Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 10, No. 1, 2022, 139-149.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Text Sentiment Classification based on Separate Embedding of Aspect and Context

Amir Lakizadeh * and Elham Moradizadeh

Computer Engineering Department, University of Qom, Qom, Iran.

Article Info

Article History:
Received 25 July 2021
Revised 07 September 2021
Accepted 12 December 2021

DOI:10.22044/JADM.2021.11022.2249

Keywords:
Aspect-level Sentiment Classification, Deep Learning, Attention Mechanism, Bidirectional Long Short-term, Memory.

**Corresponding author: lakizadeh@qom.ac.ir (A. Lakizadeh).*

Abstract

Text sentiment classification in aspect level is one of the hottest research topics in the field of natural language processing. The purpose of the aspect-level sentiment analysis is to determine the polarity of the text according to a particular aspect. Recently, various methods have been developed in order to determine sentiment polarity of the text at the aspect level; however, these studies have not yet been able to model well the complementary effects of the context and aspect in the polarization detection process. Here, we present ACTSC, a method for determining the sentiment polarity of the text based on separate embedding of aspects and context. In the first step, ACTSC deals with the separate modelling of the aspects and context in order to extract new representation vectors. Next, by combining generative representations of aspect and context, it determines the corresponding polarity to each particular aspect using a short-term memory network and a self-attention mechanism. The experimental results in the SemEval2014 dataset in both the restaurant and laptop categories show that ACTSC is able to improve the accuracy of the aspect-based sentiment classification compared to the latest proposed methods.

1. Introduction

Today, a large amount of user-generated content in a variety of social networks, stores, and communication platforms are in the form of text, which can be a valuable resource for the users and companies in order to make decisions. Due to the high volume of the accumulated data, the requirement for automatic and intelligent processing of this information and extraction of knowledge of embedding in them, more than ever, has become apparent[1].Accordingly, the attention of the researchers in the field of natural language processing to the sentiment analysis has increased [2]. Although sentiment analysis has been successful at the other levels such as the document level and the sentence level, these two approaches do not provide a detailed information of the text. The aspect-level sentiment analysis is a more efficient approach since it focuses on the useful and accurate information, and ignores the less important information[3, 4]By definition, aspect means any property or characteristic of a particular entity. For example, for a laptop, some aspects are size and price [5]. More precisely, in the phrase "rice was great but shrimp was bad", the polarity about the aspect "rice" is obviously positive, while the polarity of "shrimp" is negative [6]. The recent researches have realized the importance of aspects, and have developed various methods with the aim of accurately modelling of the texts by extracting aspect-specific representations. Most of these methods ignore the complementary effects of aspect and context on each other. Another limitation of the previous methods is the use of Word2vec word embedding, which requires a very large dataset for training, in which the emotional information is ignored. Another limitation is the lack of attention to the multi-word aspects in the recent methods. The authors in IAN [7] have tried to pay attention to the text and the aspect at the same time but their method has ignored the multi-word aspects. ATAE-LSTM [8] uses the same word-representation for both aspect and context based on the LSTM network. PBAN [9]focuses on the positions of the words in the text but this method also ignores the multi-word aspects. AEN [10] uses Attentional Encoder in order to predict the polarity of aspect. The authors in HAM [11] have introduced a hierarchical method in order to determine the sentiment polarity of aspects based on a simple attention mechanism.

In the proposed model, an attempt has been made to improve the accuracy of the aspect-based sentiment classification by reducing some of the limitations in the previous methods. In this research work, ACTSC, a method for determining the sentiment polarity of the text based on separate embedding of aspects and context is presented. In the first step, ACTSC deals with separate modellings of aspects and context in order to extract new representation vectors. Next, by combining the generative representations of aspect and context, it determines the corresponding polarity to each particular aspect using a short-term memory network and a self-attention mechanism. The experimental results on the SemEval2014 dataset show that ACTSC could improve the accuracy of the aspect-based sentiment classification compared to the latest proposed methods.

2. Related Works

The topic of SA has been studied in the literature [3, 4]. In this section, we briefly review two distinct approaches for SA: traditional machine learning and deep learning approaches. SA has been studied in various literatures [3, 4]. In this section, we will take a different perspective on SA, and examine some of the common techniques in this area.

2.1. Traditional Machine Learning Approach for SA

The Machine Learning (ML) methods are very common for SA, including most of the SA research. These methods include the two basic parts of supervised approaches and un-supervised approaches.

2.1.1 Neural Networks Methods

Supervised learning requires the labelled data. That data collection is very difficult. These algorithms require a supervisor in order to provide the input data and expected outputs [3]. Various supervised techniques have been used for SA. The techniques such as Support Vector Machine (SVM), Naïve Bayes (NB), and Maximum Entropy (ME) have achieved a great success in this field of research[12]. For example, the author in [13] has evaluated the performance of the three classifiers of SVM, NB, and ME. The success of these classifier is heavily dependent on the extracted features. For this purpose, they used unigrams, bi-grams, and a combination of unigrams and bigrams with POS. Their results showed that the NB classifier on the low-features was better than the other two classifiers but when the dimension of the feature increases, the SVM classifier gets better results than NB. The authors in [14] have used SVM, NB, and character-based n-gram model to classify sentiment in the online travel reviews.

Various combinations of classifiers have been used on the labelled data. For instance, Zhang et al. [15] have used the Support Vector Machine (SVM) and Naive Bayes (NB) approaches in order to classify the movie reviews. They used unigrams, bigrams, and trigrams as the features to train their classifiers. In addition, in [16], several approaches mainly SVM and rule-based classifiers with POS(Part-Of-Speech) tag and n-gram features have been evaluated. They have been used to classify the movie reviews and product reviews. SVM has also been used in [17, 18] for classification of the movie reviews. In order to train the classifier, various features including unigram, bigram, and word frequency have been used in these works.

2.1.2 Unsupervised Learning Approach for SA

The supervised approaches require labelled data that causes the data collection to be costly and difficult. The unsupervised learning approaches use unlabeled data in order to extract a similar pattern and a useful information from the input data. In fact, these approaches are used when collecting the annotated data is hard [19]. Turny [20]has used an unsupervised approach to classify the reviews into the two categories of recommended or not recommended. His proposed algorithm includes two basic steps. In the first step, it extracts phrases containing adjective or adverb and in the second step, Semantic Orientation (SO) calculated for the extracted. The PMI-IR algorithm is used for this purpose. The PMI-IR algorithm uses mutual information in order to calculate the semantic similarity between two phrases [21].

In [22] the authors have proposed an unsupervised approach for sentiment classification on tweets. Taras and john[23] have suggested an unsupervised approach for sentiment classification in Chinese text using a set of seed words that are automatically selected.

2.2 Deep Learning for SA

In the recent years, with the increasing power of Graphical Processing Units (GPUs) and the availability of massive amount of data, the Deep Learning (DL) techniques have been very successful in solving Machine Learning (ML) tasks. In many NLP tasks, DL is much more efficient than the traditional ML and statistical methods used a few years ago[24]. DL has been very successful in text generation, vector representation, sentence classification, sentiment analysis, sentence modelling, and feature representation [25]. One of the main advantages of DL is that there is no need for manually tune the features based on the expert knowledge and available linguistic resources [12]. However, one of the limitations of the neural network when working with the text data is that raw text cannot be given to the network because the neural network receives the data as a vector, and produces it as an output vector. Instead of using unique dimensions for each feature, they tried to embed each feature in a D-dimensional space, and represent it as a dense vector. The most important advantage of these vectors is that similar words are placed close to each other in the vector space [26]. There are several ways to create these vectors, among which Word2vec [27]and Glove [28] are more common.

DL includes various types of Artificial Neural Networks (ANNs) such as Convolution Neural Networks (CNNs), Recurrent Neural Networks (RNN), Long-Short Term Memory (LSTM), Gated Recurrent Units (GRUs), and Capsule-Net, which we will continue to focus on each one of them individually.

2.2.1 Convolutional Neural Networks (CNNs)

CNNs are a kind of Deep Neural Network (DNN) initially introduced by LeCun for object recognition [29]. Today, these networks are used for various tasks such as face recognition [30], human pose estimation [31], speech recognition [32], and NLP[33],CNNs are a kind of feedforward neural networks with the features such as convolution layers, sparse connection, parameter sharing, and pooling [12]These networks consist of three types of layer: i) Convolution layer that attempts to extract the features using kernels on the input feature map or intermediate feature map. One of the main advantages of this type of layers are sharing of parameters, which reduces the network parameter severely. ii) Pooling layer: these layers are used in order to reduce the network parameters and prevent overfitting. Max-pooling and average-pooling are the most commonly used pooling strategies. Among the layers of CNNs, pooling has been most studied, and different strategies have been proposed for it such as stochastic pooling [34]and spatial pyramid pooling [35, 36]. In max-pooling, the highest amount of input feature map that the kernel is applied to, and in average-pooling the average feature are selected [37]. iii) Fully connected layers: these layers are similar to the traditional Neural Network (NN) and form 90% of the CNNs parameters. The task of these layers is to convert a 2D feature map into a 1Dl feature vector in order to calculate score of the categories or to continue the learning process. The disadvantage of these layers are the presence of a lot of parameters in them, which is a great deal to learn [37]. CNNs in SA have also achieved a lot of success. Kim [38]has also used two models of CNN multichannel and CNN non-static for the binary classification of movie reviews.

2.2.2 Recurrent Neural Networks (RNNs)

RNNs are feed-forward networks that consider the concept of time in the modelling. This concept is defined by the edges in the adjacent steps. The edges that connect the adjacent times are called the ‘recurrent edges’. These edges may create cycles. In these networks, the status at any moment depends on the current input and the previous step [38].

In these networks, during the back-propagation process, for long time steps occurs the vanishing and exploding gradient. The Long-Short Term Memory (LSTM) network has been offered as a solution to these problems by Hochreiter and Schmidhuber [39]. These networks are similar to the RNN networks, except that each node in the hidden layer is replaced by a memory cell. Each memory cell contains a node with a self-connected recurrent edge, and assures that the gradient can pass through many time steps without gradient and vanishing gradient[39]. The LSTM networks include four gates: i) input gate, ii) output gate, iii) update gate, and iv) forget gate [40]. Another kind of recurrent networks proposed to deal with the vanishing and exploding gradient problems is the Gated Recurrent Unit (GRU) networks. These networks were first designed by Kyunghyun Cho for natural machine translation [41]. Given the structure of these networks and the nature of natural language, they are a good choice for SA, For example, Li et al. [42]have used the RNN and LSTM networks for binary classification movie reviews, respectively, reaching 74% and 78% accuracy at the fine-grained classification.

M. Zeiler et al. [34]have also proposed the Tree-LSTM, which is based on the LSTM networks, for two different tasks: i) predicting semantic relatedness of two sentences (Semeval 2014 task1) ii) classification (Stanford Sentiment Treebank). In bidirectional RNNs, two independent RNNs are placed together. The input sequence is fed in the normal time order for one network, and in the reverse time order for another. The outputs of the two networks are usually concatenated at each time step. This structure allows the networks to have both backward and forward information about the sequence at every time step. Bidirectional models such as Bi-LSTM will run inputs in two ways, one from past to future and one from future to past, and what differs this approach from unidirectional models (such as LSTM) is that in LSTM that runs backward it preserves information from the future but using the two hidden states combined in Bi-LSTM, it is able in any point in time to preserve information from both the past and future. References [43-47]show a number of recent methods based on the Bi-LSTM networks for the sentiment polarity classification task.

2.2.3 Aspect-Based Sentiment Analysis (ABSA)

ABSA takes into consideration the terms related to the aspects, and determines the sentiment polarity associated with each aspect. The ABSA model requires aspect categories and its corresponding aspect terms in order to extract the sentiment polarity for each aspect from the text corpus. The ABSA problem can be categorized as two important sub-tasks, as follows: aspect extraction (AE) and aspect sentiment classification (ASC). In AE, the aspects are extracted, and the range of those aspects is analyzed. The ASC sub-task recognizes the polarity of that aspect. Recently, various methods have been developed in order to determine the sentiment polarity of the text at the aspect level. B. Huang [48] has proposed AOA-LSTM that creates interaction between the aspect and the text. One of the limitations of this method is the lack of attention to the new representations and separate modeling. D. Tang has proposed the TD-LSTM network [49]. It has performed better than the above models. This network uses a combination of the aspect vector and text using averaging, which does not pay attention to the aspect embedding aspect information lost by passing through the network layers. K. Wang et al. have proposed the R-GAT network [50]. It uses the syntactic information and a word dependency tree in order to analyze the communication between the aspects and the context. The performance of this model compared to the previous models has been improved. The embedding of the aspect has not been considered for a long time. HAM [11]works in a hierarchically manner. Firstly, it extracts an embedding vector for the aspects. Next, it employs these aspect vectors with context information in order to determine the sentiment polarity of the text. Some of the other ABSC methods are [51-55].

3. Methods

3.1 Research Objectives

Considering the limitation of the current methods in modelling the complementary effects of the context and aspect in their sentiment polarization, we propose the ACTSC method, which firstly models the context and aspects using the LSTM networks, and then the generated representation vectors for aspects and context are combined through a Bi-LSTM network with a self-attention mechanism in order to extract the polarity of a particular aspect. Figure 1 shows the architecture of the proposed method.

3.2 Word Embedding type

3.2.1 GloVe Embedding

Glove is a count-based, unsupervised learning model that uses co-occurrence (how frequently two words appear together) statistics at a global level to model the vector representations of the words. Since the statistics are captured at a global level directly by the model, it is named as 'Global vectors' model. By sampling from uniform distribution, all the words that are out of vocabulary are initialized U (-0.01, 0.01). Suppose $\mathcal{L} \in \boldsymbol{R}^{\boldsymbol{d_{emb*|v|}}}$ to be the pre-trained GloVe embedding matrix, where d_{emb} is the dimension of the word vectors (here 300) and $|\boldsymbol{v}|$ is the vocabulary size; it maps each word $w_i \in \boldsymbol{r}^{|\boldsymbol{v}|}$ to its associated embedding vector $\boldsymbol{e_i} \in \boldsymbol{r}^{\boldsymbol{d_{emb*1}}}$, which is a column in the embedding matrix L[28]

3.2.2 Bert Embedding

BERT stands for Bidirectional Encoder Representations from Transformers. It is designed to pre-train deep bidirectional representations from the unlabeled text by jointly conditioning on both the left and right contexts. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer in order to create the state-of-the-art models for a wide range of NLP tasks. In order to make training and fine-tuning of BERT model easy, the given context and aspect are transformed to "[CLS] + context + [SEP]" and "[CLS] + aspect + [SEP]", respectively [56].

3.3 Context and Aspect Modelling

If the context and aspect have M and N words, their weights will be equal to $[w_w^1, w_w^2, \ldots, w_w^m]$ for the context and $[w_a^1, w_a^2, \ldots, w_a^n]$ for the aspect, respectively. Context and aspect are given to the Bi-LSTM network in order to extract the hidden feature vectors for both of them. By averaging the hidden vectors $[h_w^1, h_w^2, \dots, h_w^m]$ and $[h_a^1, h_a^2, \dots, h_a^n]$, the new representation vectors vw and va are obtained for the context and aspect.

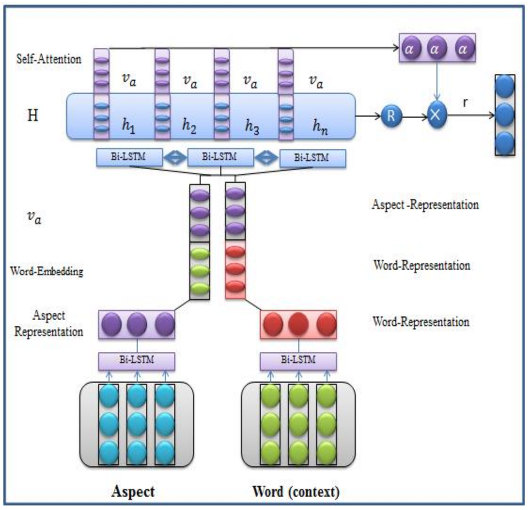

Figure 1. Architecture of the proposed method.

The averaging formula is as follows:

$$v_{\alpha} = \sum_{i=0}^{n} \frac{h^i}{n} \tag{1}$$

$$v_w = \sum_{i=0}^{m} \frac{h^i}{m} \tag{2}$$

3.3.1 Combining of Representations

After generating the suitable representation vectors for the aspect and context in terms of the Glove and BERT vectors in the previous steps, they are combined using the Bi-LSTM network based on a self-attention mechanism. Using the self-attention mechanism produces more efficient hidden states and more focus on the main parts of the text, which is effective in determining the polarity corresponding to each aspect. Combination task is done as follows:

$$i^k = \sigma(w_i^k . w^k + w_i^h . h^{(k-1)} + b_i) \tag{3}$$

$$f^k = \sigma(w_f^w . w^k + w_f^h . h^{(k-1)} + b_f) \tag{4}$$

$$ok = \sigma(w_o^w . w^k + w_o^h . h^{(k-1)} + b_o) \tag{5}$$

$$\hat{c}^k = (w_c^w . w^k + w_c^h . h^{(k-1)} + b_c) \tag{6}$$

$$c^k = f^k \odot c^{(k-1)} + i^k \odot \hat{c}^k \tag{7}$$

$$h^k = o^k \odot tanh(c^k) \tag{8}$$

where Bi-LSTM is as:

$$h^k = (\overrightarrow{h^k}, \overleftarrow{h^k}) \tag{9}$$

The input is w^k and c^{k-1} is the previous state, and also h^{k-1} is the previous hidden state; b, σ, f, and o are the input dimension, sigmoid function, and forgot and output gates.

3.3.2 Self-attention Mechanism

The self-attention mechanism gives a more accurate weight to the hidden states in the model compared to the simple mechanism. This causes more attention to the aspects and as a result is more effective in the sentiment polarity classification [57].

4. Experiments and Results

4.1 Data Collection

The data used in this work are SemEval2014[5], SemEval2015[58], SemEval2016[59],which aims to develop the research work in the field of aspect-based sentiment analysis. The goal of this task is to identify the aspects of a given entity, and determines the emotional polarity for each aspect. This collection includes the user comments in two categories: “restaurant” and “laptop”. The goal is to determine the polarity of these comments. Table 1 shows the details of the datasets used.

Table 1. Details of the used datasets in terms of three classes positive, negative, and neutral.

Property	Datasets			
	Restaurant14		**Laptop14**	
	Train	**Test**	**Train**	**Test**
#samples	1978	600	1462	411
Pos./Neg. /Neu.	2164/805 /633	728/196/ 196	987/866/ 460	341/128/ 169
Property	Datasets			
	Restaurant15		**Restaurant16**	
	Train	**Test**	**Train**	**Test**
#samples	1120	582	587	587
Pos./Neg. /Neu.	1198/403 /53	454/346/ 45	1657/749 /101	611/204/ 44

4.2. Evaluation Metrics

The quality of a classifier can be represented by a confusion matrix. Each one of its records represents the correct and incorrect example for each class. Various measure can be used in order to evaluate the efficiency of a classifier. In this work, we used the accuracy, precision, recall, and F1_score measures in order to evaluate the proposed model.

$$\text{Pr}ecision = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Re}call = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

Macro-averaging is used because there are three classes: positive, negative, and neutral. In other words, instead of averaging the classes one by one, all classes are averaged as follows:

$$Macro\ Recall = \frac{1}{|c|} \sum \frac{TP}{TP + FN} \quad (14)$$

$$Macro\ Precision = \frac{1}{|c|} \sum \frac{TP}{TP + FP} \quad (15)$$

$$MacroF1 = \frac{2 * Macro\ Recall * Macro\ Precision}{Macro\ Recall + Macro\ Precision} \quad (16)$$

4.3. Results

ACTSC determines the sentiment polarity of the text based on the separate embedding of aspects and context. In the first step, ACTSC deals with the separate modeling of aspects and context in order to extract the new representation vectors. Next, by combining the generated representations for aspect and context, it determines the corresponding polarity for each particular aspect using a short-term memory network and a self-attention mechanism. In this section, we review the results of comparison of the proposed method with some of the state-of-the-art methods. Tables 2-4 show the comparison results.

Since the proposed method uses GloVe and BERT word-embedding for the initial representation of the words in input, ACTSC-GloVe and ACTSC-BERT are used to name the proposed method in these two cases. According to Tables 2, and 3, the Memnet model [60] has the lowest accuracy in the restaurant and laptop data. This model uses a HOP, which makes the network performance to be highly dependent on HOP. One of the limitations of the Memnet model is the inflexible network performance and high dependency to the computational layer. AF-LSTM [61] performs better than the previous methods, and can also model the word similarity. The improvement rate of the proposed model compared to this model in term of accuracy measure is about 12% and 7% for both the restaurant and laptop datasets, respectively. In IAN [7], both the aspect and context are considered but in this method, also not enough attention is paid to the aspect characteristics. The proposed method can also outperform it. In the ATAE-LSTM [8], the improved aspect and context embedding lead to an improved text polarity relative to a particular aspect. However, this method ignores the separate modeling of aspect and context. Our method can also outperform the TC-LSTM/TD-LSTM[49], PBAN [9], and AOA-LSTM [48]methods.

Table 2. Comparison results of ACTSC with other methods based on laptop data (semEval-2014).

Laptop (semEval-2014)				
Method	**Acc**	**F_1**	**Precision**	**Recall**
MEMNET [60]	64.42	58.22	59.09	59.36
LSTM [39]	66.14	61.26	62.37	60.20
CNN-PF[62]	66.93	57.75	-	-
ATAE-Bi-LSTM [63]	71.94	66.41	66.84	65.99
LCRS [64]	66.46	61.95	63.15	60.81
AF-LSTM [61]	69.96	62.58	65.02	60.32
IAN [7]	68.50	63.39	64.11	62.69
ATAE-LSTM [8]	67.40	63.60	65.16	62.12
GCAE [65]	65.83	60.64	60.95	60.34
TC-LSTM [49]	67.04	62.33	62.02	62.66
CABASC [66]	70.06	64.55	66.14	63.05
AT-LSTM [8]	69.44	64.62	64.23	65.02
SHAN [67]	74.64	-	-	-
Bi-LSTM [68]	68.81	63.49	63.41	63.56
TD-LSTM [49]	68.50	62.81	62.66	62.98
HAM-GLOVE[11]	71.16	65.77	65.07	66.50
PBAN [9]	74.12	-	-	-
BERT-SOFT [69]	74.92	-	-	-
BERT-ATTENTION [70]	74.15	-	-	-
AOA-LSTM [48]	74.50	-	-	-
MGAN [71]	76.21	71.42	-	-
IGCN[72]	72.24	-	-	-
HAM-BERT [11]	76.96	72.70	72.68	72.73
R-GAT [50]	77.42	73.76	-	-
ATGCN-BERT [73]	77.96	73.80	-	-
ASMCDC [74]	74.45	70.46	-	-
ACTSC-GLOVE	**77.43**	**74.48**	**74.04**	**74.94**
ACTSC-BERT	**81.01**	**74.71**	**75.03**	**74.41**

The authors in HAM have introduced a hierarchical method in order to determine the sentiment polarity of aspects based on a simple attention mechanism [11]. In general, in the Laptop data, ACTSC-BERT has the best results in terms of accuracy compared to the other methods. Also in the Restaurant data, ACTSC-GloVe has the best results in terms of accuracy compared to the other methods. Also for a more analysis, we compared the proposed method with some recent methods based on the Restaurant data in SemEval-2015 and SemEval-2016. The results obtained (Table 4) confirm the advantage of the proposed method compared to the other methods.

Table 3. Comparison results of ACTSC with other methods based on Restaurant data (semEval-2014).

Restaurant (semEval-2014)				
Method	**Acc**	**F_1**	**Precision**	**Recall**
MEMNET [60]	73.39	61.92	62.74	61.13
LSTM [39]	74.30	65.90	67.54	64.34
CNN-PF [62]	75.15	60.25	-	-
ATAE-Bi-LSTM [63]	75.98	67.35	67.01	67.71
LCRS [64]	76.25	64.58	68.81	60.85
AF-LSTM [61]	76.46	65.54	-	-
IAN [7]	76.70	65.90	68.29	63.69
ATAE-LSTM [8]	76.79	65.28	67.93	62.84
GCAE [65]	77.28	65.06	69.01	67.18
TC-LSTM [49]	77.41	68.08	69.06	65.18
CABASC [66]	77.68	68.08	69.01	67.18
AT-LSTM [8]	78.04	69.66	70.84	68.52
SHAN [67]	81.02	-	-	-
Bi-LSTM [68]	78.30	67.52	69.11	66.01
TD-LSTM [49]	78.66	69.16	70.84	67.56
HAM-GLOVE [11]	79.11	66.84	72.35	65.67
PBAN [9]	81.16	-	-	-
BERT-SOFT [69]	-	-	-	-
BERT-ATTENTION [70]	-	-	-	-
AOA-LSTM [48]	81.20	-	-	-
MGAN [71]	81.49	71.48	-	-
IGCN[72]	81.34	-	-	-
HAM-BERT [11]	81.52	72.66	74.12	71.26
R-GAT[50]	83.30	76.08	-	-
ATGCN-BERT [73]	82.79	75.10	-	-
ASMCDC [74]	82.53	74.40	-	-
ACTSC-GLOVE	**84**	**77.46**	**77.96**	**76.97**
ACTSC-BERT	**83.13**	**75.11**	**79**	**71.59**

Table 4. Comparison results of ACTSC with other methods based on Restaurant data (SemEval-2015 and SemEval-2016).

Restaurant (semEval-2015)				
Method	**Acc**	**F_1**	**Precision**	**Recall**
Bi-LSTM[68]	78.34	53.84	52.34	54.36
GCAE[65]	76.33	55.52	57.61	53.59
LCRS[64]	75.50	56.22	59.09	53.63
AT-Bi-LSTM [63]	79.88	51.97	53.11	50.88
WORD&Clouse-Level[75]	80.90	65.50	-	-
LSTM+ASPECT+SPR[76]	81.7	66.06	-	-
ATAE-CAN[77]	81.75	-	-	-
ACTSC-GLOVE	**83.97**	**76.61**	**74.05**	**75.19**
ACTSC-BERT	**83.11**	**73.65**	**73.01**	**74.31**
Restaurant (semEval-2016)				
Method	**Acc**	**F_1**	**Precision**	**Recall**
Bi-LSTM[68]	63.81	63.48	63.41	63.56
GCAE[65]	79.98	49.97	49.95	50.00
LCRS[64]	81.62	62.41	68.60	57.26
AT-Bi-LSTM [63]	82.89	56.88	-	-
COMEM BERT-FC[78]	83.5	70.88	-	-
M-BERT[79]	81.16	83.55	-	-
ATAE-LSTM[8]	77.53	66.41	66.84	65.99
RAM[80, 81]	83.9	-	-	-
CNN[82]	71.48	60.19	61.89	58.59
ACTSC-GLOVE	**83.50**	**73.35**	**73.50**	**73.21**
ACTSC-BERT	**83.10**	**72.25**	**72.38**	**72.14**

5. Discussion

In order to more accurately evaluate the proposed method and clarify its advantages over the other methods, in this section, we access the performance of the proposed method from different aspects.

5.1 Performance on Multi-word Aspects

One of the main advantages of the proposed method is the ability to detect a correct sentiment polarity for multi-term aspects. Table 5 shows the frequency of the one-term and multi-term aspects in the Restaurant and Laptop datasets.

Table 5. Frequency of one-term and multi-term aspects in the Restaurant and Laptop datasets.

Properties	**Datasets(Train)**	
	Restaurant	**Laptop**
Single-word (len = 1)	2720/75.38%	1473/63.27%
Multi-word (len = 2)	604/6.74%	649/27.88%
Multi-word (len > 2)	284/7.87%	206/8.85%
Properties	**Datasets(Test)**	
	Restaurant	**Laptop**
Single-word (len = 1)	801/71.52%	351/52.78%
Multi-word (len = 2)	215/19.20%	209/31.43%
Multi-word (len > 2)	104/9.29%	78/11.73%

According to Table 5, the single-term aspects in the restaurant data are greater than the laptop data, and conversely, the multi-word aspects in the laptop data are greater than the restaurant data. The results obtained are illustrated in Figure 2. In these results shown for the restaurant data, the accuracy of the method in case of using GloVe embedding was better than BERT, and in the laptop data, the accuracy of the method using BERT embedding mode was better than GloVe. This is due to when we have more one-word aspects (such as restaurant data) in the sentiment classification process; there is less need to pay attention to the context, and therefore, in this case, using GloVe that does not pay attention to the context has been able to achieve a better accuracy. On the other hand, in the presence of multi-word aspects (such as laptop data), using the context-sensitive embedding such as BERT gives a greater accuracy.

5.2 Efficiency in Terms of Class Separation

The efficiency of the proposed method in terms of the evaluation criteria Precision, Recall, and F1-score for each one of the three classes of positive, negative, and neutral with the GloVe and BERT vectors are shown in Tables 6 and 7, respectively. According to these results, the proposed method has a better performance in detecting a correct sentiment polarity for the positive class.

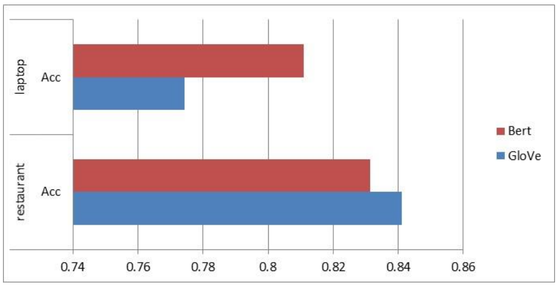

Figure 2. Comparison of results in case of using GloVe and BERT embedding in the Restaurant and Laptop data.

Table 6. Results of class separation of the proposed model with GloVe vectors in the restaurant and laptop data.

ACTSC-GLOVE		Restaurant	Laptop
ACC		**0.84**	**0.77**
Macro-F1	Negative	0.80	0.71
	Neutral	0.61	0.61
	Positive	0.91	0.87
Precision	Negative	0.77	0.60
	Neutral	0.68	0.74
	Positive	0.90	0.88
Recall	Negative	0.84	0.86
	Neutral	0.56	0.52
	Positive	0.92	0.87

Table 7. Results of class separation of the proposed model with BERT vectors in the restaurant and laptop data.

ACTSC-BERT		Restaurant	Laptop
ACC		**0.83**	**0.81**
Macro-F1	Negative	0.77	0.72
	Neutral	0.55	0.59
	Positive	0.90	0.88
Precision	Negative	0.79	0.62
	Neutral	0.72	0.77
	Positive	0.86	0.86
Recall	Negative	0.74	0.84
	Neutral	0.45	0.49
	Positive	0.96	0.90

6. Conclusions

The recent methods in determining the sentiment polarity of the text have confirmed the need for a more attention to the aspects. So far, various methods have been developed in order to determine the sentiment polarity of text in aspect level; however, these studies have not yet been able to model well complementary effects of the context. For this reason, in the proposed method, a new mechanism for text sentiment polarity classification in aspect level was proposed, called ACTSC. ACTSC is based on the separate embedding of the aspects and context using a Bidirectional Long Short-Term Memory (Bi-LSTM) network and self-attention mechanism. It works in two steps. In the first step, it deals with the separate modeling of the aspects and text in order to extract the new representation vectors. Next, combining the generated representations for aspect and context determines the corresponding polarity for each particular aspect using a short-term memory network and a self-attention mechanism. This research work showed that separate modeling of aspect and context could model complementary effects of context and aspects more effectively, and leads to an improved accuracy of text sentiment classification. Also using the Bidirectional Long Short-Term Memory network based on the attention mechanism leads to a more accurate extraction of the aspects with higher importance. The experimental results in the SemEval2014 dataset in both the restaurant and laptop categories show that ACTSC is able to improve the accuracy of aspect-based sentiment classification compared to the latest proposed methods.

Acknowledgments

The authors are grateful to the anonymous referees for scientific judgment of the manuscript

References

[1] A. Saxena, H. Reddy, and P. Saxena, "*Introduction to Sentiment Analysis Covering Basics, Tools, Evaluation Metrics, Challenges, and Applications*," in Principles of Social Networking: Springer, 2022, pp. 249-277.

[2] B. Torabian, "*Sentiment classification with case-base approach*," 2016.

[3] B. Liu, "*Sentiment analysis and opinion mining*," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.

[4] M. Dragoni and G. Petrucci, "*A fuzzy-based strategy for multi-domain sentiment analysis*," International Journal of Approximate Reasoning, vol. 93, pp. 59-73, 2018.

[5] M. Pontiki et al., "*Semeval-2016 task 5: Aspect based sentiment analysis*," in International workshop on semantic evaluation, 2016, pp. 19-30.

[6] B. Xing et al., "*Earlier attention? aspect-aware LSTM for aspect-based sentiment analysis*," arXiv preprint arXiv:1905.07719, 2019.

[7] D. Ma, S. Li, X. Zhang, and H. Wang, "*Interactive attention networks for aspect-level sentiment classification*," arXiv preprint arXiv:1709.00893, 2017.

[8] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "*Attention-based LSTM for aspect-level sentiment classification*," in Proceedings of the 2016 conference on empirical methods in natural language processing, 2016, pp. 606-615.

[9] S. Gu, L. Zhang, Y. Hou, and Y. Song, "*A position-aware bidirectional attention network for aspect-level*

sentiment analysis," in Proceedings of the 27th international conference on computational linguistics, 2018, pp. 774-784.

[10] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "*Attentional encoder network for targeted sentiment classification,*" arXiv preprint arXiv:1902.09314, 2019.

[11] A. Lakizadeh and Z. Zinaty, "*A Novel Hierarchical Attention-based Method for Aspect-level Sentiment Classification,*" Journal of AI and data mining, vol. 9, no. 1, pp. 87-97, 2021.

[12] S. Vohra and J. Teraiya, "*A comparative study of sentiment analysis techniques,*" Journal Jikrce, vol. 2, no. 2, pp. 313-317, 2013.

[13] B. Pang, L. Lee, and S. Vaithyanathan, "*Thumbs up? Sentiment classification using machine learning techniques,*" arXiv preprint cs/0205070, 2002.

[14] Q. Ye, Z. Zhang, and R. Law, "*Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,*" Expert systems with applications, vol. 36, no. 3, pp. 6527-6535, 2009.

[15] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "*Sentiment classification of Internet restaurant reviews written in Cantonese,*" Expert Systems with Applications, vol. 38, no. 6, pp. 7674-7682, 2011.

[16] R. Prabowo and M. Thelwall, "*Sentiment analysis: A combined approach,*" Journal of Informetrics, vol. 3, no. 2, pp. 143-157, 2009.

[17] E. Riloff, S. Patwardhan, and J. Wiebe, "*Feature subsumption for opinion analysis,*" in Proceedings of the 2006 conference on empirical methods in natural language processing, 2006, pp. 440-448.

[18] C. Whitelaw, N. Garg, and S. Argamon, "*Using appraisal groups for sentiment analysis,*" in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, pp. 625-631.

[19] W. Medhat, A. Hassan, and H. Korashy, "*Sentiment analysis algorithms and applications: A survey,*" Ain Shams engineering journal, vol. 5, no. 4, pp. 1093-1113, 2014.

[20] P. D. Turney, "*Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews,*" arXiv preprint cs/0212032, 2002.

[21] K. Church and P. Hanks, "*Word association norms, mutual information and lexicography. I: ACL 27th Annual Meeting 76–83,*" Vancouver. Halvautomatisk ekserpering av anglisismer i norsk, vol. 85, 1989.

[22] M. Fernández-Gavilanes, T. Alvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. J. González-Castano, "*Gti: An unsupervised approach for sentiment analysis in twitter,*" in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 533-538.

[23] T. Zagibalov and J. A. Carroll, "*Automatic seed word selection for unsupervised sentiment classification of Chinese text,*" in Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), 2008, pp. 1073-1080.

[24] D. W. Otter, J. R. Medina, and J. K. Kalita, "*A survey of the usages of deep learning for natural language processing,*" IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 604-624, 2020.

[25] Y. Zhang, J. E. Meng, R. Venkatesan, N. Wang, and M. Pratama, "*Sentiment classification using comprehensive attention recurrent models,*" in 2016 International joint conference on neural networks (IJCNN), 2016, pp. 1562-1569: IEEE.

[26] K. Vyas and D. Naik, "*Language Model Fine-Tuning with Second-Order Optimizer,*" in Advances in Speech and Music Technology: Springer, 2021, pp. 355-366.

[27] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "*Distributed representations of words and phrases and their compositionality,*" in Advances in neural information processing systems, 2013, pp. 3111-3119.

[28] J. Pennington, R. Socher, and C. D. Manning, "*Glove: Global vectors for word representation,*" in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532-1543.

[29] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "*Object recognition with gradient-based learning,*" in Shape, contour and grouping in computer vision: Springer, 1999, pp. 319-345.

[30] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, "*Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis,*" in CVPR 2011, 2011, pp. 3361-3368: IEEE.

[31] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "*Joint training of a convolutional network and a graphical model for human pose estimation,*" Advances in neural information processing systems, vol. 27, pp. 1799-1807, 2014.

[32] L. Deng, "*Farewell editorial: Keeping up the momentum of innovations,*" vol. 22, ed: IEEE Press Piscataway, NJ, USA, 2014, pp. 1687-1687.

[33] Y. Kim, "*Convolutional neural networks for sentence classification. arXiv 2014,*" arXiv preprint arXiv:1408.5882, 2019.

[34] M. D. Zeiler and R. Fergus, "*Stochastic pooling for regularization of deep convolutional neural networks,*" arXiv preprint arXiv:1301.3557, 2013.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "*Spatial pyramid pooling in deep convolutional networks for visual recognition,*" IEEE transactions on pattern

analysis and machine intelligence, vol. 37, no. 9, pp. 1904-1916, 2015.

[36] W. Ouyang et al., "*Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection,*" arXiv preprint arXiv:1409.3505, 2014.

[37] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "*Deep learning for visual understanding: A review,*" Neurocomputing, vol. 187, pp. 27-48, 2016.

[38] Z. C. Lipton, J. Berkowitz, and C. Elkan, "*A critical review of recurrent neural networks for sequence learning,*" arXiv preprint arXiv:1506.00019, 2015.

[39] S. Hochreiter and J. Schmidhuber, "*Long short-term memory,*" Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[40] M. Dragoni and G. Petrucci, "*A neural word embeddings approach for multi-domain sentiment analysis,*" IEEE Transactions on Affective Computing, vol. 8, no. 4, pp. 457-470, 2017.

[41] D. Bahdanau, K. Cho, and Y. Bengio, "*Neural machine translation by jointly learning to align and translate,*" arXiv preprint arXiv:1409.0473, 2014.

[42] J. Li, M.-T. Luong, D. Jurafsky, and E. Hovy, "*When are tree structures necessary for deep learning of representations?,*" arXiv preprint arXiv:1503.00185, 2015.

[43] L. Stappen, A. Baird, E. Cambria, and B. W. Schuller, "*Sentiment analysis and topic recognition in video transcriptions,*" IEEE Intelligent Systems, vol. 36, no. 2, pp. 88-95, 2021.

[44] M. Rahman, Y. Watanobe, and K. Nakamura, "*A Bidirectional LSTM Language Model for Code Evaluation and Repair,*" Symmetry, vol. 13, no. 2, p. 247, 2021.

[45] K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "*A novel context-aware multimodal framework for persian sentiment analysis,*" arXiv preprint arXiv:2103.02636, 2021.

[46] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "*SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis,*" in Proceedings of the 29th ACM international conference on information & knowledge management, 2020, pp. 105-114.

[47] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "*A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks,*" Neurocomputing, vol. 380, pp. 1-10, 2020.

[48] B. Huang, Y. Ou, and K. M. Carley, "*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, 2018, pp. 197-206: Springer.

[49] D. Tang, B. Qin, X. Feng, and T. Liu, "*Effective LSTMs for target-dependent sentiment classification,*" arXiv preprint arXiv:1512.01100, 2015.

[50] K. Wang, W. Shen, Y. Yang, X. Quan, and R. Wang, "*Relational graph attention network for aspect-based sentiment analysis,*" arXiv preprint arXiv:2004.12362, 2020.

[51] C. Yang, H. Zhang, B. Jiang, and K. Li, "*Aspect-based sentiment analysis with alternating coattention networks,*" Information Processing & Management, vol. 56, no. 3, pp. 463-478, 2019.

[52] D. Meškelė and F. Frasincar, "*ALDONAr: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model,*" Information Processing & Management, vol. 57, no. 3, p. 102211, 2020.

[53] S. S. Sharma and G. Dutta, "*SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination,*" Information Processing & Management, vol. 58, no. 1, p. 102412, 2021.

[54] I. A. Farha and W. Magdy, "A comparative study of effective approaches for arabic sentiment analysis," Information Processing & Management, vol. 58, no. 2, p. 102438, 2021.

[55] Y. Tay, L. A. Tuan, and S. C. Hui, "*Dyadic memory networks for aspect-based sentiment analysis,*" in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 107-116.

[56] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "*Bert: Pre-training of deep bidirectional transformers for language understanding,"* arXiv preprint arXiv:1810.04805, 2018.

[57] C. Olah, *"Understanding LSTM Networks.[(accessed on 25 February 2021)],"* ed, 2015.

[58] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, *"Semeval-2015 task 12: Aspect based sentiment analysis,"* in Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), 2015, pp. 486-495.

[59] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, *"Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews,"* Journal of computational science, vol. 27, pp. 386-393, 2018.

[60] K. S. Tai, R. Socher, and C. D. Manning, *"Improved semantic representations from tree-structured long short-term memory networks,"* arXiv preprint arXiv:1503.00075, 2015.

[61] Y. Tay, L. A. Tuan, and S. C. Hui, *"Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis,"* in Proceedings of the AAAI

conference on artificial intelligence, 2018, vol. 32, no. 1.

[62] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, *"Sentiment analysis using deep learning approaches: an overview,"* Science China Information Sciences, vol. 63, no. 1, pp. 1-36, 2020.

[63] J. Zhou, J. X. Huang, Q. Chen, Q. V. Hu, T. Wang, and L. He, *"Deep learning for aspect-level sentiment classification: survey, vision, and challenges,"* IEEE access, vol. 7, pp. 78454-78483, 2019.

[64] S. Zheng and R. Xia, *"Left-center-right separated neural network for aspect-based sentiment analysis with rotatory attention,"* arXiv preprint arXiv:1802.00892, 2018.

[65] W. Xue and T. Li, *"Aspect based sentiment analysis with gated convolutional networks,"* arXiv preprint arXiv:1805.07043, 2018.

[66] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, *"Content attention model for aspect based sentiment analysis,"* in Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1023-1032.

[67] X. Wang, G. Xu, J. Zhang, X. Sun, L. Wang, and T. Huang, *"Syntax-directed hybrid attention network for aspect-level sentiment analysis,"* IEEE Access, vol. 7, pp. 5014-5025, 2018.

[68] J. Liu and Y. Zhang, *"Attention modeling for targeted sentiment,"* in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 2017, pp. 572-577.

[69] Z. Lin et al., *"A structured self-attentive sentence embedding,"* arXiv preprint arXiv:1703.03130, 2017.

[70] Y. Song, J. Wang, Z. Liang, Z. Liu, and T. Jiang, *"Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference,"* arXiv preprint arXiv:2002.04815, 2020.

[71] F. Fan, Y. Feng, and D. Zhao, *"Multi-grained attention network for aspect-level sentiment classification,"* in Proceedings of the 2018 conference on empirical methods in natural language processing, 2018, pp. 3433-3442.

[72] A. Kumar, V. T. Narapareddy, V. A. Srikanth, L. B. M. Neti, and A. Malapati, *"Aspect-based sentiment classification using interactive gated convolutional network,"* IEEE Access, vol. 8, pp. 22445-22453, 2020.

[73] J. Liu, P. Liu, Z. Zhu, X. Li, and G. Xu, *"Graph Convolutional Networks with Bidirectional Attention for Aspect-Based Sentiment Classification,"* Applied Sciences, vol. 11, no. 4, p. 1528, 2021.

[74] J. Cheng and H. Yi, *"Aspect-Based Sentiment Analysis Based on Multi-Channel and Dynamic Weight,"* in IOP Conference Series: Earth and Environmental Science, 2021, vol. 693, no. 1, p. 012065: IOP Publishing.

[75] J. Wang et al., *"Aspect Sentiment Classification with both Word-level and Clause-level Attention Networks,"* in IJCAI, 2018, vol. 2018, pp. 4439-4445.

[76] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, *"Effective attention modeling for aspect-level sentiment classification,"* in Proceedings of the 27th international conference on computational linguistics, 2018, pp. 1121-1131.

[77] M. Hu et al., *"Can: Constrained attention networks for multi-aspect sentiment analysis,"* arXiv preprint arXiv:1812.10735, 2018.

[78] A. Essebbar, B. Kane, O. Guinaudeau, V. Chiesa, I. Quénel, and S. Chau, *"Aspect Based Sentiment Analysis using French Pre-Trained Models,"* in ICAART (1), 2021, pp. 519-525.

[79] H. Le et al., *"Flaubert: Unsupervised language model pre-training for french,"* arXiv preprint arXiv:1912.05372, 2019.

[80] C. Sindhu and G. Vadivu, *"Fine grained sentiment polarity classification using augmented knowledge sequence-attention mechanism,"* Microprocessors and Microsystems, vol. 81, p. 103365, 2021.

[81] P. Chen, Z. Sun, L. Bing, and W. Yang, *"Recurrent attention network on memory for aspect sentiment analysis,"* in Proceedings of the 2017 conference on empirical methods in natural language processing, 2017, pp. 452-461.

[82] Y. LeCun and Y. Bengio, *"Convolutional networks for images, speech, and time series,"* The handbook of brain theory and neural networks, vol. 3361, no. 10, p. 1995, 1995.