Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **1**, March **2025**

Enhanced Aspect-level Sentiment Analysis of User Reviews Using RoBERTa and Data Augmentation

Li-Zhu Ye[1,2,*]

[1]School of Graduate Studies,
Management and Science University, 40100 Shah Alam, Selangor, Malaysia
[2]School of Information Technology and Engineering,
Guangzhou College of Commerce, Guangzhou 511363, P. R. China
lizhuye2018@gcc.edu.cn

Md Gapar Md Johar[a]

[a]Software Engineering and Digital Innovation Center,
Management and Science University, 40100 Shah Alam, Selangor, Malaysia
mdgapar@msu.edu.my

Mohammed Hazim Alkawaz[b]

[b]Department of Computer Science,
College of Education for Pure Science, University of Mosul,
Mosul 41002, Nineveh, Iraq
mohammed.ameen@uomosul.edu.iq

[*]Corresponding author: Li-Zhu Ye

Received December 9, 2024, revised January 8, 2025, accepted January 16, 2025.

Abstract. *Aspect-level sentiment analysis of user reviews is essential to the area of multi-sentiment analysis. Accurately recognizing the many sentiment polarities present in several elements of a single phrase is its fundamental function. Existing models, however, frequently contain flaws. They frequently struggle with the issue of either aspect-level semantics, sentence-wide semantics, or both being absent. Furthermore, throughout the analysis process, aspect terms' positional variations and contextual correlations are frequently overlooked. The precision of sentiment analysis is impacted by these flaws. To increase sentiment classification performance, this study suggests an aspect-level sentiment analysis model that blends data augmentation technology with RoBERTa.It is evident from trials on particular datasets that this model outperforms the AE-LSTM model in terms of accuracy rate, marking an advancement in the field of sentiment analysis.*
Keywords: Aspect-level Sentiment Analysis; RoBERTa; data augmentation;

1. Introduction.

Text sentiment evaluation can routinely analyze and apprehend thoughts and emotions in text, enabling humans to successfully draw close the emotional developments at the back of massive quantities of textual data. This technological know-how is tremendously treasured for commercial enterprise decision-making as it can furnish insights based totally on actual customer feedback, assisting groups to optimize products, enhancing carrier quality, and regulating market strategies. Coarse-grained sentiment evaluation can decide the sentiment tendency of archives and sentences, whilst fine-grained sentiment evaluation can decide the sentiment tendency of particular factors in sentences. This article researches element degree sentiment evaluation (ALSA), which includes finer-grained sentiment classification of objects or entities in the corpus

[1, 2], which can higher meet the wishes of corporations [3]. For example, in the remark "The taste of the dishes is super and the fees are reasonable, however, the mindset of the personnel is disappointing.", the thing and its corresponding emotional polarity are "taste of food" positive, "price" positive, and "service" negative, respectively. In regular laptop mastering methods, function extraction is based totally on a guide diagram and requires expert information and trip [4], whilst the coaching and optimization of classifiers are primarily based on basic machine learning (ML) algorithms such as naive Bayes, etc. Although these techniques can obtain widespread outcomes in the subject of sentiment analysis, their effectiveness is significantly restrained by using first-rate features and frequently requires exceptional guide characteristic determination and parameter tuning to obtain the greatest overall performance [5].

Currently, with the improvement of deep mastering technology, though current strategies have made some growth in this field, There are nonetheless some limitations. Firstly, regular BERT fashions dealing with unregistered users' impact are now not suitable when the usage of OOV and nested entities. The full name of OOV is out-of-vocal, which means "Out-Of-Vocabulary" in Chinese. Secondly, attention mechanism and though the aggregate of CNN can enhance the interplay awareness between entities, it is no longer positive in coping with complicated tasks. The nested shape is nonetheless inadequate when combined [6, 7, 8]. Finally, current records augmentation strategies are proposed. There is nonetheless room for enhancement in phrases of mannequin robustness.

To overcome the challenges previously highlighted, this study introduces a cutting-edge aspect-level sentiment analysis model. The model innovatively incorporates dynamic word embeddings derived from RoBERTa and augments these with embeddings for location information and part-of-speech features [10, 11]. Furthermore, it integrates data augmentation strategies and bidirectional long short-term memory networks to bolster the detection of nuanced emotional sentiments within sentences. The primary innovations and contributions of this research are as follows:

(1) A data augmentation module has been developed to generate adversarial samples and perturb the original input, thereby enhancing the model's generalization capabilities through adversarial training. To achieve data augmentation, increase the diversity of input samples, bolster the model's robustness, and improve its ability to recognize ambiguous entity boundaries, a technique involving the random substitution of synonyms or hypernyms for non-stop words has been implemented.

(2) Enhanced semantic insights are obtained through the utilization of RoBERTa for generating word embeddings with more robust representational capabilities. Additionally, the model's performance is amplified by a multi-attention mechanism that leverages positional information, leading to a more profound understanding of the content correlation between the text and the target vocabulary. The multi-head self-attention approach effectively identifies the correlations among aspect words in commentary texts, thereby refining the understanding and assessment of aspect emotions.

(3) In this context, a multi-head self-attention mechanism is deployed to distill keyword information from aspects and to assess their contextual influence within the broader text. This mechanism assigns each head the role of an independent attention unit, enabling the learning of diverse sentence representations across multiple representational subspaces. By treating aspects and comment texts with equal importance, the model introduces a multi-head self-attention mechanism that interactively learns a range of semantic spatial information representations for sentences and aspects, thereby capturing a richer array of emotional features. Introducing a multi head self attention mechanism to interactively learn different semantic spatial information representations of sentences and

aspects, modeling aspects and comment texts equally, capturing richer emotional feature information.

(4) An improved RoBERTa model, which performs better than its BERT equivalent and acts as the word embedding layer, supports the model. It is stacked from the Transformer architecture's Encoder component. The BERT model has been improved by removing the NSP task, using bigger corpora for training, increasing batch sizes, and extending time steps. With every sequence input, a new masking strategy is generated by the dynamic masking mechanism. The embedding layer of our model is formed by the pre-trained RoBERTa model, which can be accessed through the Hugging Face platform. The embedding vectors are then obtained from this layer.

2. Relevant technologies.

2.1. Attention mechanism.

An essential tool in computer science, attention mechanisms are particularly vital in machine learning (ML) and natural language processing (NLP) applications [9]. Simulating human visual and cognitive processes enables models to analyze sequential data by concentrating on distinct informational components, improving the representation of the connections and significance among data. In natural language processing tasks, attention methods are employed to intelligently highlight the most pertinent information in the input sequence while lowering the influence of irrelevant information. This greatly improves the model's performance and accuracy. Each input element is given a weight by the attention mechanism during the computation process, indicating the content that the model should focus on when predicting the output [9]. The attention method in sentiment analysis tasks can automatically focus on terminology that significantly affects the sentence's sentiment, accurately determining the sentence's sentiment polarity. In sentiment analysis tasks, the attention mechanism can automatically focus on vocabulary that has a significant impact on the sentiment of the sentence, thereby accurately identifying the sentiment polarity of the sentence.

2.2. Long Short-Term Memory Network (LSTM).

A bidirectional long short-term memory (BiLSTM) network is employed in this article. Utilizing the surface's semantic information (BiLSTM) to extract text and remark text. To obtain a more thorough understanding of sequence data and to facilitate deep-level feature extraction of input context information, it can concurrently use the past and future. This allows for the building of context dependencies using forward LSTM and backward LSTM models. Sentences are processed in reverse order by backward LSTM and from left to right by forward LSTM [10]. Therefore, through BiLSTM, two results can be obtained: different hidden representations, then move the word forward and backward in the hidden state. Connect the hidden states of text to obtain a more comprehensive semantic understanding of information. Assuming that $X^c = (x_1^c, x_2^c, \ldots, x_n^c)$ represents input related terminology. After using the RoBERTa model, the sentence embedding is obtained as $E^c = (e_1^c, e_2^c, \ldots, e_n^c)$. Embedding aspects are represented as $E^a = (e_1^a, e_2^a, \ldots, e_m^a)$, among them $E^c \in \mathbb{R}^{d\times n}$ and $E^a \in \mathbb{R}^{d\times m}$, where d is the dimension of word embedding. Let the position embedding of each word be P_t, part of speech embedding as pos_t, it is necessary to concatenate data augmentation, word embedding, part of speech embedding, and positional embedding of the sentence, and the result of the concatenation is $z_t' = [E^c; P_t^c; pos_t^c]$. Then, input z_t' into the BiLSTM on the text to obtain the position operation. As shown in Figure 1, the hidden state h_t^a of aspect words can be obtained in the left BiLSTM. The specific calculation formula is:

$$\bar{h}_t^c = \mathrm{LSTM}(z_t^c) \tag{1}$$

$$\hat{h}_t^c = \mathrm{LSTM}(z_t^c) \tag{2}$$

$$h_t^c = \begin{bmatrix} \rightarrow & \leftarrow \\ h_t^c, & h_t^c \end{bmatrix} \tag{3}$$

Where $\bar{h}_t^c$ and $\hat{h}_t^c$ represent the forward and backward LSTM's hidden states at time t respectively, and h_t^c is the result of concatenating the two states mentioned above.

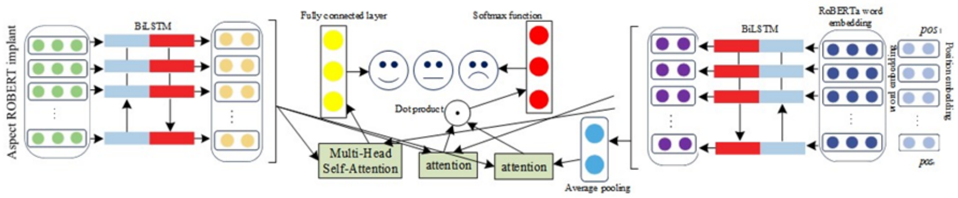

FIGURE 1. Model architecture diagram

Similarly, the hidden state of aspect words can be obtained through BiLSTM:

$$h_t^a = [\bar{h}_t^a; \hat{h}_t^a] \tag{4}$$

2.3. **Word Embedding.** Word embedding is one of the key technologies in NLP that maps discrete vocabulary to a continuous vector space, enabling computers to better understand and process textual data. It encodes the semantic information of words into vector representations, making similar words closer in the vector space, thereby capturing the semantic relationships and contextual meanings between words. The transformation of low-dimensional vector space enables a more effective representation of semantic and syntactic connections in high-dimensional vocabulary space, reducing computational complexity while preserving the main semantic information. By using multiple embedding techniques, embedding vectors with different expressive abilities can be obtained.

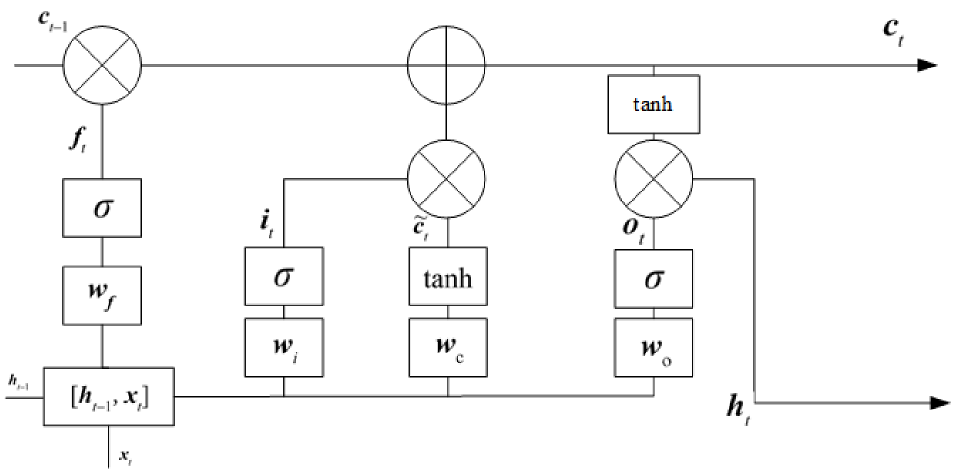

FIGURE 2. LSTM unit structure

where h_t is the hidden state of time t; c_{t-1} and c_t are the cellular state at the point $t-1$ and t, respectively; w_f, w_i, w_c and w_o are the trainable weight matrix; tanh and σ are the time input tanh and non-linear activation functions; o_t shows the output gate; $\tilde{c}_t$ shows the update gate; f_t is the forgotten gate.

3. **The model's general architecture.** As illustrated in Figure 1, an aspect-level sentiment analysis model utilizing RoBERTa and positional characteristics will be further explained in this section. First, dynamic word vectors will be created using a pre-trained language model called RoBERTa. Second, the model will be able to use part of speech to enhance contextual representation and focus more on words that are closer to the aspect words by combining the interaction notions between aspects and context as well as aspect location information. The model can capture a variety of semantic traits by utilizing a multi-head self-attention method, which guarantees the acquisition of rich semantic content. The emotional polarity of particular phrase elements is then determined by connecting the emotional features that were acquired in various methods and sending them to the fully linked layer for classification using the Softmax algorithm.

3.1. **RoBERTa embedding.** In this article, the improved RoBERTa model with better performance based on the BERT model is used as the word embedding layer, which is also stacked from the Encoder part of the Transform model. The main improvements to the BERT model are as follows: removing the NSP (next sentence prediction) task; using more corpus and training with larger batch sizes and time steps; adopting a dynamic masking mechanism, which generates a new masking method every time a sequence is input into the model [11]. To retrieve the embedding vector and a pre-trained RoBERTa model as the model's embedding layer, this article makes use of the Hugging Face platform.

3.2. **Location features.** First, establish a position index sequence with a length equal to the phrase, in which each word's position index corresponds to the aspect word's location. A word's position index is marked as "0" when it is an aspect word, t_{start} and t_{end} the start and end indexes of the aspect are represented by two indications. A positional index sequence is produced by determining the relative distance between each word in the sentence and the aspect word. In "The food is great, but the service is like a dream", for example, the positional index sequence for the aspect word "food" is $[1, 0, 1, 2, 3, 4, 5]$, but the positional index sequence for the aspect word "service" is $[4, 3, 2, 1, 0, 1, 2]$. Various word position embedding representations can be produced by querying the position embedding matrix $p \in \mathbb{R}^{d\times n}$ [13]. The PyTorch framework is used to randomly initialize position embedding, which is then continuously trained to provide more accurate representations. The precise formula for calculation is

$$p_c = \begin{cases} |c - t_{\text{start}}|, & c < t_{\text{start}} \\ 0, & t_{\text{start}} \le c \le t_{\text{end}} \\ |c - t_{\text{end}}|, & c > t_{\text{end}} \end{cases} \tag{5}$$

where c is a word's position index in the sentence; p_c indicates the word's location with aspect words.

3.3. **BiLSTM.** In this article, the purpose of using a BiLSTM network is to extract semantic information from comment texts and their specific aspects [14]. It can simultaneously utilize past and future contextual information to completely grasp the characteristics of sequence data and can perform deep-level feature extraction on input contextual information. It models contextual dependencies using forward LSTM and reverse LSTM. The sentence is processed from front to back by forward LSTM, but the backward LSTM processes them in reverse order. Therefore, by using BiLSTM, two different hidden representations can be obtained, and subsequently, by concatenating the forward and backward hidden states of each word, richer semantic information about the text can be obtained. Assuming the sentence represents the input, indicating the input-related terminology. The RoBERTa model is used to obtain the aspect and sentence embeddings, where d is the word embedding dimension. The concatenation result is, and then input it into the BiLSTM on the right to obtain the position and part of speech context hidden state of each word. ";" represents the vector concatenation operation. The hidden state of aspect words can be obtained in the left BiLSTM.

3.4. **Data augmentation module.** To increase the diversity of current datasets and, consequently, the generalization capacity of models, data augmentation techniques are commonly used in machine learning to improve model performance. Data augmentation can greatly increase the model's recognition efficiency for unusual or new samples in named entity identification tasks. This article's data augmentation module is primarily divided into two sections. First, a synonym dictionary is utilized to improve a few samples in the original data during the model's data preprocessing phase. Adversarial training techniques are then added to the model's training phase to improve the diversity of data, slightly alter the data features, and maximize the model's resilience and recognition efficiency through backpropagation.

(1) Synonym enhancement. This article employs Synonym Replacement (SR) as a data augmentation technique, choosing n words from the non-entity portion of the sentence that does not belong to the stop word set at random and replacing them with synonyms chosen at random. In addition to improving the data's semantic coverage, this technique helps the model retain its recognition effectiveness even as natural language changes. As a result, the corpus may be effectively increased, raising the variety of context and language encountered during model training and successfully resolving the OOV cold start issues when sparse labels are present.

(2) Adversarial training. In adversarial learning, the goal of adversarial training is to increase the model's dependability when faced with oppositional data. Adversarial training involves creating oppositional samples by introducing small disruptions into the input data. These samples are then used to train the model to better withstand adversarial attacks. This is a more focused and specialized application that is usually used to improve the stability and security of models. Equation (6) illustrates the adversarial learning formula that was defined in [17]:

$$\min_{\theta} \mathbb{E}_{(x,y)\sim D}\left[\max_{r_{\text{adv}}\in S} L(\theta, x + r_{\text{adv}}, y)\right] \tag{6}$$

Where x refers to the original input sample, y represents the corresponding true value label, D represents the training set, r_{adv} is the perturbation, S represents the perturbation space, $L(\theta, x, y)$ represents the loss of a sample, and finding a set of opposing samples in the sample space that maximizes the loss is what the function in parenthesis signifies. When confronted with this data distribution, the outer function refers to altering the model parameters to lower the expected loss of the model on this adversarial sample set. To improve the model's robustness and recognition efficiency, this paper creates an adversarial training that leverages Projected Gradient Descent (PGD) to create adversarial samples, add perturbations to the model's embedding layer, and train the model using these samples. PGD's fundamental concept is to move the current sample a little bit toward the gradient of the loss function with each iteration, then map it back to a predetermined range to make sure the final sample stays inside that range. Until the predetermined number of iterations is reached or a particular termination criterion is met, this process is repeated. The following formulas can be used to represent the process:

$$r_{\text{adv}}^{t+1} = \prod_{r_{\text{adv}}\in S\leq\delta} \left(r'_{\text{adv}} + \alpha \cdot g\left(r'_{\text{adv}}\right) / g\left(r'_{\text{adv}}\right)_2\right) \tag{7}$$

$$g\left(r'_{\mathrm{adv}}\right) = \nabla_{r_{\mathrm{adv}},L}\left(f_\theta\left(x + r'_{\mathrm{adv}}, y\right)\right) \tag{8}$$

Where $r_{\mathrm{adv}}, F \leq \delta$ is the constraint space of the disturbance, α is the step size of the disturbance. If the disturbance is too large, it will be pulled back to the boundary of the sphere.

3.5. **Multi attention mechanism. (1) A multi-head self-attention method for interaction between aspects and sentences.** Here, a multi-head self-attention mechanism is employed to extract keyword information from aspects and interact it with context to determine the impact of keywords on context in aspects. In the multi-head self-attention mechanism, each attention head operates independently and is capable of capturing unique sentence representations from different feature spaces. By modeling aspects and comment texts equally, a multi-head self-attention mechanism is implemented to dynamically collect different semantic and spatial information representations of sentences and aspects, capturing deeper emotional feature information [15].

The hidden states of aspect terms h_t^a and contextual hidden states h_t^c are obtained through left and right BiLSTM, concatenated in the sequence dimension, and represented by

$$h_t^{ac} = h_t^{ac} = [h_t^a; h_t^c]$$

The output of multi-head self-attention can be obtained by feeding h_t^{ac} into a multi-head self-attention network. The aspect and sentence interaction representation can then be acquired by performing an average pooling operation; the precise calculation method is provided below.

$$\mathrm{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{Concat}(\mathrm{head}_1; \mathrm{head}_2; \ldots; \mathrm{head}_n)\mathbf{W} \tag{9}$$

$$\mathrm{head}_i = \mathrm{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) \tag{10}$$

Where $\mathbf{W}$ is the weight matrix and head_i is the scaled dot product attention mechanism's output.

(2) Attention from aspect to sentence. The functions of the various words in the context vary in how they depict the sentence. First, get the hidden context representation of the aspect h_t^a using the BiLSTM on the left. Second, the hidden context representation of the phrase h_t^c is obtained by concatenating the word embedding vector, position embedding vector, and part of the speech embedding vector through the BiLSTM on the right. Lastly, determine each word's attention weight distribution in the aspect terms that match the sentence [16]. The relevant calculation formula is

$$S_i = \sum_{j=1}^{n} \alpha_{ij} h_j^c \tag{11}$$

$$\alpha_{ij} = \frac{\exp\left[f(h_j^c, h_i^a)\right]}{\sum_{j=1}^{n} \exp\left[f(h_k, h_i^a)\right]} \tag{12}$$

$$f(h_j^c, h_i^a) = \tanh(h_j^c W_m h_i^a + b_m) \tag{13}$$

where tanh is a nonlinear activation function, W_m and b_m are trainable parameters, and α_{ij} is the attention weight from the i word in the aspect terminology to the j word in the sentence.

(3) Attention from sentence to aspect. The hidden representation of comment text can be obtained through the BiLSTM on the right, and then the attention weights of different words in the aspect can be calculated using location information, part-of-speech information, and semantic information. Different weights are further allocated to the hidden representation of attention output from aspect to sentence, reducing the number of learnable parameters, accelerating model convergence, and filtering out redundant information and noise in the original text, further improving classification accuracy. This process can be expressed as:

$$c_r = \sum_{j=1}^{m} \beta_i S_i \tag{14}$$

$$\beta_i = \frac{\exp[f(h_{\text{avg}}, h_i^a)]}{\sum_{j=1}^{m} \exp[f(h_{\text{avg}}, h_j^a)]} \tag{15}$$

$$f(h_{\text{avg}}, h_i^a) = \tanh(h_{\text{avg}}^T W_n h_j^a + b_n) \tag{16}$$

$$h_{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} h_i^c \tag{17}$$

Where β_i represents the attention weight of words in the sentence to aspect terminology; h_{avg} is obtained by averaging the hidden states of BiLSTM on the right side through pooling; W_n is the weight matrix; b_n is the bias; c_r is the output that has been attended to.

3.6. **Emotion classification.** The final aspect-focused sentiment feature representation x is created by combining the hidden state h and the attention output c_r learned from the phrase to aspect by the multi-head self-attention mechanism for aspect and context interaction. The projected dimension of x, which is then put into a dense layer, corresponds to the number of label categories. There are three categories in this study. Lastly, a Softmax function is employed to ascertain the probability distribution of the text sentiment polarity classes. The precise formula for calculation is:

$$y = \text{softmax}(W_x x + b_x) \tag{18}$$

3.7. **Model training.** The aim of the model is the cross-entropy loss function, which is expressed as follows:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} \log \hat{y}_{ij}) + \lambda \|\theta\|^2 \tag{19}$$

Where n is the instance number and m is the category count; Sample i's label is y_{ij}, and $\hat{y}_{ij}$ forecasts the likelihood that it will fall into category j. The regularization coefficient is denoted by λ, whereas the collection of all model parameters is represented by θ. Utilizing gradient descent, optimize and modify the model parameters to lower the loss function when training the model by comparing the loss magnitude between the label value and the predicted value.

4. **Experiment.**

4.1. **Hyperparameter setting.** Using a hidden unit count of 420 dimensions, embedding representations of aspects and phrases are produced, both of which are 768 dimensions, building on the pre-trained RoBERTa model. Both the part-of-speech and positional embedding dimensions are set to 100. The model can process phrases up to 90 lengths, and the batch size is 64, including 6 multi-head attention heads. Every weight matrix is produced by evenly allocating (-0.1,0.1). The PyTorch deep learning framework is used to implement the model that is suggested in this research. Additionally, the model has L2 regularization and a Dropout mechanism, and the Adam optimizer is utilized with a defined learning rate. There are three sentiment polarity categories, the regularization coefficient is [value], and the dropout rate is set at 0.3.

4.2. **Dataset.** The restaurant and laptop datasets from the SemEval 2014 challenge are used to evaluate the suggested model's performance, as well as the SemEval 2015 and 2016 tasks' restaurant 15 and restaurant 16 datasets, respectively. Four benchmark aspect-level sentiment categorization datasets are made accessible to the public for experimentation. Emotional categories, attributes, and comment statements make up each piece of data. The dataset's statistics are shown in Table 1. An example of the Restaurant 14 dataset can be found in Table 2.

Table 1. Dataset statistics

Data Sets	**Aggressive**	**Negative**	**Neuter**
Restaurant 14 (training)	2164	807	637
Restaurant 14 (Test)	728	196	196
Laptop 14 (training)	994	870	464
Laptop 14 (Test)	341	128	169
Restaurant 15 (training)	1178	382	50
Restaurant 15 (Test)	439	328	35
Restaurant 16 (training)	1620	709	88
Restaurant 16 (Test)	597	190	38

Table 2. Sample restaurant 14 dataset

Comment statements	**Aspect**	**Emotion**
I have to say they have one of the fastest delivery times in the city.	delivery times	Aggressive
In fact, this was not a Nicoise salad and was barely eatable.	Nicoise salad	Negative

4.3. **Evaluation Indicators.** Accuracy A and macro average ($macro - F_1$) are used as performance metrics in this experiment. The percentage of samples that are correctly identified out of the total is known as accuracy, and the method used to calculate it is

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \times 100\% \tag{20}$$

where the number of positive cases that the model accurately detects is denoted by TP (True Positives). The number of negative occurrences that the model accurately classifies as negative is indicated by TN (True Negatives). The number of negative cases that the model mistakenly classifies as positive is known as FP (False Positives). The number of positive occurrences that the model mistakenly classifies as negative is known as FN (False Negatives). F_1 is a comprehensive evaluation metric that strikes a balance between recall (R) and precision (P). The harmonic mean of P and R is used to determine the final evaluation score.

Firstly, calculate the F_1 values of each category (positive, neutral, negative), and then average the F_1 values of different categories to obtain the final evaluation metric *Macro*-F_1. *Macro*-F_1 considers the differences between imbalanced categories and categories, which can accurately examine the model's effectiveness.

4.4. **Experimental results.** The suggested model was compared to several contemporary classic multidimensional sentiment analysis models to assess its efficacy.

(1) AE-LSTM (LSTM with aspect embedding): AE-LSTM is a variation of LSTM that uses a unidirectional LSTM network to represent sentences. The sole distinction is that AE-LSTM links the average of aspect features to the sentence embedding's end.

(2) ATAE-LSTM (Attention-based LSTM with aspect embedding): ATAE-LSTM shares a similar model structure to AE-LSTM, emphasizing the significance of aspect words in aspect-level sentiment analysis by embedding them into each context vector. Additionally, create a weight distribution using an attention mechanism to give crucial emotional vocabulary greater attention.

(3) IAN (Interactive Attention Network): IAN is a type of neural network that interacts through attention mechanisms. Make the model treat aspect information and contextual features equally.

(4) MemNet (memory network): uses deep memory networks rather than gated recurrent units and long short-term memory networks for modeling. It obtains the final emotional feature representation by paying greater attention to the significance of various words in context through multi-layer attention methods.

(5) AOA (attention over LSTM): Capturing aspect and sentence information through fine-grained attention. By using row Softmax and column Softmax to automatically focus on important vocabulary in the sentence, the precision of sentiment classification is boosted.

(6) PBAN (position-aware bidirectional attention network): It models using bidirectional attention and position information. The model becomes more focused on sentiment words that are closer to the aspect words by using position information, which can indicate that words at different distances have varied effects on the sentiment of aspect words.

(7) AEN (Attention Encoder Network): To overcome the difficulties of parallelization and long-term dependence loss in RNN variations, an attention encoding network was proposed. Additionally, to lessen label unreliability problems, implement label smoothing regularization.

The findings of aspect-level sentiment analysis of user reviews are compared between this model and the traditional models presented in Table 3. Table 3 shows the comparison of aspect-level sentiment analysis results between this model and the classic models discussed.

According to Table 3's findings, the AE-LSTM model performs comparatively poorly. Using the average aspect information embedding as part of the attention calculation can result in some information loss, even when the significance of aspect information for

Table 3. Classification results of different models

Model	Dining room14		Laptop		Dining room15		Dining room16	
	A/%	F1/%	A/%	F1/%	A/%	F1/%	A/%	F1/%
AE-LSTM	74.36	60.12	66.37	60.15	74.96	48.55	83.27	52.33
ATAE-LSTM	75.16	61.05	68.82	62.05	75.92	50.09	84.61	56.29
IAN	77.19	64.63	71.98	66.65	77.36	53.48	82.36	62.31
MemNet	78.25	65.72	70.13	63.17	77.65	59.42	83.45	57.68
AOA-LSTM	77.65	64.13	69.31	60.52	76.14	53.61	82.66	56.32
PBAN	79.94	71.55	74.35	61.89	79.31	56.74	81.96	60.85
AEN	81.97	70.28	76.59	62.65	82.16	62.52	83.11	68.76
The model	82.77	72.48	79.19	67.19	83.12	70.79	85.57	71.85

emotional expression is taken into account. To incorporate aspect information into the hidden vectors, the ATAE-LSTM model joins the average of word and aspect embeddings as the LSTM's input. The attention mechanism is also used to help the model focus on words in the comments that have important emotional meaning. As a result, it has been 0.8%, 2.45%, 0.96%, and 1.34% more accurate than AE-LSTM. Compared to the ATAE-LSTM model, the IAN model performs better. It uses two long short-term memory networks to acquire hidden representations of aspect terms and context, treats aspect and context embedding equally, and then uses an interactive attention method to obtain the final sentence representation. The contribution of contextual words associated with aspect sentiment polarity is more likely to be captured by the MemNet model thanks to its several layers of memory networks. The performance is superior to that of the IAN and AOA-LSTM models. To emphasize the significance of aspect and context word distance for sentiment classification, the model first uses RoBERTa to obtain dynamic word embeddings with a stronger representation ability. Next, it takes into account the interaction between aspect information and context using multiattention. Finally, it incorporates location information. The model may also learn semantic representations of aspects and remark sentences in different subspaces thanks to the multi-head self-attention mechanism. Therefore, compared to the other models presented, the model suggested in this article performs better. The accuracy in the four datasets mentioned above has increased by 8.41%, 12.82%, 8.16%, and 2.3% when compared to the AE-LSTM model.

5. Model Analysis.

5.1. **Ablation experiment.** Use ablation experiments to quantitatively examine the significance and contributions of the various modules in the suggested model. Table 4 and Figure 3 present the results of the ablation experiment. The multi-head self-attention mechanism, position feature module, RoBERTa word embedding module, aspect and sentence intersecting attention module, and other components were gradually eliminated from the benchmark model presented in this study. The multi-head self-attention mechanism, the position feature module, the RoBERTa word embedding module, and the attention module that interacts between aspects and sentences were all eliminated one after the other to assess the contributions of the various parts of the suggested model. A, B, C, and D are the respective designations for the new models that resulted. Table 4 shows that the accuracy has dropped by 2.66%, 3.19%, 3.13%, and 3.71%, respectively, when the multi-head self-attention mechanism is used for aspect and sentence interaction. Aspect-level sentiment analysis requires the use of a multi-head self-attention mechanism to interactively learn various semantic spatial information representations of sentences and aspect items, modeling both aspect and comment phrases at the same time. The RoBERTa embedding module, which has a similar effect to the multi-head self-attention module and performs better in downstream tasks with powerful feature representation, is the module with the second highest contribution. The model's sentiment classification performance can be improved by incorporating location data into the semantic learning procedure. Lastly, the model will perform worse if the attention module that interacts with aspects and sentences is removed.

TABLE 4. Ablation results

Model	Restaurant 14		Laptop 14		Restaurant 15		Restaurant 16	
	A/%	F1/%	A/%	F1/%	A/%	F1/%	A/%	F1/%
A	80.11	68.24	76.00	65.16	79.99	67.63	81.96	69.42
B	81.97	69.45	78.21	65.97	82.46	69.63	83.55	69.21
C	80.23	68.56	76.36	66.32	79.56	68.57	82.29	68.74
D	81.78	69.12	77.25	66.46	81.96	69.32	84.51	70.37
Model of this article	82.77	70.28	79.19	67.19	83.12	70.79	85.57	71.85

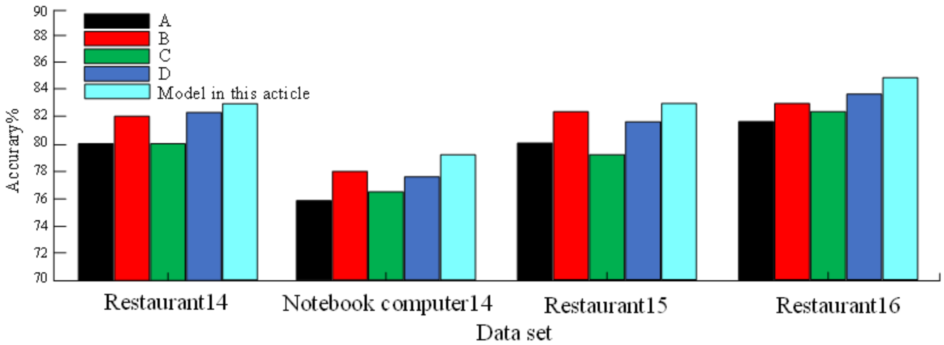

FIGURE 3. Visualization of ablation experiment

5.2. **The impact of the number of attention heads.** To learn semantic information of various representation subspaces in aspects and comment statements, the model in this paper employs a multi-head self-attention technique. To determine the model's performance under various H, the number of heads (H) of the multi-head self-attention mechanism can be changed. The experimental outcomes of the model on four datasets are displayed in Figure 4. The accuracy trend across many datasets is quite comparable as the number of heads in multi-head attention grows. This self-attention technique is typical when the number of heads is 1. The model's performance keeps improving as the number of heads rises. The model performs best when $H = 6$, after which there is a declining tendency. Therefore, merely adding more heads with multi-head attention does not always result in better performance; on the contrary, it causes our model to perform worse. This article has concluded that six heads are the ideal number for the multi-head self-attention mechanism based on the experimental data.

5.3. **Three phrases illustrating the effects of various elements.** Figure 5 illustrates how the suggested model exhibits an increasing trend across all datasets when a phrase has less than three features. On the contrary, increasing the number of aspects from 3 to 7 will lead to a decrease in the fluctuation of model performance. Therefore, the increase

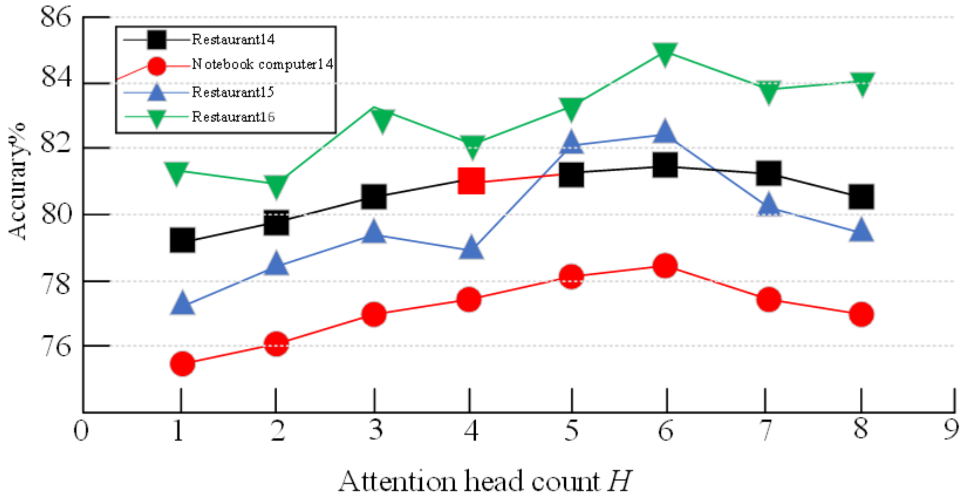

Figure 4. The effect of attention head counts on model performance

in the number of aspects in a sentence hinders the learning of the model. However, the model proposed in this article can still process multiple aspects of sentences with high accuracy, proving the superior performance of the aspect level sentiment analysis model proposed in this article. Conversely, if you increase the number of aspects from three to seven, the model's performance fluctuation will diminish. As a result, understanding the model is hampered by a sentence's increasing number of characteristics. Nonetheless, the model suggested in this article can still accurately analyze several phrase features, demonstrating the aspect level sentiment analysis model's superior performance.

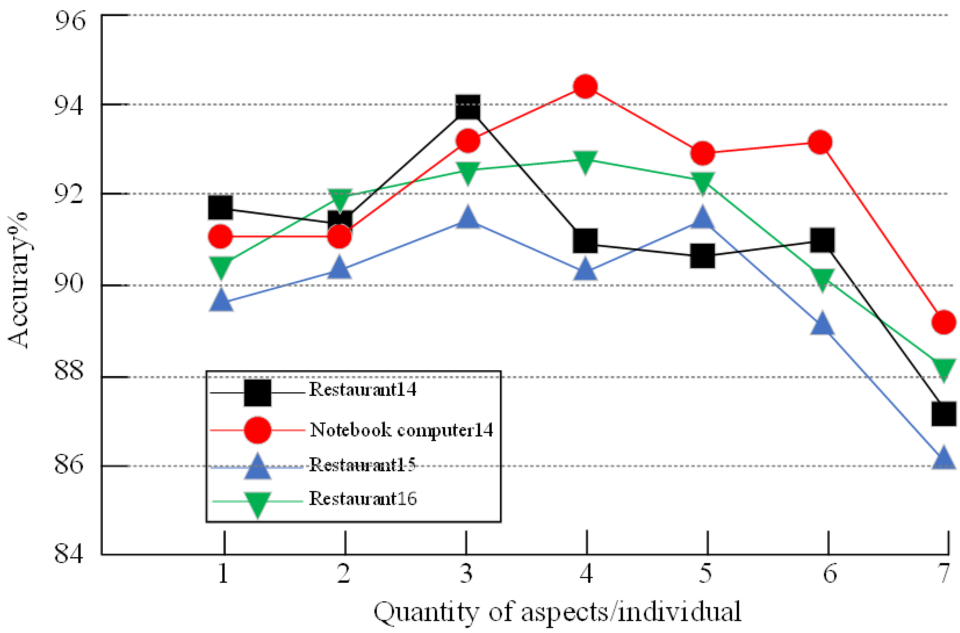

Figure 5. The influence of different aspects of the sentence

6. **Conclusion.** For aspect-level sentiment analysis tasks, a multi-attention model is suggested, which is built on RoBERTa, data augmentation, and location features. The primary concept of this approach is to gain more semantic information by using RoBERTa to build word embedding with higher representation capabilities. Furthermore, the model

has a multi-attention mechanism that can extract more detailed semantic information between aspect words and sentences. Positional information may also be used to enhance the model's performance. The evaluation of aspect emotions can be further enhanced by introducing a multi-head self-attention system that can capture the relationships between aspects and comment texts. The results of experiments on four publicly accessible datasets showed that the model suggested in this study can outperform the benchmark model and develop efficient emotional feature representations.

Acknowledgment. The authors are grateful to the Management and Science University in Malaysia, the Guangzhou College of Commerce in China, and the University of Mosul in Iraq for their support. This work is funded by The Teaching Quality and Teaching Reform Project of Guangdong Province in 2022 (No: 2022JXGG01); Higher Education Research Project of the Year 2024 in the 14th Five-Year Plan of Guangdong Higher Education Association (No: 24GYB75); School-level quality Engineering Construction project of Guangzhou College of Commerce in 2024 (No: 2024ZLGC11); Teaching Quality Management Professional Committee of Private Colleges and Universities of Guangdong Province Teaching Management Association 2024 Project (No: GDZLGL2428); "Artificial Intelligence+ "Pilot Course Project "Data Collection Technology"(No: 2024rgznsdkc06) .

REFERENCES

[1] H.-j. Park, M. Song, and K.-S. Shin, "Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories," Knowledge-Based Systems, vol. 187, p. 104825, 2020.

[2] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey," IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 845-863, 2022.

[3] N. Majumder, R. Bhardwaj, S. Poria, A. Gelbukh, and A. Hussain, "Improving aspect-level sentiment analysis with aspect extraction," Neural Computing and Applications, vol. 34, no. 11, pp. 8333-8343, 2020.

[4] X. Wang, X. Chen, M. Tang, T. Yang, and Z. Wang, "Aspect-Level Sentiment Analysis Based on Position Features Using Multilevel Interactive Bidirectional GRU and Attention Mechanism," Discrete Dynamics in Nature and Society, vol. 2020, pp. 1-13, 2020.

[5] B. Yang, D. Han, R. Zhou, D. Gao, and G. Wu, "Aspect opinion routing network with interactive attention for aspect-based sentiment classification," Information Sciences, vol. 616, pp. 52-65, 2022.

[6] A. Naresh and P. Venkata Krishna, "An efficient approach for sentiment analysis using machine learning algorithm," Evolutionary Intelligence, vol. 14, no. 2, pp. 725-731, 2020.

[7] F. M. Rammo and M. N. Al-Hamdani, "Detecting the speaker language using CNN deep learning algorithm," Iraqi Journal for Computer Science and Mathematics, vol. 3, no. 1, pp. 43-52, 2022.

[8] A. A. Qanbar and Z. Y. Algamal, "Improving Support vector machine for Imbalanced big data classification," Journal of Intelligent Systems & Internet of Things, vol. 11, no. 2, 2024.

[9] Y. Wan, Y. Chen, L. Shi, and L. Liu, "A knowledge-enhanced interactive graph convolutional network for aspect-based sentiment analysis," Journal of Intelligent Information Systems, vol. 61, no. 2, pp. 343-365, 2022.

[10] B. Yu and S. Zhang, "A novel weight-oriented graph convolutional network for aspect-based sentiment analysis," The Journal of Supercomputing, vol. 79, no. 1, pp. 947-972, 2022.

[11] X. Zhang and T. Gao, "Multi-head attention model for aspect level sentiment analysis," Journal of Intelligent & Fuzzy Systems, vol. 38, no. 1, pp. 89-96, 2020.

[12] Z. Zhou and F. a. Liu, "Filter gate network based on multi-head attention for aspect-level sentiment classification," Neurocomputing, vol. 441, pp. 214-225, 2021.

[13] J. Shi, W. Li, Q. Bai, Y. Yang, and J. Jiang, "Syntax-enhanced aspect-based sentiment analysis with multi-layer attention," Neurocomputing, vol. 557, p. 126730, 2023.

[14] G. Pang et al., "Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model," Wireless Communications and Mobile Computing, vol. 2021, no. 1, 2021.

[15] X. Wang, F. Li, Z. Zhang, G. Xu, J. Zhang, and X. Sun, "A unified position-aware convolutional neural network for aspect based sentiment analysis," Neurocomputing, vol. 450, pp. 91-103, 2021.
[16] M. Wankhade, C. S. R. Annavarapu, and A. Abraham, "MAPA BiLSTM-BERT: multi-aspects position aware attention for aspect level sentiment analysis," The Journal of Supercomputing, vol. 79, no. 10, pp. 11452-11477, 2023.
[17] J. Li et al., "A SAR Ship Detection Method Based on Adversarial Training," Sensors, vol. 24, no. 13, 2024.