The Journal of Supercomputing (2023) 79:11452–11477
https://doi.org/10.1007/s11227-023-05112-7

MAPA BiLSTM-BERT: multi-aspects position aware attention for aspect level sentiment analysis

Mayur Wankhade[1] · Chandra Sekhara Rao Annavarapu[1] · Ajith Abraham[2]

Accepted: 9 February 2023 / Published online: 28 February 2023

Abstract
Sentiment categorization at the aspect level tries to provide fine-grained sentiment polarities for distinct aspects inside a sentence. Some issues remain unresolved in the previous work. First, the specific position context is not fully addressed. Second, the distinct aspect of an opinionated sentence is evaluated independently. Also, the present, attentive approaches neglect the disruption caused by all the aspects in the same sentence while measuring the current aspect attention vector. We proposed multi-aspect-specific position attention bidirectional long short-term memory (MAPA BiLSTM)-bidirectional encoder representations from transformers (BERT) model to address these issues. The MAPA BiLSTM BERT introduces the explicit multiple aspect position-aware attention between the aspect word and the closest context words, also BERT aspect-specific attention investigates how to model multiple aspects using the aspect position attention mechanism. The parallel fused MAPA BiLSTM-BERT gains the multiple aspect contextual knowledge in a sentence for aspect classification. We conduct an empirical evaluation of the proposed method on the laptop review, restaurant review (SemEval2014 datasets), Twitter review, and multi aspect multi-sentiment (MAMS) dataset; results indicate a significant performance improvement over state-of-the-art methods.

✉ Mayur Wankhade
mayur.18dr0078@cse.iitism.ac.in

Chandra Sekhara Rao Annavarapu
acsrao@iitism.ac.in

Ajith Abraham
abraham.ajith@gmail.com

1 Department of Computer Science, Indian Institute of Technology (Indian School of Mines), Dhanbad 826004, Jharkhand, India

2 Machine Intelligence Research Labs, Auburn, WA 98071, USA

Keywords Sentiment analysis · Position attention · Deep learning · Machine learning · Aspect base sentiment analysis

1 Introduction

As electronic information technology evolves, a growing number of users share their thoughts or experiences on social media platforms which helps to identify opinions about a product or service. As a result, how to mine public opinion for vital information has become a trending topic in natural language processing (NLP) research. So to identify public opinion is also called sentiment analysis. The goal of sentiment classification is to identify the polarity of a text's sentiment as positive, negative, or neutral. It is employed in almost every industry including healthcare, consumer items, criminal investigations, political elections, and the identification and prevention of fraud. As a part of sentiment analysis, aspect-based sentiment analysis (ABSA) is a subcategory that tries to determine at aspect level sentiment classification inside the sentence into the positive or negative or neutral. Its purpose is to extract and summarize people's opinions about entities and their aspects, referred to as targets [1].

ABSA is a subtask of sentiment analysis that seeks to discern the sentiment polarities associated with each aspects in a sentence. It is helpful to identify opinions at aspect level sentiment classification [2]. It is a fine-grained sentiment analysis task that requires the detection of a specific aspect's sentiment polarity. ABSA is subdivided into two parts first one is aspect term sentiment analysis (ATSA), and the second is aspect category sentiment analysis (ACSA). Sentiment analysis at the aspect level seeks to determine the sentiment polarity of certain aspect words in the sentence. Traditional approaches for aspect-level sentiment analysis (ALSA) relied heavily on constructing a sentiment lexicon-based dictionary. In general, an ALSA must address two issues: the analysis of aspect sentiment polarity and the extraction of aspect entities. However, lexical dictionaries and feature-based techniques are time-consuming. The ability to handle sentences with complicated structures deep learning approaches have recently become the de facto standard for sentiment categorization [1, 3]. Furthermore, it has been shown that the attention mechanism effectively captures potential semantic relationships between the aspect and the context word.

Sentiment analysis at the aspect level considers that sentences may contain descriptions of multiple aspect opinion on a single object, and such an analysis must determine the sentiment polarity of a particular object or angle. Numerous serialization models incorporating attention mechanisms have been suggested in [4] conjunction with deep neural networks, including ABCDM, attention emotion-enhanced convolutional (AEC)-LSTM, attention-based models such as convolution neural networks (CNN) [5], ATAE-LSTM [6], target dependent (TD)-LSTM [7], and position attention LSTM [8]. However, these sentiment analysis at the phrase level cannot immediately apply to the ABSA issues, as the sentence sentiment reflects inconsistent beliefs about certain aspect words[9]. These models are primarily based on serialized data and employ word-level, part of speech (POS), and data fusion as inputs. Several studies have concentrated on forecasting sentiment intensity via stacked ensembles when an aspect word is dissociated from its sentiment polarity [10].

ABSA can be a challenging task in categorization because of the ambiguity of the sentence. The context-based characteristic is frequently a significant factor in sentiment classification, implying that the comprehension of a word is mainly dependent on the position of the aspect word and the context words. Thus, both aspect word's and context words position become significant features for sentiment classification [9].

The attention mechanism has demonstrated remarkable performance in neural models for various NLP tasks, including semantic entailment [11], machine translation [12], ALSA [13]. The attention mechanism prioritizes the most relevant features for desired actions. It's a technique to extract the most significant context words for subsequent model training as supervised attention signals to circumvent. Although prior attention-based techniques have been demonstrated to be successful and helpful in classifying sentiment at the aspect level.

The research motivation is aspect-level sentiment classification that can be used to examine consumer feedback by correlating distinct sentiments with various aspects of a product or service. ABSA is the process of determining the sentiment polarity toward the multiple aspects, which are typically expressed directly or implicitly in the form of preset aspect categories in user-generated natural language texts. The objective is to determine the sentiment associated with a specific aspect category. The majority of recent investigations in this subject have followed this structure. Given that a review text may reference many targets, it is vital to analyze the sentiment associated with various features of each target. Influenced by the current success of the attention mechanism, our model enhances the standard word embedding by including context data and enhances the capability of aspect level sentiment analysis.

The majority of existing techniques manage categorization by evaluating the value of context words concerning each particular aspect sentiment to the benefits provided by aspect relations. The goal is to classify sentiment polarity for various aspects of the sentence. In the prior work, there are still unresolved concerns. Firstly, the position-specific context is not adequately handled. Second, each part of an opinion sentence is independently examined. In addition, the contemporary, attentive methods disregard the disturbance generated by all the aspects in the same sentence while evaluating the current aspect. To address this issue, we propose a novel MAPA BiLSTM BERT approach for enhancing aspect-level sentiment classification. Our work jointly used content words with respective position attention mechanisms for each aspect, which is helpful for a more precise analysis of each specific aspect-level sentiment categorization. The BiLSTM attention employs aspect disagreement regularization to improve the identification of aspect-specific features from overlapped representations. The proposed approach automatically extracts supervision information from the training corpus and gradually aid in training ABSA models' attention processes.

The rationale behind our approach is as follows: if a trained model can accurately forecast its sentiment for a training instance, its most significant context words should be continually considered during the succeeding model training phase.

1.1 Research contribution of our work as mentioned below

- We propose a novel multi-aspect attention learning network for relation construction that uses the bi-attention mechanism to capture bidirectional semantic knowledge between the context word and the aspect word.
- Aspect-specific attention improves by integrating content and position attention, which allows for a more precise analysis of the impact of each context word on a specific aspect.
- BERT superior context modeling capability enables it to build more accurate aspect-specific sentence representations, enabling our model to make more accurate predictions.
- We validate the proposed model using various datasets result demonstrated outperformed the baseline model in the experiments.

Remaining research paper is organised as follows. Section 2 discuss the related work. Section 3 presents our proposed method. Section 4, Result analysis and comparison. Finally, Sect. 5 conclusions our work.

2 Related work

This section is divided into three subsections. The first section discusses ABSA, the second section discusses aspect-based position attention, and the third section aspect-specific attention mechanisms.

2.1 Aspect-based sentiment analysis

Traditional sentiment analysis approaches rely heavily on typical machine learning algorithms to predict the sentiment polarity of text [14]. Earlier studies used feature-based approaches, which are feature engineering based on frequency. ALSA task is more precise as compared to document-level, sentence-level. Wang et al. [15] suggested a sentiment-aligned topic method to predict product aspect ratings. In [16] investigated method for extraction and development of a sentiment dictionary through various extended dictionaries and then determined the sentiment value associated with a microblog text. Previous research work relied heavily on discriminative classifiers with human label feature with support vector machine (SVM) approach [17], ABSA approach based on neural networks have gradually gained supremacy in recent years [18, 19]. Compared to earlier conventional models, neural ABSA models perform better because they can learn aspect-related semantic representations of input texts. The majority of research has concentrated on sentiment associated with a whole piece document or sentence [20, 21]. Individuals may refer to many target entities (or various qualities of a target) in a single document or sentence in the real world. In [22] suggested weak supervised neural network for learning a set of sentiment clusters embedding from the target domain's phrase global representation.

The deep learning technique for sentiment analysis focuses on developing and deploying for predicting the underlying sentiment in texts. Numerous neural network approaches have been suggested in this field of study, including convolution neural networks (CNN), LSTM, recursive neural networks (RCNN), and recurrent neural networks (RNN) [23]. Among these methods, LSTM is a well-known design that is well-suited for language representation learning. The gated approach in the sequence encoder allows LSTM to capture present significant and long-distance dependence for sentiment analysis, which is particularly effective for long-span texts [24]. Based on the aforementioned neural frameworks, an attention mechanism is described that enables models to pay attention to significant word patterns that may signal emotion polarity via attention weights [25].

The deep learning-based approach gained substantial popularity in ABSA and is capable of dynamic feature extraction. This approach automatically encode the original features as low-dimensional vectors. Syntactical information can provide information about the relationship between a specific aspect and the sentiment expression in ABSA [26]. As a result, grammatical relationships between contextual and aspect terms should be addressed when predicting sentiment polarity for a particular aspect. Several graph networks based on dependency trees have been introduced in the structure of a sentence using the sentence dependency tree [27]. Thus, these approaches can assign syntactically significant words to a particular aspect and, via convolutions operations, capture long-term syntactic dependencies. In addition, Zhang et al. [28] highlighted the advantages of employing graph convolutional networks (GCN) to extract the syntactical relationships inside the context words.

2.2 Aspect base position attention

Positional embedding-based ABSA models incorporate the location information for the provided aspect word [9, 29]. These methods use the position of the aspect attention word about the context word by calculating the number of words or by utilizing tree structure dependency. The concept that the nearest contextual word to the aspect term is more significant for memory-based models promotes attention processes [30]. Similarly, it is used to offset the memory network's disadvantages[31]. Additionally, the position-based ABSA issue is illustrated in [32] by concatenating position and word embedding to create a position-aware context representation. In the work of [33] ABSA is broken down into the following three tasks: Sentiment tagging goes through N rounds of sequence tagging. It takes the *i*th word w_i as a possible opinion target in the *i*th round, and it gives each word in the sentence a tag called sentiment tag associated with the opinion target. Then, it moves on to the next word considering an opinion target comprises more than one word. The structure of the opinion encoder is the same as the structure of the target encoder. As a result, several neural networks have gotten a boost in popularity.

LSTM-based models need more time during the training process. In [34] used a position-based bidirectional GRU network for the ABSA task. Other effective networks have been employed to extract information from a sentence's words, including memory and CNNs. Tang et al. [35] suggested a deep memory network that explicitly remembers each context word's essential features in terms of an aspect. Multiple deep memory networks were used to construct the crucial degree evaluations. Xue et al. [5], on the other hand, suggested a solution based on CNN and gate units. The gated tanh unit, a one-of-a-kind gate mechanism, generated emotion characteristics in response to a specific aspect word. Individuals may refer to many target entities (or various target characteristics) inside a single sentence or document in the real world. By focusing on the overall mood of the sentence or document, we miss out on attitudes toward specific target entities [36, 37]. A range of techniques for target-dependent or aspect-level sentiment categorization has been developed [38]. For instance, in [39], product characteristics were used to predict the polarity of attitude toward the target object. Jiang et al. [36] advocated integrating syntactic characteristics to enhance target-dependent sentiment analysis.

2.3 Aspect-specific attention mechanisms

The attention mechanism [40], which is frequently used in deep learning-based approaches, enhances the classification capabilities of the models. When someone reads a text, they will focus on individual words or phrases rather than the entire documents [41]. Ma et al. [42] described an interactive attention neural network model obtained attention ratings in contexts and aspects via interactive attention and then built representations for each independently. Huang et al. [43] suggested ABSA task using neural network based on an double attention neural network. They combined aspect and context concepts and captured essential qualities shared by words in a phrase. Previous attention mechanisms are neural network models such as CNN, memory networks, and RNN in order to learn different levels of attention for different aspects and generate attention-based sentence embeddings. In [44] used the GRU to perform aspect-based sentiment analysis on complex sentences in conjunction with a content attention method.

Recently, the attention mechanism has demonstrated potential in NLP applications [45]. In the attention approach [46, 47], which is extensively used in machine translation [48] and picture recognition [49]. It is included to compel the model to prioritize context words with more semantic relationships than aspect words. The attention method receives numerous weighted ratings and develops new sentence representations by interactively modeling context and aspect representations [50, 51]. Deep neural networks allowing to improve their performance by learning where to direct their attention. In [52, 53] show instances of recent research on sentiment analysis based on attentiveness. The attention-based bi-directional CNN-RNN deep model employs an attention mechanism that extracts both previous and current contexts while taking temporal knowledge flow into account [54]. In ABCDM, the attention mechanism is used to the bidirectional layer outputs to change the prominence more or less on various words. Additionally they added common sense knowledge than language modeling. For instance, SenticNet is a concept that integrates top-down as well as bottom-up learning through the use of a combination of symbolic and sub-symbolic artificial intelligence methods [55]. Neural networks have become prominent in various language applications in the last decade, including sentiment analysis. Regularized multi-task learning technique contain multiple domain models are bound to be comparable to their average model [56].

In previous work has resulted in considerable advancements in ALSA because attention processes have been used to acquire semantically meaningful aspect-specific representations. Numerous scholars have attempted to solve the ABSA problem using a variety of ways and have achieved some impressive achievements. However, previous attention approach in the ABSA model have a significant shortcoming shared with neural models of other NLP tasks. Existing research has identified two distinct patterns in neural network learning: "apparent patterns" that are over-learned and "inapparent patterns" that are not appropriately learned [15, 57]. As a result, attention ABSA models advantage to over-focus on high-frequency phrases with strong sentiment polarities and under-focus on low-frequency terms, resulting in suboptimal performance [58].

3 The proposed method

The building of the proposed method for aspect-level sentiment classification is explained, and our workflow is shown in Fig. 1. The BiLSTM network and BERT are used to acquire a more profound semantics knowledge of each word and contextual

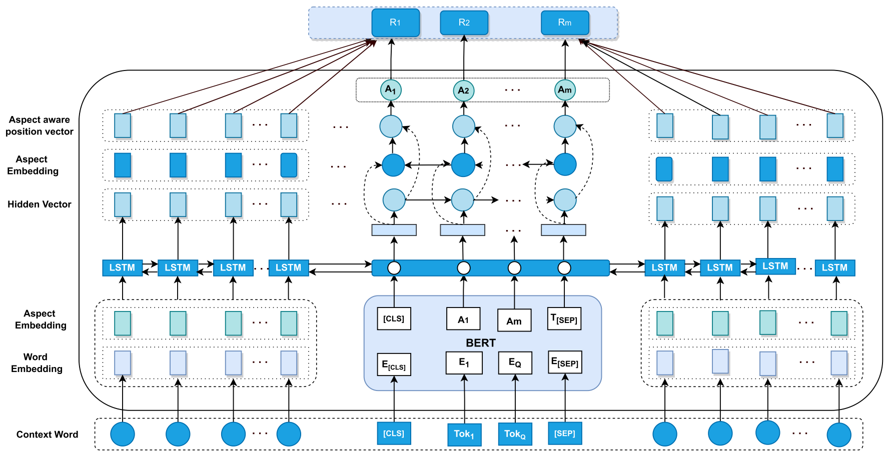

Fig. 1 The proposed model workflow for aspect-based sentiment classification

sentimental correlation. The BiLSTM networks are constructed of forward and backward sequence models. Advantage BERT models with the benefit of sequence-to-sequence have demonstrated decent performance in NLP. We selected them as the pre-trained model for aspect embedding. The parallel fused MAPA BiLSTM-BERT gains the multiple aspect contextual knowledge in a sentence for aspect classification. ABSA specific objective is to classify sentiment polarity toward various aspects of a sentence in order to acquire more insight into user reviews via fine-grained sentiment analysis of affective texts.

3.1 Task definition

The objective of ALSA is to forecast a sentiment classification for multiple aspect in the sentence (aspect, sentence). In sentence having $S = [w_1, w_2, \dots, w_n]$ consist of 'n' word and 'm' aspect $A = [A_1, A_2, \dots, A_m]$ where the ith aspect $A_i = [w_i, w_{i+1}, \dots, w_n]$ is a sub-sequence. The objective is to classify each aspect A_i with sentiment polarity (negative, neutral, positive). For instance, the statement " The [food] taste is pleasant, but very high [price]." has two aspect terms: [A_1 : food] and [A_2 : price] which is supposed to produce [food]: positive and [price]: negative outputs. In second case " The taste of [food] is delicious also the [price] reasonable, but the [service] is worst." consist of three aspect terms: [A_1 : food], [A_2 : price], and [A_3 : service] which is supposed to produce [food]: positive, [price]: positive and [service]: negative outputs. Our task is to find out the multiple aspect-level sentiment polarity classification in sentence, as shown in Fig. 2.

3.2 BiLSTM multi-aspect attention

BiLSTM has been demonstrated to be an efficient method for incorporating contextual knowledge into word embedding. Also, the hidden layer is performed as a sequence of combined context knowledge. The hidden state in LSTM holds knowledge regarding the current context word and the preceding moment. BiLSTM incorporates bidirectional knowledge in both directions, and the resulting aspect representation of a particular word is derived by appending both directions representations. Thus, combining the two hidden states can better comprehend the context than the unidirectional LSTM. In the sentence current input word vector $w_t \in R^d$ vectorizes the tth sequence words of the sentence. The forward and backward state at different level are represent as the current cell state c_t, current hidden state h_t of the BiLSTM can be computed as below:

$$\vec{i_t} = \sigma_g \left(W_i w_t \cdot W_i h_{(t+1)} + b_i \right) \tag{1}$$

The taste of food is delicious also price reasonable, but the service is worst.

Aspect : food, Polarity : positive ; Aspect : price, Polarity : positive ; Aspect : service, Polarity : negative

Fig. 2 Example of aspect level sentiment analysis for finding multiple aspect category in the sentence

$$\overleftarrow{i_t} = \sigma_g\left(W_i w_t \cdot W_i h_{(t-1)} + b_i\right) \tag{2}$$

$$\overrightarrow{f_t} = \sigma_g\left(W_f w_t \cdot W_f h_{(t+1)} + b_f\right) \tag{3}$$

$$\overleftarrow{f_t} = \sigma_g\left(W_f w_t \cdot W_f h_{(t-1)} + b_f\right) \tag{4}$$

$$\overrightarrow{o_t} = \sigma_g\left(W_o w_t \cdot W_o h_{(t+1)} + b_o\right) \tag{5}$$

$$\overleftarrow{o_t} = \sigma_g\left(W_o w_t \cdot W_o h_{(t-1)} + b_o\right) \tag{6}$$

$$\overrightarrow{c_t} = f_t \odot c_{(t+1)} + i_t \odot \tanh\left(W_c \cdot [w_t, h_{(t+1)}] + b_c\right) \tag{7}$$

$$\overleftarrow{c_t} = f_t \odot c_{(t-1)} + i_t \odot \tanh\left(W_c \cdot [w_t, h_{(t-1)}] + b_c\right) \tag{8}$$

$$h_t = o_t \odot \tanh(c_t) \tag{9}$$

The sigmoid function is utilized in the equations to regulate the input and output of data during each loop. The variables to be acquired during training are $W_i, W_f, W_o, W_c, b_i, b_f, b_o, b_c$ hidden values have the same dimension as a word vector representation. Inspired by aspect embedding [59] to automatically enable the neural network to learn representation of aspect terms. Also, the aspect embeddings are viewed as model parameters that must be learned throughout the training process. In our proposed approach, BiLSTM is used to input both the context word and word vector. Also, the aspect embedding with the BERT model is used in our work to direct the multi-aspect attention mechanism, which has been empirically demonstrated to be effective toward additional information about the aspect. The following equations are computed as a i_t,f_t, o_t, and c_t :

$$\overleftarrow{i_t} = \sigma\left(W_i w_t \cdot W_i v_a \cdot W_i h_{(t-1)} + b_i\right) \tag{10}$$

$$\overrightarrow{i_t} = \sigma\left(W_i w_t \cdot W_i v_a \cdot W_i h_{(t+1)} + b_i\right) \tag{11}$$

$$\overleftarrow{f_t} = \sigma\left(W_f w_t \cdot W_f v_a \cdot W_f h_{(t-1)} + b_f\right) \tag{12}$$

$$\overrightarrow{f_t} = \sigma\left(W_f w_t \cdot W_f v_a \cdot W_f h_{(t+1)} + b_f\right) \tag{13}$$

$$\overleftarrow{o_t} = \sigma\left(W_o w_t \cdot W_o v_a, h_{(t-1)}] + b_o\right) \tag{14}$$

$$\overrightarrow{o_t} = \sigma\left(W_o w_t \cdot W_o v_a, h_{(t+1)}] + b_o\right) \tag{15}$$

$$\overleftarrow{c_t} = f_t \odot c_{t-1} + i_t \odot \tanh\left(W_c \cdot [w_t, v_a, h_{(t-1)}] + b_c\right) \tag{16}$$

$$\overrightarrow{c_t} = f_t \odot c_{t+1} + i_t \odot \tanh\left(W_c \cdot [w_t, v_a, h_{(t+1)}] + b_c\right) \tag{17}$$

where $v_a \in R^{d_a}$ derive the aspect embedding matrix M. The functions such as

$$\Theta^{\text{BiLSTM}} = \left\{W_i \in R^{(2\times d^2+d\times d_a)}, W_f \in R^{(2\times d^2+d\times d_a)}, W_c \in R^{(2\times d^2+d\times d_a)}, \quad W_o \in R^{(2\times d^2+d\times d_a)}\right\}$$

and model parameter such as $\{b_i \in R_d, b_f \in R_d, b_o \in R_d, b_o \in R^d\}$ are indicate as the learning parameters.

In sentence having $S = [w_1, w_2, \dots, w_n]$ consist 'n' word as the input to the BiLSTM networks in the GloVe+BiLSTM model, the output of the BiLSTM hidden layer contains the forward state $\overrightarrow{h_f} = [h_{f_1}, h_{f_2}, \dots, h_{f_n}]$ and backword $\overrightarrow{h_b} = [h_{b_1}, h_{b_2}, \dots, h_{b_n}]$ output of each context. Whereas $\overrightarrow{h_f}, \overrightarrow{h_b} \in \mathbb{R}^{n*dh}$ dh represent hidden state dimension. The combining the forward and backward output the final outcomes represented as the following equation can be used to obtain the concatenated output in the BiLSTM layer:

$$h_i = [\overrightarrow{h_f}, \overleftarrow{h_b}] \tag{18}$$

The position attention towards each aspect word in sentence $w = (w_1, w_2, \dots, w_n)$ consist 'n' word and 'm' aspect $A = [A_1, A_2, \dots, A_m]$ where the ith aspect $w_i = [w_{A_1}, w_{A_2}, \dots, w_{A_m}]$ is a sub-sequence. In the aspect word feature vector $\overrightarrow{h_a} = [h_{i_1}, h_{i_2}, \dots, h_{i+m-1}]$ which contain the semantic correlation between the aspect word and closest context words. A BiLSTM evaluates a sequence characteristic in both forward and backward orientations. We define the ith token's hidden states as follows:

$$\overleftarrow{h_i} = \text{LSTM}(w_i, \overleftarrow{h_{i-1}}) \tag{19}$$

$$\overrightarrow{h_i} = \text{LSTM}(w_i, \overrightarrow{h_{i+1}}) \tag{20}$$

Given a sentence that not all words in the context contribute equally to sentiment categorization, an attention layer is used to give priority to the most significant aspect in the context. BiLSTM aspect attention layer is wrapped to learn a weight of α_t for every hidden state h_t received at time step t. Because the padded sequences have k inputs, the time step t will be between 1 and k. The weighted vector for attention mechanism adopted by [60], $\alpha_t = [\alpha_1, \alpha_2, \dots, \alpha_k]$, is determined using the output sequence $H = [h_1, h_2, \dots, h_k]$.

3.2.1 BERT aspect attention

We begin the input sentence with a special token [CLS] and end it with a special token [SEP] for an aspect encoder that is structurally identical to the target encoder. We employ a BERT to capture the aspect characteristics of opinion expressions: each token is turned into a vector by adding its token, segment, and position embeddings. We add a task-specific layer to BERT and fine-tune the parameters of the original BERT to match the aspect embedding attention. The BERT model uses the masked language technique to forecast some randomly masked tokens out of a sequence with a predefined input structure. [CLS] $w = [w_1, w_2, \dots, w_n]$ [SEP] and 'm' aspect [CLS] $A = [w_i, w_{i+1}, \dots, w_{i+m-1}]$[SEP]. The vector sequence v_i is then passed through the stacking of transformer layers to encode contextual information:

$$h_i^N = \mathrm{BERT}\{v_i^N\} \tag{21}$$

where $h_i \in \mathbb{R}^{dh}$ which represent the contextual information of ith word in the sentence and dh represent hidden state dimension. Inspired by BERT[59] aspect embedding mechanism. We have fine-tune the ABSA BERT. The parallel fusion aspect attention BiLSTM-BERT gains the multiple aspect contextual knowledge in a sentence for aspect classification. The input mentioned above representation is subsequently supplied into L consecutive Transformer encoder blocks. To get the fixed dimensional fix representation of the input sequence, we use h_0 as the first token $[CLS]$ of the last hidden state as per the input sequence to next layer.

3.2.2 Multi-aspect position attention mechanism

The original motivation for incorporating attention mechanisms [61] into ABSA tasks is that each word in a phrase affects a different aspect. This stage focuses on determining the position of the words expressing the sentence's aspects (called aspect words). We use attention techniques in this work to determine the sentiment contribution of each word to various elements. Consider the line "delicious food but terrible service at this restaurant". The context term "delicious" is more significant for the "food" aspect than the term "terrible". On the contrary, the context term "terrible" is more significant for the "service" aspect than it is for the "delicious" aspect. Multiple computational layers are capable of learning higher-level representations and performing remarkably well at the aspect level for sentiment analysis. We use multiple attention in this study to discover how to describe attributes. This enables the Bi-directional deep memory network to learn multiple-level representations of text.

We have an aspect word $A = [A_1, A_2, \dots, A_m]$, where m indicate the number of aspects in the given opinions, the position embedding of words w_i in sentence s, indicate by $p_i \in \mathbb{R}^m$, is constructed by the following processes.

The attention mechanism is employed in this work to combine the positions of words representing aspects and to combine their representations. This technique can be used to emphasize the distinctions between words expressing aspect-level sentiment without expressing aspect-level sentiment explicitly. Multi-aspect target-dependent sentiment categorization can improve by measuring words proximity to target entities. Intuitively, a closer context word has a more significant influence on aspect words. For example, in the statement "The food is delicious but not worth the price," the aspect word [food] and closest sentiment word "delicious" should be given a higher weight also the aspect word [price] and closed context word are "not worth" should be given a higher weight. To improve the performance of the aspect vectors, we must measure the distance between the aspect word and each context word. It is used to determine the weight assigned to each context word. We use absolute distance as a metric and give closer terms a higher weight. To be more precise, we begin by calculating the position weight p_i of the ith word:

$$p_i = \begin{cases} 1 - \frac{t+k-1-i}{l} & \text{for } i < t + k - 1 \\ 1 - \frac{i-t}{l} & \text{for } t + k - 1 \leqslant i \leqslant n \\ 0 & \text{for } i > n \end{cases} \tag{22}$$

where 'l' could be a sentence length, t indicates the position of the first word in aspect, and k indicates the aspect word length. In the case of the length, the sentence is less than the conditional threshold, then padded with zeros. As a result, the position correlation coefficient for this section is 0. The position attention method is used to improve the differentiation of useful and unimportant elements in terms of aspect sentiment. The following equation is used to define a score function f that captures the significance of the information in the sentence.

$$f(c_i, p_i) = \tanh\Big(F \cdot [c_i, p_i] + b\Big) \tag{23}$$

The relevant information utilized to determine the sentiment polarity towards the aspect entity in aspect-dependent categorization is frequently derived from words close to the aspect entity. Additionally, to maximize the utility of the aspect knowledge, we examine the position of context words and the aspect word. The influence of each context word we can analyze using the kernel function as given below:

$$\text{Kernel}(\mu) = \exp\left(\frac{-\mu^2}{2\lambda^2}\right) \tag{24}$$

where λ indicates the propagation range, μ indicates the distance between aspect word and current context words. The closer the distance, the more quickly the position-aware impact increases. The impact of a particular distance is calculated using equation 22 to follow a Gaussian distribution in each dimension. As a result, the effect is stretched into a matrix denoted by the symbol P. The influence on the distance of in the ith dimension μ is calculated as mentioned below:

$$P(i,\mu) \sim N\left(\text{Kernel}(\mu), \sigma'\right) \tag{25}$$

where, $P(i,\mu)$ as the normal distribution, kernal mean value μ and standard deviation of σ'.

$$A_i^j = \frac{\exp\left(e(h_j, v_a^i, p_j)\right)}{\sum_{k=1}^{n} \exp\left(e(h_k, v_a^i, p_k)\right)} \tag{26}$$

where $e(h_j, v_a^i, p^j)$ is an scoring functions that measures the semantic relationships between the jth closest context word and the aspect word. $\Theta^{att} = \{W_h \in R^{d\times d}, W_p \in R^{d\times d_p}, W_a \in R^{d\times d_a}, b \in R^d, \eta \in R^d\ \}$ are training parameters determined the aspect-specific sentence representation is mentioned below as:

$$e(h_j, v_a^i, p_j) = \eta^T \tanh\left(W_h h_j + W_p p_j + W_a v_a^i + b\right) \tag{27}$$

$$r_i = \sum_{j=1}^{n} A_i^j h_j \tag{28}$$

Also, aspect specific position aware attention attentive represent 'r' is mapped into sentiment class 'C'.

$$r_i = \tanh\left(W_r r_i + b_r\right) \tag{29}$$

$\Theta^{\text{classifier}} = W_r \in R^{d\times C}, b_r \in R^C$ are indeed the parameter estimates. Then, a softmax function is used to calculate the sentiment dispersion:

$$g_c = \frac{exp(\hat{r}_i^c)}{\sum_{j=1}^{c}(\hat{r}_i^z)} \tag{30}$$

This regularization term ensures that the attention weight vectors of multiple aspects are orthogonal, ensuring that separate aspect categories pay attention to distinct parts of the supplied text with little overlap. The regularization is used to the layer of attention with multiple perspectives. In sentence $A = \{A_1, A_2, \dots, A_m\}$ and attention weight $\alpha_1, \alpha_2, \dots, \alpha_m$ is a matrix formed by based on attention-weight in the sentence, and $A_i = \{A_i^1, A_i^2, \dots, A_i^n\}$ denotes the aspect probability distribution of attention for each ith aspects of the sentence as measure by equation 22 mentioned position attention. To keep each aspect focused on specific sentence words and to reduce the disruption caused by all of the aspect, the penalization term approach is:

$$P = \|(AA^T - I)\|_{\mathrm{FBN}}^2 \tag{31}$$

where as 'FBN' denotes the Frobenius norm, and identity matrix indicate as 'I'. We fine-tune the BERT multiple aspect embedding stage with MAPA BiLSTM stage. As a result, the ith attentive representation calculated as mentioned below:

$$R_i = \sum_{j=1}^{n} A_i^j h_j \tag{32}$$

Finally, the multiple aspects A_m the attentive representations for the input computed as classification.

3.3 Model training

Sentiment classification based on aspect is used to predict the sentiment classification of a given aspect in a sentence s using the trained model. After performing the multi-layer calculations, the final output vector is obtained. After passing through a linear layer, the output vector is fed into the softmax function, which calculate the probability distribution. The loss function is calculated as:

$$L = -\sum_{d=1}^{D} \sum_{c=1}^{C} y_c(d) \cdot \log(g_c(d)) + \lambda L_2(\Theta) \tag{33}$$

where D denotes the dataset, d denotes a single sample, $y_c(d)$ denotes the sentiment polarity distribution, and L_2 is the regularization coefficient. It is worth noting that while the double BiLSTM position aware aspect attention mechanism stage models have the same network design, the BERT aspect attention model benefits from the Frobenius norm fine-tune based penalization term.

Table 1 The statistical information of the laptops review, restaurant reviews, Twitter reviews and MAMS datasets

Dataset	Parameter	Positive	Negative	Neutral
Laptop	Training	994	870	464
	Testing	3417	128	169
Restaurant	Training	2164	807	637
	Testing	728	196	196
Twitter	Training	1561	1560	3127
	Testing	173	173	346
MAMS	Training	3380	2764	5042
	Testing	400	329	607

4 Experimental results and analysis

4.1 Dataset

We analyze the effectiveness of our proposed method with four benchmark datasets: laptops reviews, restaurant reviews (SemEval 2014 Task), Twitter reviews, and MAMS. Table 1 summarizes the statistical characteristics of datasets. The Restaurant data set from SemEval 2014 is frequently utilized in ACSA tasks. There are various aspect in dataset include "food", "price", "service", "ambiance", and "miscellaneous". Sentiment polarities can be categorized into positive, negative, or neutral. Multi-aspect Multi-sentiment (MAMS) is a challenging dataset for ABSA, where each sentence contains at least two aspects with contrasting sentiment polarity. MAMS Jiang et al. [62] present a dataset for a recent issue with 4,297 sentences and 11,186 characteristics. The average MAMS sentence contains 2.62 aspect terms and 2.25 aspect categories. Each sentence in MAMS comprises many components with opposing emotional valences. We have used a smaller version of the MAMS dataset (named MAMS-small) created by randomly picking training instances from the MAMS training set.

4.2 Experiment setup

The experiments were performed using an NVIDIA Tesla GPU to train our model. We discuss the proposed approach experimental setup in details. Our investigations are based on SemEval2014 dataset ABSA datasets that contains a single sentence from a product review and the related aspect term and sentiment label. The datasets are omitted from "null" aspect terms, and "dispute" or multiple sentiment labels are also excluded from the analysis's aspect terms. The rest of the sentences contain at least one characteristic of the word that has been labeled as positive, neutral, or negative. The hyper-parameter setting in proposed approach as of dropout rate = 0.1, regularization weight = 0.001, learning rate = 0.001, dimensions of hidden size BiLSTM is = 300. Model parameters used in our work with respective content attention and position attention for all four datasets as Restaurant (109.51 M BERT-ATT case, 110.01 M Our case), Laptop (109.54 M BERT-ATT, 110.00 M Our case), Twitter (4.54 M BERT-ATT, 110.37 M, 113.94 M Our case), MAMS (2.96 M BERT-ATT, 109.59 M 110.01 M Our case). We have 50 epochs of fine-tuning in our case, and the batch size is fixed at 16. When training the BERT prediction model, we employ an adam optimizer with a learning and dropout rate. Furthermore, the final classification model has been trained for 10 epochs with a batch size of 16. We have used several parameters based on various datasets with different cases. We conducted experimental results on five random seeds and reported the average accuracy and macro-F1 values.

4.3 Baseline methods

4.3.1 Feature-based methods

- *SVM* classifier uses n-gram, lexicon, and parse characteristics[63].
- *SVM-feature* A multi-class SVM classifier has been used to train this model. The groups feature were used to represent each sentence: n-grams, non-contiguous n-grams, part of speech tags, cluster n-grams, and lexicon features[64].

4.3.2 Attention-based methods

- *TD-LSTM* which divides the text sequence in two parts and a new text sequence is formed by combining the left part and aspect terms, and the right part accomplishes the same function. The next model the two new phrases using two distinct LSTM networks and concatenate the target-specific predictions for estimating the sentiment polarity of the aspect [7].
- *MemNet* uses a multi-hop memory network to direct attention on specific aspect words in a sentence using word and position attention [65].
- *ATAE-LSTM* Attention-based LSTM network with aspect embedding. It uses aspect embeddings to calculate the attention weights for LSTM hidden states. Prediction is performed using the weighted sum of the hidden states [6].
- *RAM* Recurrent attention network architecture based on MemNet. Suggested the BiLSTM and GRU to generate the context representations by integrating the outputs of many attention units [8].
- *IAN* This approach provides an interactive attention technique for modeling target and contextual representations independently [66].
- *AOA* employs an attention-over-attention model to learn the relationship between aspect and contextual words and focus on the essential parts of sentences [43].
- *MGAN* Multi grained attention network was designed to capture the word level interaction between the aspect and context. Also, construct the final context representation by leveraging two attention processes. Aspect alignment loss was created to study the relationships between the aspect and its context thoroughly [67].
- *TNet* extracts significant characteristics from altered word representations generated by a bi-directional RNN layer using a CNN layer rather than attention [68].
- *CABASC* This method offered a content attention approach for ABSA that can capture critical information about provided aspect from a global perspectives while also taking into account the sequence of the words and their associations [69].
- *Two-stage paradigm* The first stage explicit relationship between aspect words and their context sequence words through the application of a positional attention mechanism. The second stage investigates how to model many aspect words simultaneously in a comment text using a positional attention method [70].
- *MAN* uses multiple attention methods to determine the sentiment polarity of various aspect phrases. It reduces the model's complexity by employing a transformer encoder rather than an RNN-based neural network. Additionally, an inter-

active attention layer, is introduced to collect differentially grained interacting knowledge between aspect and context word [71].

4.3.3 Syntax-based methods

- *PBAN* focuses on the position information of aspect terms and mutually models the relationship between aspect terms and phrases [30].
- *ASGCN-DG* is the same as ASGCN-DT, except the graph's adjacency matrix is un-directional [28].
- *SK-GCN* Syntax knowledge are exploited via graph convolution networks (GCN) modeling the syntactical dependency trees and common-sense information graphs [72].
- *R-GAT* reconfigures and prunes an ordinary dependency parse tree to build an aspect-based dependency tree structure rooted in a target aspect. It then encodes the tree structure for emotion prediction using a relational graph attention network [73].
- *BiGCN* Hierarchical syntactic and lexical graphs, create a concept hierarchy on the syntactic and lexical graphs to distinguish distinct forms of dependence connections or lexical word pairings [74].

4.3.4 BERT-based methods

- *BERT* model that accepts the input "[CLS] sentence [SEP] aspect [SEP]" [75].
- *BERT-ADA* suggested a two stage technique based on a pre-trained BERT language model that includes self-supervised domain-specific BERT model fine-tuning and cross-domain analysis [76].
- *AEN-BERT* a network of attention encoders was proposed to measure the context self-attention and the aspect external attention to the context representation of a particular aspect and introduce label smoothing regularization [77].
- *MGMD* is a BERT-based multi-source data fusion approach which acquires sentiment information from a variety of sources [78].
- *PyABSA* open source framework comprises the capabilities of ATE, aspect sentiment classification, and text classification, while its modular architecture allows it to be modified to numerous ABSA subtasks [79].
- *SGGCN+BERT* In this study, the representation vectors of the aspect terms are used to generate graph-based deep learning model gate vectors used toward the aspect terms [80].
- *Sentic GCN-BERT* LSTM approach uncased pre-trained BERT. In other words, the GCN layers in sentic GCN-BERT use the sentence's outputs encodings and the accompanying graph as input for building graph representation [81].

4.4 Model comparison

For baseline methods, LSTM-based methods perform better because they are capable of producing semantic feature representations without requiring manual feature engineering; By factoring in the importance of aspects, TD-LSTM, AE-LSTM, LSTM-based method; ATAE-LSTM performs poorly due to its reliance on a simple LSTM structure. IAN outperforms ATAE-LSTM by adding two information channels to last stage sentence representation. However, MemNet outperforms IAN because it employs several computational layers, and numerous attention layers are more effective than individual attention layers. RAM and TNet beat these established benchmarks. TNet approach improves the accuracy and f-1 score on mentioned datasets. Because TNet incorporates many CPT module and CNN layer capable of extracting non-sequential and long-distance information. AEN-BERT does not rely solely on pre-trained BERT for initials word embedding; it uses numerous attention mechanisms. As a result, AEN-BERT outperforms BERT-SPC. Finally, because MGMD derives sentiment information from various sources, it outperforms baseline models. Attention-based system performance for the following reasons: When compared to the state-of-the-art methods, our MAPA BiLSTM BERT obtained an result of 80.78% and 87.33% on the SemEval2014 (laptop & Restaurant) datasets, 75.31% for Twitter review and 85.73% for MAMS dataset respectively; When compared to the state-of-the-art methods, our proposed approach achieves an better result. The performance of the proposed approach shows that our model produces superior results than RGAT-BERT and SKGCN-BERT in a majority of the case. The proposed approach performs better than RGAT-BERT on Laptop, Restaurant, and MAMS in terms of accuracy by 1.68%, 1.36%, and 1.13%, respectively. This observation validates our model that fine-grained relationship information is beneficial. When combined with context-position attention and relation-aware feature aggregation, accuracy improved by 1.23%, 1.73%, 0.96%, and 1.42% on Restaurant, Laptop, Twitter, and MAMS, respectively. The improvement in F1 score amounts to 1.73%, 2.56%, 0.98%, and 2.61%, respectively. Compared with the Twitter review case, RGAT-BERT performs better than the proposed approach. The proposed approach comparison with all four datasets results obtained in Table 2 indicates that our model performed better in most cases.

4.5 Performance analysis

As previously stated, the sentiment of an aspect term is dependent on the context words immediately preceding it. The majority of prior work treated the average of all aspect representations of words as such interpretation of the aspect, which did not accurately represent the aspects. It resulted in suboptimal aspect-based representations generated via the attention-based method for sentiment classification. Existing approaches are primarily concerned with optimizing neural network structures. We overlook the reality that insufficient training corpora severely limits these model performances. Also, the majority of previous attention-based approaches employ LSTM to encode sentences, our model generates aspect-specific sentence

Table 2 Result analysis and performance comparison on laptops reviews, restaurant reviews, Twitter reviews, and MAMS datasets

Models	Laptop		Restaurant		Twitter		MAMS	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
SVM-Feature	70.50	–	80.20	–	–	–	–	–
TD-LSTM	71.83	68.43	78.00	68.43	70.80	69.00	–	–
IAN	72.10	–	78.61	–	–	–	76.16	–
RAM	74.49	71.35	80.23	70.82	–	–	–	–
ATAE-LSTM	68.70	–	77.20	–	–	–	–	–
AOA	72.62	67.52	79.97	70.42			77.25	–
Two-stage paradigm	73.10	–	80.10	–	–	–	–	
PBAN	74.12	–	81.16	–	–	–	–	-
AT-LSTM	77.43	73.70	81.63	71.94	73.51	71.93		
GAT	73.04	68.11	78.21	67.17	71.67	70.13		
CABASC	75.04	–	80.89	–	–	–	–	-
MemNet	70.33	64.09	78.16	65.83	69.65	67.68	–	–
BiGCN	74.59	71.84	81.97	73.48	74.16	73.35	–	–
ASGCN-DG	74.14	69.24	80.86	72.19	72.15	70.40	–	–
TNET	76.54	71.75	80.69	71.27	74.90	73.60	–	–
MAN	78.13	73.20	84.38	71.31	–	–	–	–
SK-GCN	79.00	75.57	83.48	75.19	–	–	–	–
BERT	77.59	73.28	84.11	76.68	–	-		
Transformer	80.78	72.10	74.09	69.42	72.78	70.23	79.63	78.92
R-GAT + BERT	79.73	75.50	85.50	79.33	**75.28**	**75.25**	84.02	83.03
AEN-BERT	79.93	76.31	83.12	73.76	73.55	72.10	–	–
MGMD	79.18	74.86	86.16	78.83	–	–	–	-
DGEDT-BERT	79.80	75.60	86.30	80.0	–	–	–	–
PyABSA	80.10	78.11	–	–	–	–	–	–
SGGCN+BERT	80.08	**80.01**	**87.20**	**82.50**	–	–	–	–
Sentic GCN-BERT	**82.12**	79.05	86.92	81.03	–	–	–	–
MAPA BiL-STM-BERT	**80.78**	**79.68**	**87.33**	**80.17**	**75.31**	**74.11**	**85.73**	**85.45**

The best two result are highlighted in bold

representations. Compared to some previous attention-based mechanism mentioned in Sect. 4.3.2, BiLSTM-BERT can model context and has higher computational efficiency. In the proposed work, BERT aspect attention embedding based on nearest word context meaning solves the single word problem.

As noted per previous it found that a single layer attention cannot extract all sentiment relationships included in the sentence. Multiple processing layers are required to process the complex semantic information contained in words. With more attention layers, the accuracy and macro-f1 score increase. This demonstrates that stacking BiLSTM and BERT attention layers can equip the model with more complicated sentiment relationships between words. This should be recognized that several unusual circumstances

Table 3 Effect of with and without content attention (CA), position attention (PA) comparison on laptops reviews, restaurant reviews (Semeval2014), Twitter reviews, and MAMS dataset

Models	Laptop		Restaurant		Twitter		MAMS	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
MAPA BiLSTM-BERT without CA-PA	76.11	72.06	85.18	79.11	74.07	70.17	76.33	77.25
MAPA BiLSTM-BERT with CA	76.98	72.23	85.38	80.68	74.39	70.16	79.25	76.02
MAPA BiLSTM-BERT with PA	77.94	76.17	86.10	80.16	74.01	70.12	83.82	79.57
MAPA BiLSTM-BERT with CA-PA	80.78	76.68	87.33	80.17	75.31	74.11	85.73	85.45

Table 4 The proposed approach training and testing time in second comparison on laptops reviews, restaurant reviews, and Twitter reviews

Models	Laptop		Restaurant		Twitter	
	Train	Test	Train	Test	Train	Test
MAPA BiLSTM-BERT without CA-PA	631.47	7.54	1293.04	10.49	1666.14	7.17
MAPA BiLSTM-BERT with CA	677.18	7.59	1309.08	11.54	1701.39	7.56
MAPA BiLSTM-BERT with PA	678.14	8.01	1315.10	11.59	1754.01	7.29
MAPA BiLSTM-BERT with CA-PA	679.58	8.18	1326.33	12.07	1778.31	7.33

constitute exceptions. As a result, we perform studies to determine the influence of layer count on parallel fusion intra level attention. To ascertain the rationality of each element in the MAPA BiLSTM BERT, we create variations of the MAPA BiLSTM BERT model and undertake ablation experiments. MAPA BiLSTM BERT denotes the proposed model using only content attention and position attention in the content attention layers and position attention layer, respectively. The proposed method various variations result obtained indicated in Table 3 shows the significance of the position attention and content attention approach and time required in Table 4, indicating negligible variations in timing. Table 2 summarises the findings. In our work proposed MAPA BiLSTM BERT model with several version modifications mentioned in Table 3. The proposed approach consider multiple aspects within sentences with closed context words in aspect with respective position attention mechanism.

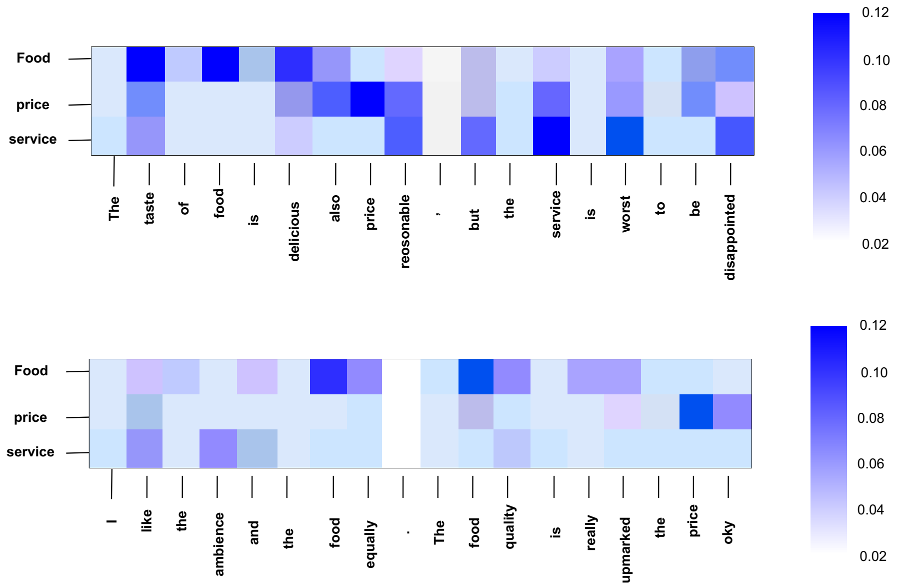

Fig. 3 Case study: examples learned by our proposed approach are reported. The dark blue color indicates more attention towards the word. Light blue indicates less attention respectively (color figure online)

4.6 Case study

To better understand why our proposed model outperforms simple design baselines, we choose examples from the datasets and visualize attention heat maps based on the trained model. As discussed previously, ABSA's primary issue is to discern distinct sentiment polarity for different parts inside the same sentence. Our case study examples show that dark blue indicates more attention towards the word, light blue indicates less attention, and darker colors represent strong sentiment. To illustrate the proposed methodology, we create heat maps for a sentence that includes three aspects, namely "food," "price," and "service." In the opinion-based sentence shown in Fig. 3, indicate closer the context word shows more impact. For example, the food quality indicates positive intent, and the price is satisfactory, while the restaurant is criticized for its poor service. Our case study examples show that dark blue indicates more attention towards the word, light blue indicates less attention, and darker colors represent strong sentiment. This demonstrates that our model is capable of capturing the relationship between a multiple aspect word and a sentiment word in the same sentence. Our approach used the CA-PA mechanism, so one of the limitations of our work is assigning the weight under various scenario-challenging than prior syntax-free work because of the modeling of structural dependencies. Furthermore, our model contains more computations than earlier work.

To highlight qualitatively how the proposed MAPA BiLSTM BERT model enhances performance when predicting aspect-based sentiment polarity, we depict the attention weights using two representative examples (i.e., a single-aspect instance and a multiple-aspects instance). As shown Fig. 3, illustrates the results accompanying sentiment terms. The model can pay much more attention to crucial effective words when extracting sentiment features associated with specific attributes. The multiple-aspects example demonstrates that the proposed method may distinguish between distinct aspects and focus on distinct contextual affective terms based on those distinctions.

5 Conclusion

In this work, we propose a framework for extracting aspect-level sentiment polarity from short text using multiple aspect attention. We proposed MAPA mechanisms utilizing the BiLSTM and BERT. Long-term memory information has been stored using memory network components, whereas sequence information and long-distance reliance are captured using BiLSTM. A bidirectional multi-aspect position attention mechanism determines the relationship between the context word and aspect term. The proposed approach investigates multiple aspects within a single sentence. To be more precise, the BiLSTM model is used to process individual text components sequentially, and the BERT model processes multiple aspects simultaneously. Finally, comprehensive trials on Semeval 2014 datasets (laptops reviews, restaurant reviews), Twitter review and MAMS dataset reveal that our MAPA BiLSTM-BERT approach, which models several aspects concurrently, can reduce the disruption between the multiple aspects and direct distinct aspects attention to different portions of the sentence. The proposed work is helpful for aspect-level sentiment classification that can examine consumer feedback by correlating distinct sentiments with various aspects of a product or service.

Author contributions MW: Data curtion, conceptualization, validation, implementations, methodology, writing—original draft. AA: Methodology, conceptualization, editing, review, validation, supervision. ACSR: Review, validation, editing, methodology, supervision, conceptualization.

Funding Not applicable.

Data availability Data used in the research article are available from a reasonable research point of view request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable

References

1. Wankhade M, Rao ACS, Kulkarni C (2022) A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55:5731–5780
2. Majumder N, Bhardwaj R, Poria S, Gelbukh A, Hussain A (2020) Improving aspect-level sentiment analysis with aspect extraction. Neural Comput Appl 1–11
3. Wu T, Tang S, Zhang R, Cao J, Zhang Y (2020) Cgnet: a light-weight context guided network for semantic segmentation. IEEE Trans Image Process 30:1169–1179
4. Nazir A, Rao Y, Wu L, Sun L (2020) Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. IEEE Trans Affect Comput
5. Xue W, Li T (2018) Aspect based sentiment analysis with gated convolutional networks. arXiv preprint arXiv:1805.07043
6. Wang Y, Huang M, Zhu X, Zhao L (2016) Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp 606–615
7. Tang D, Qin B, Feng X, Liu T (2015) Effective LSTMs for target-dependent sentiment classification. arXiv preprint arXiv:1512.01100
8. Chen P, Sun Z, Bing L, Yang W (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 452–461
9. Chen P, Sun Z, Bing L, Yang W (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 452–461
10. Li X, Bing L, Lam W, Shi B (2018) Transformation networks for target-oriented sentiment classification. arXiv preprint arXiv:1805.01086
11. Rockt aschel T, Grefenstette E, Hermann KM, Kocisky T, Blunsom P (2015) Reasoning about entailment with neural attention. arXiv preprint arXiv:1509.06664
12. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473
13. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900
14. Wankhade M, Annavarapu CSR, Verma MK (2021) CBVoSD: context based vectors over sentiment domain ensemble model for review classification. J Supercomput 1–37

15. Wang Y, Huang M, Zhu X, Zhao L (2016) Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp 606–615
16. Wang S, Mazumder S, Liu B, Zhou M, Chang Y (2018) Target-sensitive memory networks for aspect sentiment classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 957–967
17. Kiritchenko S, Zhu X, Cherry C, Mohammad S (2014) NRC-Canada-2014: detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp 437–442
18. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900
19. Guo X, Zhang G, Wang S, Chen Q (2020) Multi-way matching based fine-grained sentiment analysis for user reviews. Neural Comput Appl 32(10):5409–5423
20. Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Hum lang Technol 5(1):1–167
21. Yang M, Qu Q, Shen Y, Lei K, Zhu J (2020) Cross-domain aspect/sentiment-aware abstractive review summarization by combining topic modeling and deep reinforcement learning. Neural Comput Appl 32(11):6421–6433
22. Ahmed M, Chen Q, Li Z (2020) Constructing domain-dependent sentiment dictionary for sentiment analysis. Neural Comput Appl 32(18):14719–14732
23. Chen P, Sun Z, Bing L, Yang W (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 452–461
24. Jin Z, Yang Y, Liu Y (2020) Stock closing price prediction based on sentiment analysis and LSTM. Neural Comput Appl 32(13):9713–9729
25. Tang D, Qin B, Feng X, Liu T (2015) Effective LSTMs for target-dependent sentiment classification. arXiv preprint arXiv:1512.01100
26. Wu H, Zhang Z, Shi S, Wu Q, Song H (2022) Phrase dependency relational graph attention network for aspect-based sentiment analysis. Knowl Based Syst 236:107736
27. Zhang B, Hu Y, Xu D, Li M, Li M (2022) SKG-learning: a deep learning model for sentiment knowledge graph construction in social networks. Neural Comput Appl 1–20
28. Zhang C, Li Q, Song D (2019) Aspect-based sentiment classification with aspect-specific graph convolutional networks. arXiv preprint arXiv:1909.03477
29. Li X, Lam W (2017) Deep multi-task learning for aspect term extraction with memory interaction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 2886–2892
30. Gu S, Zhang L, Hou Y, Song Y (2018) A position-aware bidirectional attention network for aspect-level sentiment analysis. In: Proceedings of the 27th International Conference on Computational Linguistics, pp 774–784
31. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900
32. Gu S, Zhang L, Hou Y, Song Y (2018) A position-aware bidirectional attention network for aspect-level sentiment analysis. In: Proceedings of the 27th International Conference on Computational linguistics, pp 774–784
33. Xu L, Li H, Lu W, Bing L (2020) Position-aware tagging for aspect sentiment triplet extraction. arXiv preprint arXiv:2010.02609
34. Gu S, Zhang L, Hou Y, Song Y (2018) A position-aware bidirectional attention network for aspect-level sentiment analysis. In: Proceedings of the 27th International Conference on Computational Linguistics, pp 774–784
35. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900
36. Jiang L, Yu M, Zhou M, Liu X, Zhao T (2011) Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp 151–160
37. Kiritchenko S, Zhu X, Cherry C, Mohammad S (2014) NRC-Canada-2014: detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp 437–442

38. Dong L, Wei F, Tan C, Tang D, Zhou M, Xu K (2014) Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (volume 2: Short Papers), pp 49–54
39. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 168–177
40. Ren F, Feng L, Xiao D, Cai M, Cheng S (2020) Dnet: a lightweight and efficient model for aspect based sentiment analysis. Expert Syst Appl 151:113393
41. Wankhade M, Rao ACS (2022) Bi-directional lstm attention mechanism for sentiment classification. In: 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), pp 1–6. IEEE
42. Ma D, Li S, Zhang X, Wang H (2017) Interactive attention networks for aspect-level sentiment classification. arXiv preprint arXiv:1709.00893
43. Huang B, Ou Y, Carley KM (2018) Aspect level sentiment classification with attention-over-attention neural networks. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp 197–206. Springer
44. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. arXiv preprint arXiv:1406.1078
45. Wankhade M, Rao ACS (2022) Opinion analysis and aspect understanding during COVID-19 pandemic using BERT-Bi-LSTM ensemble method. Sci Rep 12(1):17095
46. Cai Y, Huang Q, Lin Z, Xu J, Chen Z, Li Q (2020) Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: a multi-task learning approach. Knowl Based Syst 203:105856
47. Wei J, Liao J, Yang Z, Wang S, Zhao Q (2020) Bilstm with multi-polarity orthogonal attention for implicit sentiment analysis. Neurocomputing 383:165–173
48. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473
49. Mnih V, Heess N, Graves A et al. (2014) Recurrent models of visual attention. In: Advances in Neural Information Processing Systems, pp 2204–2212
50. Chen P, Sun Z, Bing L, Yang W (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 452–461
51. Wang X, Chen X, Tang M, Yang T, Wang Z (2020) Aspect-level sentiment analysis based on position features using multilevel interactive bidirectional GRU and attention mechanism. Discrete Dynamics in Nature and Society 2020
52. Lou Y, Zhang Y, Li F, Qian T, Ji D (2020) Emoji-based sentiment analysis using attention networks. ACM Trans Asian Low Resour Lang Inf Process (TALLIP) 19(5):1–13
53. Xi C, Lu G, Yan J (2020) Multimodal sentiment analysis based on multi-head attention mechanism. In: Proceedings of the 4th International Conference on Machine Learning and Soft Computing, pp 34–39
54. Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR (2021) ABCDM: an attention-based bidirectional CNN–RNN deep model for sentiment analysis. Future Gener Comput Syst 115:279–294
55. Cambria E. Li Y, Xing FZ, Poria S, Kwok K (2020) Senticnet 6: ensemble application of symbolic and subsymbolic ai for sentiment analysis. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp 105–114
56. Ribani R, Marengoni M (2019) A survey of transfer learning for convolutional neural networks. In: 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pp 47–57. IEEE
57. Xu H, Li S, Hu R, Li S, Gao S (2018) From random to supervised: a novel dropout mechanism integrated with global information. arXiv preprint arXiv:1808.08149
58. Li X, Bing L, Lam W, Shi B (2018) Transformation networks for target-oriented sentiment classification. arXiv preprint arXiv:1805.01086
59. Zhao A, Yu Y (2021) Knowledge-enabled BERT for aspect-based sentiment analysis. Knowl Based Syst 227:107220
60. Tay Y, Tuan LA, Hui SC (2018) Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32

61. Mehta S, Islam MR, Rangwala H, Ramakrishnan N (2019) Event detection using hierarchical multi-aspect attention. In: The World Wide Web Conference, pp 3079–3085
62. Jiang Q, Chen L, Xu R, Ao X, Yang M (2019) A challenge dataset and effective models for aspect-based sentiment analysis. 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), pp 6280–6285
63. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? sentiment classification using machine learning techniques. arXiv preprint cs/0205070
64. Kiritchenko S, Zhu X, Cherry C, Mohammad S (2014) Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp 437–442
65. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900
66. Ma B, Yuan H, Wu Y (2017) Exploring performance of clustering methods on document sentiment analysis. J Inf Sci 43(1):54–74
67. Fan F, Feng Y, Zhao D (2018) Multi-grained attention network for aspect-level sentiment classification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp 3433–3442
68. Li X, Bing L, Lam W, Shi B (2018) Transformation networks for target-oriented sentiment classification. arXiv preprint arXiv:1805.01086
69. Liu Q, Zhang H, Zeng Y, Huang Z, Wu Z (2018) Content attention model for aspect based sentiment analysis. In: Proceedings of the 2018 World Wide Web Conference, pp 1023–1032
70. Ma X, Zeng J, Peng L, Fortino G, Zhang Y (2019) Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis. Future Gener Comput Syst 93:304–311
71. Xu Q, Zhu L, Dai T, Yan C (2020) Aspect-based sentiment classification with multi-attention network. Neurocomputing 388:135–143
72. Zhou J, Huang JX, Hu QV, He L (2020) SK-GCN: modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. Knowl Based Syst 205:106292
73. Wang K, Shen W, Yang Y, Quan X, Wang R (2020) Relational graph attention network for aspect-based sentiment analysis. arXiv preprint arXiv:2004.12362
74. Zhang M, Qian T (2020) Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 3540–3549
75. Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
76. Rietzler A, Stabinger S. Opitz P, Engl S (2019) Adapt or get left behind: domain adaptation through BERT language model finetuning for aspect-target sentiment classification. arXiv preprint arXiv:1908.11860
77. Song Y, Wang J, Jiang T, Liu Z, Rao Y (2019) Attentional encoder network for targeted sentiment classification. arXiv preprint arXiv:1902.09314
78. Chen F, Yuan Z, Huang Y (2020) Multi-source data fusion for aspect-level sentiment classification. Knowl Based Syst 187:104831
79. Yang H, Li K (2022) PyABSA: open framework for aspect-based sentiment analysis. arXiv preprint arXiv:2208.01368
80. Veyseh APB, Nour N, Dernoncourt F, Tran QH, Dou D, Nguyen TH (2020) Improving aspect-based sentiment analysis with gated graph convolutional networks and syntax-based regulation. arXiv preprint arXiv:2010.13389
81. Liang B, Su H, Gui L, Cambria E, Xu R (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowl Based Syst 235:107643