And η represents the correlations based on row-wise softmax.

The final weights of all the aspects are obtained by taking the row-wise average of γ. Similarly, the final weights of the sentences are determined by taking the column-wise average of η.

$\gamma_{avg} \in R^{m\times 1}$ and $\eta_{avg} \in R^{n}$ are the final weights. After that the weighted sum of $values_S$ and $value_A$ is computed to get the weighted representation of aspects and contexts. Aspects a after calculating γ_{avg} multiply with $value_A$ and contexts s attained after η_{avg} multiply with $values_S$.

The multiple groups of pairs of $< keys_S, values_S >$ and $< key_A, value_A >$ were generated for the operations of B-ACIA. It is based on the concept of Multi-head attention in B-ACIA, called Multi-BACIA. The aspects-contexts were projected via various groups of fully connected layers. So, various sets of $< key_S^L$, value $_S^L >$ and $< key_A^L$, value $_A^L >$ were generated. After generating multiple sets, the operation of B-ACIA is performed in parallel. With this, multiple outputs are produced which are concatenated to generate the final vector.

Numerous blocks of B-ACIA were stacked to produce the sequence-to-sequence representation of contexts and aspects. The residual connection and fully connected layers were implemented between two blocks of B-ACIA. The training parameters doesn't increase by the residual connections but they serve the purpose of solving the degradation problem. the degradation problem arises when the number of layers in the network increases.

The normalization layer is also applied to enhance the training speed and also it improves the ability to generalize. In between the two structures of B-ACIA, a feed-forward network which is fully connected is implemented for mapping of representations of aspects and contexts. The feed-forward network contains two fully connected layers with the activation function between them.

The activation function ReLU is used in the layers. The output S_{next} and A_{next} are used as input to the next structure. Finally the sequences representations are averaged and concatenated to form the final vector. This vector is used as the input to the machine learning based classifiers.

V. RESULTS AND DISCUSSION

We used two datasets to experiment with the proposed method. These datasets are already annotated by their relavant aspect and context terms. The authors of [57] manually annotated the datasets and we just used this annotated dataset for our experiments. The Restaurant dataset and Laptop dataset from Sam Eval task 4 [57] were used for experimentation purposes. These datasets contain reviews of users. Moreover, the reviews are annotated with aspects. The Sam Eval task 4 is already annotated by using the existing corpus, no separate or additional dictionary is defined for the aspects and context polarities. Although, authors manually annotated the datasets according to its aspects and context terms. The distribution of sentiment polarity categories for the two datasets is shown in Table 2.

TABLE 2. Description of two datasets.

Datasets		Positive	Negative	Neutral
Restaurant	Train	2164	807	637
	Test	728	196	196
Laptop	Train	994	870	464
	Test	341	128	169

After applying the basic pre-processing steps in our experiments, we employed the transformer-based word embedding BERT [34]language model to get the word embedding. Each word was pre-processed and fed into $BERT_{BASE}$ to get the real-valued vector $b_w \in R^{d_w \times |V|}$ where $|V|$ is the size of vocabulary and d_w is the embedding dimension. The embedding dimension in $BERT_{BASE}$ was 768. Aspects and contexts of interactive attention representations were formulated using a heap of four layers, and the model parameters were set to random initialization.

All the experiments are done using NVIDIA Quardo RTX 8000 GPUs. The default batch size is used (i.e., 32). While training all iterations complete their whole epochs and there is not any stop condition in this phase. Other experimental settings are in the following table 3:

TABLE 3. Experimental settings for BACIA.

pre-trained model	BERT
context-based dimensions	768
optimizer	Adam
max len	512
pooling layer approach	mean/max
traing task	classification

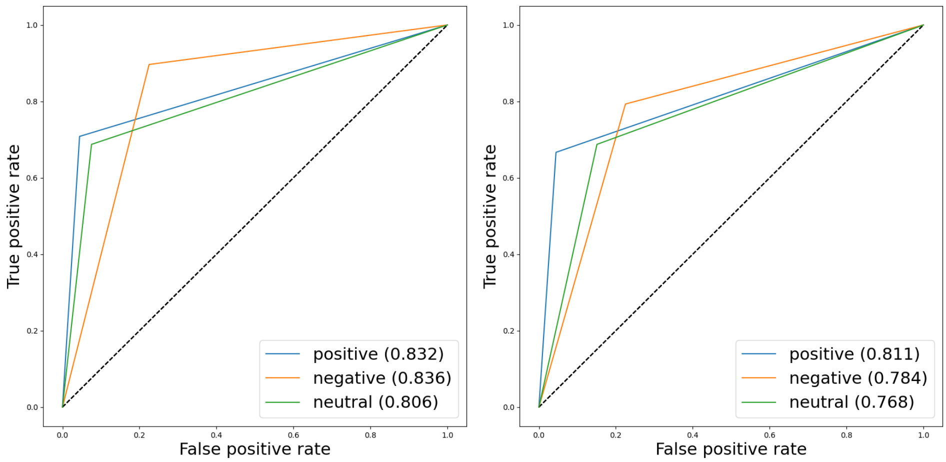
The BERT embedding was used to get the contextualized meanings of words from the sentences. Here, BERT also captures the polysemous words and provides the semantic relationships among words in a sentence. The proposed model was trained on both data sets by deploying pre-trained based MultiACIA, and in the last layer, DL classifiers are used using Adam optimizer with a learning rate set to 0.001. The first experiment was done on the Restaurant data set, and the acquired accuracy, i.e., 0.78, decreased by 2%. However, macro-F1 (average F1 score of all three classes) on all the classes of this data set is significantly improved by almost 1%. Afterward, the model is run on the Laptop data set, and it can interpret that accuracy is competitive with the existing model by a little increment of 0.4%. 3% significantly improves the overall F1 of this data set on all three classes. Our proposed model impacts the overall performance of all three classes of sentiment polarities, and the analysis of the following results table 2 also interprets this impact.

Figure 4 depicts the comparisons of positive, negative and neutral sentiment polarity, True Positive Rate (TPR) with False Negative Rate (FNR) on both data sets by using AUC-ROC curves. It can be seen in the figure 4a that negative polarities have higher TPR than the other two polarities for the Restaurant data set. Among all three classes of sentiment classification, the negative class is strongly classified by 83.6%. However, the positive class is also significantly competitive, and its TPR is at 83.2%. However, neutral is not classified as strongly as a positive and negative class, and 80.6% classifies its TPR. In the figure 4b for the Laptop data set, positive polarities have higher TPR than the other two polarities. Among all three classes of sentiment classification, the positive class is strongly classified by 81.1%. However, the negative and neutral classes are classified close to each other, just by the difference of 1.6% (Negative TPR = 78.4% and Neutral TPR = 76.8%).

TABLE 4. State-of-the-art Models' accuracy and F1 score Comparisons. All the baseline results are retrieved from the original papers. The actual multiB-ACIA and BERT-MultiACIA model results are marked with "$*$". It shows that BERT-MultiACIA is competitively at the same level on both datasets. However, macro F1 is significantly increased. In bold, ML general classifiers results are mentioned here; among them, SVM (marked with $\star$) performed significantly better.

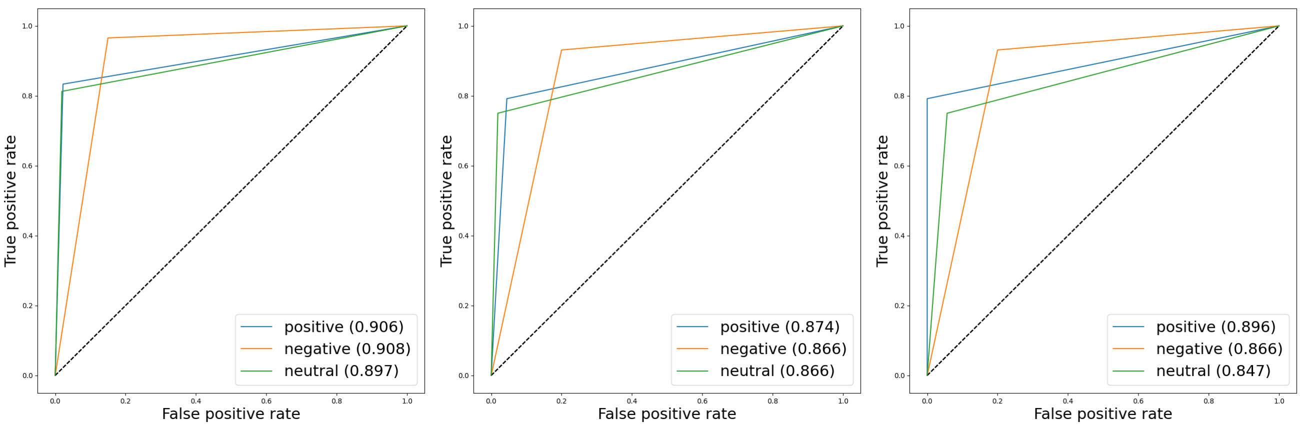
Models	Restaurant		Laptop	
	Accuracy	macro-F1	Accuracy	macro-F1
LSTM [13]	0.7430	0.6413	0.6650	0.6398
MultiACIA* [11]	0.7991	0.6987	0.7246	0.6800
IAN [15]	0.7860	0.6912	0.7210	0.6714
IAN-MultiACIA [42]	0.8044	0.7141	0.7324	0.6845
AOA-LSTM [35]	0.8120	0.7105	0.7405	0.6812
AOA-MultiACIA [58]	0.8259	0.7213	0.7527	0.7024
BERT-MutliACIA*	0.7826	0.7709	0.7245	0.7148
RF-MultiACIA	0.8405	0.8338	0.8115	0.8105
NB-MultiACIA	0.8405	0.8385	0.8260	0.8209
SVM-MultiACIA $\star$	**0.8840**	**0.8814**	**0.8550**	**0.8534**

To make our aspect-context level feature extraction outcomes more concise, we used machine learning-based classifiers for the classification. The extracted features by deploying the deep learning-based model, i.e., BACIA, were



(a) Restaurant dataset (b) Laptop dataset

FIGURE 4. DL classifier TPR and FNR results on both datasets.



(a) SVM-ML Classifier (b) NB-ML Classifier (c) RF-ML Classifier

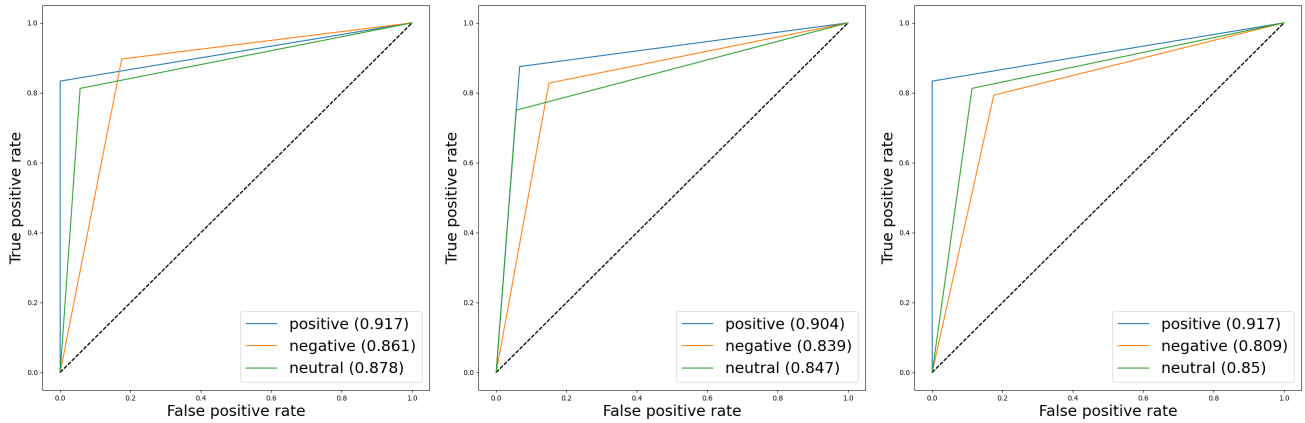
FIGURE 5. SVM, NB and RF classifier TPR and FNR results on Restaurant datasets.

extracted and then general ML-based classifiers were used for the further classification outcomes comparisons. The existing ML classifiers are used with all the default hyper-parameter settings.

It turned out that ML-based classifiers outperformed. The three classifiers, namely Support Vector Machine (SVM), Naive Bayes (NB) and Random Forest (RF), were deployed [22]. All three models performed better than the deep learning-based classifier. Out of the three models, the SVM classifier significantly outperformed by providing approximately 88% and 86% accurate results on Restaurant and Laptop datasets, respectively. Moreover, both datasets' three classifiers' results are mentioned in the Table 4. Their TPR and FNR classification are also discussed below with relevant AUC-ROC outcomes.

So far, SVM proved to be the state-of-the-art classification model when applied on the aspect-context level feature extracted data frame. by giving more than 90% TPR of classification on all three classes. Figure 5a shows that negative polarities of the Restaurant data set are classified more accurately than the other two polarities. However, the positive class is at the competitive level just by the difference of 0.2%. The neutral class of the said data set is at 89.7%, so it can be seen that SVM is giving significant results in all three classes. Although on the Laptop dataset, only positive class classification results are improved by 91.7% other two classes are at the same level and do not give the improved TPR of classification as compared to the Restaurant dataset TPR of classification.

NB classifier achieved the highest TPR of positive classification, and it is 89.6%. Although, the other two, i.e., negative and neutral classification results, are decreased by 3% and 5% compared to the positive one. Figure 5b shows the three classes classification ratio by applying the NB classifier to Restaurant dataset. However, it can be seen in the figure 6b that 90.4% TPR more accurately classifies positive polarities in Laptop dataset as compared to the Restaurant dataset's positive polarities. Here, negative and neutral polarities



(a) SVM-ML Classifier (b) NB-ML Classifier (c) RF-ML Classifier

FIGURE 6. SVM, NB and RF classifier TPR and FNR results on Laptop datasets.

classification is competitive with each other by just 0.8% difference. Overall, it can be interpreted from the table 2 that NB-MultiACIA accuracy is diminishing by approximately 4.5% and nearly at 3% respectively, for both datasets.

RF-MultiACIA overall results are less significant among all three general ML classifiers, and table 4 shows the accuracy comparisons of all ML classifiers along with MultiACIA. It can be evaluated that both datasets' accuracy by RF-MultiACIA is either competitive with NB-MultiACIA or less improved with SVM-MultiACIA. Furthermore, the RF classifier on the Restaurant dataset gives a comprehensive TPR classification for positive class, i.e., 89.6%. The class classification results are observable in the figure 5c. The negative and neutral classification of the said dataset is nearly classified with a little difference (3% and 5% respectively). Despite that, the Laptop dataset achieved higher positive class classification results than the dataset mentioned earlier, which is at 91.7% TPR. However, neutral and negative classes are far less accurately classified than positive ones because both are decreasing by approximately 11% and 7% TPR, respectively.

We deduced the experimental outcomes from two perspectives; the first is BERT-MutliACIA (the transformer-based interactive attention approach), and the other is general ML classifier classification results on the aspect-context level feature extracted data. We have compared the state-of-the-art models and analysis evaluation inference with the same sequence shown in Table 4.

VI. CONCLUSION

It is a challenging task to formulate a solution for sentiment classification. The reason is that a single sentence can have multiple aspects. These aspects can be varied to their relevant context. To find the mixed sentiments, we proposed an aspect-context series interactive attention representation model based on the transformer technique. Our enhanced model BERT-MulitACIA can capture the contextualized meanings from the sentences according to their relevant aspects and context. This approach is collecting the semantic relationships among the whole review sentence, which will assist when getting the sequences for aspect-context-based information from the text and the same for context-aspect-based information. At first, to validate our proposed model, we have compared the outcomes with the existing MultiACIA, and it can infer that BERT-MultiACIA either provides better or more competitive results. The proposed model attained better performance on the Laptop dataset regarding its macro-F1 score, but its accuracy and Restaurant dataset accuracy proved that the BERT-MultiACIA is competitive.

To make our results more concise, we have also deployed the final representations from the BERT-MultiACIA model in the general ML classification models. SVM outperforms all three general ML classifiers (the other two are NB and RF). However, this idea can be explored more by deploying the aspect-context representations for other general ML classifiers. In the end, our proposed framework achieves better results. However, it will improve by considering more sequences of aspect-context while deploying the transformer encoder approaches in interactive attention mechanisms. Moreover, our study can be experimented on the other datasets as well, specifically on those which so far do not have their aspect-context relevant terms.

REFERENCES

[1] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, Mar. 2017.

[2] L. M. Rojas-Barahona, "Deep learning for sentiment analysis," *Lang. Linguistics Compass*, vol. 10, no. 12, pp. 701–719, 2016.

[3] A. Rajput, "Natural language processing, sentiment analysis, and clinical analytics," in *Innovation in Health Informatics*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 79–97.

[4] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, May 2012.

[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[6] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[7] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image Vis. Comput.*, vol. 65, pp. 3–14, Sep. 2017.

[8] R. Talib, M. Kashif, S. Ayesha, and F. Fatima, "Text mining: Techniques, applications and issues," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 11, pp. 1–6, 2016.
[9] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016.
[10] M. R. Talib, M. K. Hanif, Z. Nabi, M. U. Sarwar, and N. Ayub, "Text mining of judicial system's corpora via clause elements," *Int. J. Inf. Technol. Secur.*, vol. 9, no. 3, pp. 31–42, 2017.
[11] Z. Wu, Y. Li, J. Liao, D. Li, X. Li, and S. Wang, "Aspect-context interactive attention representation for aspect-level sentiment classification," *IEEE Access*, vol. 8, pp. 29238–29248, 2020.
[12] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent Twitter sentiment classification," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2011, pp. 151–160.
[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[14] B. Huang and K. M. Carley, "Syntax-aware aspect level sentiment classification with graph attention networks," 2019, *arXiv:1909.02606*.
[15] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," 2017, *arXiv:1709.00893*.
[16] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.
[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
[18] N. Ayub, M. R. Talib, M. K. Hanif, and M. Awais, "Aspect extraction approach for sentiment analysis using keywords," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 6879–6892, 2023.
[19] Y. Liang, H. Liu, and S. Zhang, "Micro-blog sentiment classification using Doc2vec + SVM model with data purification," *J. Eng.*, vol. 2020, no. 13, pp. 407–410, Jul. 2020.
[20] R. Asgarnezhad and S. A. Monadjemi, "Nb vs. SVM: A contrastive study for sentiment classification on two text domains," *J. Appl. Intell. Syst. Inf. siences*, vol. 2, no. 1, pp. 1–12, 2021.
[21] S. Ullah, M. R. Talib, T. A. Rana, M. K. Hanif, and M. Awais, "Deep learning and machine learning-based model for conversational sentiment classification," *Comput., Mater. Continua*, vol. 72, no. 2, pp. 2323–2339, 2022.
[22] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of naïve Bayes, support vector machine, decision trees and random forest on sentiment analysis," in *Proc. KDIR*, vol. 1, 2019, pp. 525–531.
[23] Y. Xue, X. Wang, and Z. Gao, "Multi-classfication sentiment analysis based on the fused model," in *Proc. IEEE 31st Int. Conf. Tools With Artif. Intell. (ICTAI)*, Nov. 2019, pp. 1771–1775.
[24] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Inf. Fusion*, vol. 59, pp. 44–58, Jul. 2020.
[25] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Cham, Switzerland: Springer, 2012, pp. 415–463.
[26] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, "AI based emotion detection for textual big data: Techniques and contribution," *Big Data Cognit. Comput.*, vol. 5, no. 3, p. 43, Sep. 2021.
[27] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora, and I. Pappas, "A review on text-based emotion detection—Techniques, applications, datasets, and future directions," 2022, *arXiv:2205.03235*.
[28] F. Fatima, R. Talib, M. K. Hanif, and M. Awais, "A paradigm-shifting from domain-driven data mining frameworks to process-based domain-driven data mining-actionable knowledge discovery framework," *IEEE Access*, vol. 8, pp. 210763–210774, 2020.
[29] K. Rizwan, S. Babar, S. Nayab, and M. K. Hanif, "HarX: Real-time harassment detection tool using machine learning," in *Proc. Int. Conf. Modern Trends Inf. Commun. Technol. Ind. (MTICTI)*, Dec. 2021, pp. 1–6.
[30] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. ICML*, 2011, pp. 513–520.
[31] Z. Nabi, R. Talib, M. Kashif Hanif, and M. Awais, "Contextual text mining framework for unstructured textual judicial corpora through ontologies," *Comput. Syst. Sci. Eng.*, vol. 43, no. 3, pp. 1357–1374, 2022.
[32] *Attention Mechanism in Deep Learning | Attention Model Keras*. Accessed: Jul. 7, 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/
[33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[35] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 452–461.
[36] Z. Lei, Y. Yang, M. Yang, and Y. Liu, "A multi-sentiment-resource enhanced attention network for sentiment classification," 2018, *arXiv:1807.04990*.
[37] B. Liang, Q. Liu, J. Xu, Q. Zhou, and P. Zhang, "Aspect-based sentiment analysis based on multi-attention CNN," *J. Comput. Res. Development. Chin.*, vol. 54, no. 8, pp. 1724–1735, 2017.
[38] X. Feng, B. Qin, and T. Liu, "A language-independent neural network for event detection," *Sci. China Inf. Sci.*, vol. 61, no. 9, pp. 1–12, Sep. 2018.
[39] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," 2018, *arXiv:1805.07043*.
[40] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," 2016, *arXiv:1605.08900*.
[41] T. Hayashi and H. Fujita, "Word embeddings-based sentence-level sentiment analysis considering word importance," *Acta Polytechnica Hungarica*, vol. 16, no. 7, pp. 7–24, 2019.
[42] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Model. Predict. Behav. Represent. Modeling Simulation*. Cham, Switzerland: Springer, 2018, pp. 197–206.
[43] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "Exploiting document knowledge for aspect-level sentiment classification," 2018, *arXiv:1806.04346*.
[44] Z. Chen and T. Qian, "Transfer capsule network for aspect level sentiment classification," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 547–556.
[45] J. Sarzynska-Wawer, A. Wawer, A. Pawlak, J. Szymanowska, I. Stefaniak, M. Jarkiewicz, and L. Okruszek, "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Res.*, vol. 304, Oct. 2021, Art. no. 114135.
[46] A. Rietzler, S. Stabinger, P. Opitz, and S. Engl, "Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification," 2019, *arXiv:1908.11860*.
[47] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," 2019, *arXiv:1903.09588*.
[48] D. Zhang, Z. Zhu, Q. Lu, H. Pei, W. Wu, and Q. Guo, "Multiple interactive attention networks for aspect-based sentiment classification," *Appl. Sci.*, vol. 10, no. 6, p. 2052, Mar. 2020.
[49] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, S. Mishra, and A. Abraham, "AI-based conversational agents: A scoping review from technologies to future directions," *IEEE Access*, vol. 10, pp. 92337–92356, 2022.
[50] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting BERT for end-to-end aspect-based sentiment analysis," 2019, *arXiv:1910.00883*.
[51] J. Zhan, J. Mao, Y. Liu, M. Zhang, and S. Ma, "RepBERT: Contextualized text embeddings for first-stage retrieval," 2020, *arXiv:2006.15498*.
[52] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "LCF: A local context focus mechanism for aspect-based sentiment classification," *Appl. Sci.*, vol. 9, no. 16, p. 3389, Aug. 2019.
[53] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," 2018, *arXiv:1805.01086*.
[54] B. Chen, R. Sun, Y. Dai, H. Zheng, and R. Zhang, "LordBERT: Embedding long text by segment ordering with BERT," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2022, pp. 1–8.
[55] X. Ma, Z. Wang, P. Ng, R. Nallapati, and B. Xiang, "Universal text representation from BERT: An empirical study," 2019, *arXiv:1910.07973*.
[56] F. Meng, J. Feng, D. Yin, and M. Hu, "A new fine-tuning architecture based on BERT for word relation extraction," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.* Cham, Switzerland: Springer, 2019, pp. 327–337.
[57] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 27–35.
[58] H. T. Nguyen and M. Le Nguyen, "Effective attention networks for aspect-level sentiment classification," in *Proc. 10th Int. Conf. Knowl. Syst. Eng. (KSE)*, Nov. 2018, pp. 25–30.