Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Deep learning based Text Emotion Recognition for Chatbot applications

Mounika Karna[a], Sujitha Juliet D.[b], R.Catherine Joy[c]
[a] BTech Student, Dept of CSE, Karunya Institute of Technology and Sciences, Coimbatore
[b]Assistant Professor, Dept of CSE, Karunya Institute of Technology and Sciences, Coimbatore
[c]Assistant Professor, Dept of ECE, Karunya Institute of Technology and Sciences, Coimbatore

Abstract— **Emotions play a vital role in human interaction. We recognize emotion of a person from their speech, face gesture, body language and sign actions. Since humans use many text devices to make interactions these days, emotion extraction from the text has drawn a lot of importance. It is therefore crucial that emotions in textual conversation need to be well understood by the machines, which ultimately provide users with emotional awareness feedback. This paper investigates the effectiveness of deep learning based Long Short-Term Memory mechanism for identification of textual emotions. The study was carried out on 'Emotion classification' dataset with six emotional groups. The experimental results proved that LSTM based text emotion classification provides relatively higher accuracy compared to the existing learning methods.**

Index Terms— **Emotion recognition, Text based analysis, Online Chatbot, text mining, LSTM, deep learning**

I. INTRODUCTION

Emotion is a part of Human life. It persists in various situations such as inspiration, thoughts, awareness, imagination, focus, thinking, experience and effective decision- making. People occasionally express their emotions through conscious or subconscious from their face countenances, words and text – messaging. Whereas, most of them choose the text in the form of emails, commenting on items, weblogs as an interface for communication. To politicians, economists, market analysts and social scientists, identification of emotions from text can be of interest. Detection of emotional status which uses text has many practical knowledge in fields such as ITS [Intelligent tutoring system], authentication of a user, email, filling out online forms, review for a product and moreover emotional developmental applications.

Our current research is focused on casual interactions with users, in the context of an online chat-bot. In this case, we note that users often show a wide range of emotions, such as being nervous about studying, being desperate for a new career, feeling depressed about wanting to break up, etc. The software-human variations vanish in such circumstances and participants expect software to thoroughly interpret people's psychology and feelings. Understanding these emotions and developing an emotionally attentive approach enable a deep sustaining relationship with users and to know their emotional desires.

This paper explores the effectiveness of LSTM based deep learning method for text emotion recognition. An LSTM has a control flow which processes data that is pass ed on as it propagates. The variations are operations inside the cells of the LSTM. Such operations are used to encourage the LSTM to keep information or to forget it. Without using expressions from face and speech transitions, interpreting text based emotion identification is a demanding issue.

The continuation of the paper is arranged as follows: Section 2 provides a summary of the related works. Our investigation is described in detail in Section 3. Section 4 discusses the findings and our experimental setup. The paper ends up in Section 5.

II. RELATED WORKS

Generally, Emotion recognition methods are classified into three subgroups: Methods based on lexicons, Methods based on rules and Methods based on learning. Methods based on

Mounika Karna[a] is the B.Tech student from the Dept of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India, e-mail: mounikakarna@karunya.edu.in
Sujitha Juliet D. [b] is the Assistant Professor from the Dept of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India, e-mail: sujitha@karunya.edu
R. Catherine Joy[c] is the Assistant Professor from the Dept of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India, e-mail: catherinejoy@karunya.edu

lexicons are applied at vocabulary stage, using lexical effective techniques such as the Word Net Affect. Methods based on rules differ in syntax and grammar of the language and language properties are used by learning based methods.

Lexicon based approaches are ways of defining emotions using only one or many lexical tools. One of those methods is focused on keywords which use a predetermined set of words to categorize text into emotional groups. Hatzilygeroudis and Perikos [1] have developed a predictive algorithm which examines the emotional words of the sentence and determines the level in the text representing those psychological terms and determines their strength depending on the ranking obtained. Balahur et al. [2] proposed an ontology-related framework which EmotiNet used as a guide to differentiate implicit emotions from context. The research methods also originate from lexicon-focused models. Gill et al. [3] used co-occurrence semantic domain techniques such as Latent Semantic Analysis (LSA) and Hyperspace Analog to Language (HAL) to dynamically calculate the semantic similarity among the records and terminology. Rachman et al. [4] proposed a context-based approach in which the data set is built using the Word Net Concern Feelings and the Psychological Standards for Words and phrases. They used the Allocation of Latent Dirichlet to dynamically expand the data collection. The Ekman emotional model was used for annotation of the corpus in their proposed method.

Given the importance of the language design, such as the key aspect of the sentences, solutions based on the lexicon are very simple but rarely used. For example, the phrase "This book is unbelievably and hilariously wonderful" implies that the usage of negative words is twice as many positive words, but the phrase has a very positive sense. Moreover, words may also have various meanings, and may vary as per the context and use its context.

Language learners define some of the policies based on grammatical structure in rule-based system. Many methods set up rules by using a dictionary that includes effective words, and others without using such a dictionary. The drawbacks of this approach may be a lot of ambiguity in designing and rewriting the rules, absence of exploring feelings from defined rules and dependency on language. Shaheen et al. [5] suggested a framework based on the grammatical and contextual structure of the input sentence. Current approach of their framework uses different ontologies to create rules from the input sentences, such as Word Net and Concept Net. Instead texts are categorized into psychological groups by contrasting the principles created to a list of standard guidelines for the obtained training. Sandhya and Anusha. [6] proposed a hybrid, rules-based approach to learning. We also defined certain rules based on the grammatical characteristics of text. Their current method used deep learning techniques (including SVM and Naïve Bayes) to increase the efficiency of classifiers based on learning. Ekman Emotional Model as well as ISEAR database have been used for their research.

Training-based techniques approach the issue of sentiment classification as an identification issue, and do not have to understand the sentence structure of the statement. Machine learning algorithms that are both monitored and unmonitored, used to classify emotions. In the monitored methods, sets of data input must be compiled. Use of the emotional tags with text annotation will be done. Annotation is a key drawback in this field of the supervised methods. In addition, the latest work related to sentiment analysis in Twitter messages is that classification of these messages which is done manually via hashtags. Balabantaray et al. [7] created a psychological classifier that distinguishes emotions in the Ekman model using dual-class SVM. Rao and at.[8] a hybrid method was applied on the basis of dual monitored, complex topic frameworks. The very first method allocated concepts to specific feelings, and the second version described the topics relating to dual terms and emotional states to each other.

In order to create a prototype for emotional classification, unmonitored architectures attempt to find the embedded frameworks in unsorted information. The drawbacks of such techniques are the shortage of conceptual and semantic knowledge and monitored techniques. This approach even uses dual models namely categorical as well as dimensional. In the case of non-monitored approaches using dimensional model, referring to "Feelings in text: categorical and dimensional models" uses the three dimensional (3-D) model:

They used BOW [Bag of Words] techniques to create feature vectors and reduce the size of the dimensions used by methods such as LSA, PLSA and NMF. Firstly, the body language of a phrase is derived from the sentimental parts using a key phrase-based approach. The phrase concepts and dependencies are observed using knowledge-based approach. Essentially, using methods of classification predictive machine learning (Maximum Entropy and Naïve Bayes) determines the sentence if state is emotional.

Though many researchers have investigated on the emotion techniques, there is still a room for further development because of its demand. Therefore this paper aims at developing a method on LSTM based emotion recognition.

III. Proposed Method

A. Pretreatment

There are measures necessary to identify emotion. The first step as shown in Fig 1 provides pretreatment for the given dataset. Punctuations are omitted in this analysis and words are translated into lowercase letters. The punctuation marks such as [! # $ % &" (*) +,-/\: "< = > < @\^ ' { } ~)] are also eliminated from the pre-training process.

Every word throughout the expression is represented by an integer in which the integer is unique for each word. Instead, in the beginning, add the padding "0" so that each sentence is of the same length. When that phase is done, the numerical will be an input to the neural network. In this preprocessing phase, we also calculate the actual word count in the training data and the maximum word count in a single sentence. On the other hand, to be able to do the classification, SVM's benchmark method needs an extraction of a feature. For the SVM we are using TF-IDF as features.

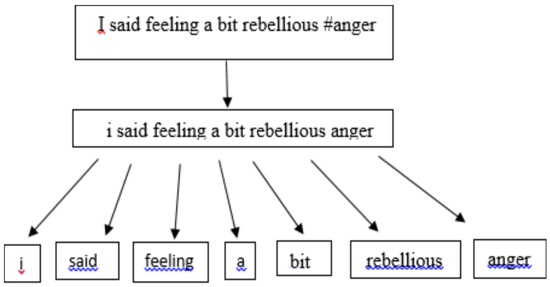

Fig. 1. Preprocessing for LSTM

B. *Emotion Classification*

Fig. 2 Displays a LSTM simulation diagram using KERAS software. The following layers forms LSTM for emotion classification: Embedding layer, LSTM stratum and output layer.

Some parameters are necessary for the embedding layer as its input. There is only one neuron to the embedding layer. The word transmitted into this neuron will be converted into a true-valued vector (output dimension) of required lengths. Once the network has been equipped we will obtain the masses of the embed layer. It implies that there is a true value length vector for every term. The embed process is implemented as perword2vec embedding approach from Mikolov et al.

FOR EXAMPLE:

Consider the below input sentence which is given. Here firstly each sentence is segmented into a word in which further processing of the sentence happens through LSTM with the help of Softmax process and finally emotion recognition from the input sentence will be obtained.

INPUT GIVEN:

I believe that I am more sensitive to other people feelings and tend to be more compassionate

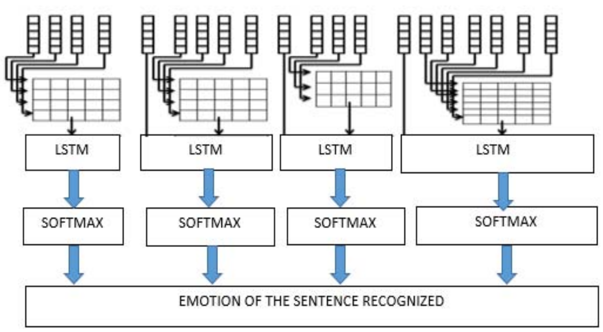

Fig.2 Deep Model Structure

The person's emotional recognition from any of his / her outputs, including text is a nonfixed behaviour, so we can infer from the text that the text's predominant emotion is one of them.
The key shortcoming of existing methods is considering the precise and defined parameters for emotions to identify emotions from the text.

In other words, the existing methods only assign each text to one of the e motional categories.
The aim of this paper is to recognize all existing emotions in the text, and to determine the predominant emotion of the sentence only with minimal feature technology.
To achieve that goal, we are proposing a hybrid deep learning model to a utomatically learn features.
Our proposed model uses phrase structural knowledge and deep learning models in a hierarchical manner to provide sentence level functionality.
Our proposed multilabelling approach has the following steps:
1. Segmentation of sentences done for identification of emotional recognition through the sentence.
2. Word representation through use of word embedding method
3. The primary determination of emotion by the sentence using a Long-Short-Term-Memory (LSTM) network.

Formal, a sentence consists of n sections as{ p1, p2,..., pn} and m words as{ x1, x2,..., xm} are included in each portion.
According to the language grammar, different parts of the sentence are usually linked together in combination with words and the word used in conjunction specifies the form of relation parts together.

In the second phase, we transform words{ x1, x2,..., xn} into continuous vector representations by using word embedding methods where d is the embedding dimension of the term. Each component can now be interpreted as a matrix of embedding m d M × ∈ℜ . To create dense representations of each component we used three convolutionary filters. we set the first, second and third filters { d K1 ∈ℜ , d K 2 2 ∈ℜ , d K 3 3 ∈ℜ } to represent unigrams,bigrams & trigrams respectively.Convolution performs on embedding matrix M and is genera ted by equation (1) with features $C_{i,j}$.

$$c_{j,i} = f(x_{i:i+j-1}.K_j^T + b) \qquad (1)$$

Where f is the non-linear activation function and b is the bias term.
Average pooling layer is used (Equation (2)) to combine the varying number of features from the convolution layer into a vector.

$$C_j = \frac{1}{m} \sum_{i=1}^{m} c_{j,i} \qquad (2)$$

Finally, the concatenation of three filters is used as part representation and decoded into probabilities by a softmax layer for each group of emotions.

In the last step LSTM is used to assess the prevailing emotion of a sentence. LSTM will catch the sentence's long-range dependencies, and let us use the connections between each component. In other words, we can capture the type of connection within the sentence for each portion. The extracted representation of each component that uses CNN and the current conjunction word embedding is fed into a forward LSTM network as input for a sentence with n parts and n-1 conjunction words (Si={ p1, p2,..., pn, con1, con2-, conn-1}). LSTM calculates the hidden state by taking the following combination of three gates: input (it), forget (ft) and output (ot).

$$x = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$

$$f_t = \sigma(W_f . X + b_f)$$
$$i_t = \sigma(W_t . X + b_t) \quad (3)$$
$$o_t = \sigma(W_O . X + b_O)$$
$$c_t = f_t \otimes c_{t-1} + i_t \tanh(W_c . X + b_c)$$
$$h_t = o_t \otimes \tanh(c_t)$$

Where $\otimes$ is element-wise multiplication, Wand b select and remove history state vectors and input vectors.

The last LSTM output contains the information from previous sections, and it is used as the final vector. The extracted feature vector is decoded into probabilities for each emotion category by a linear layer and a softmax layer, and the higher-probability category is chosen as the predominant emotion of the sentence. Fig.3 sets out an example of our deep model structure.

IV. Experiment And Result

A. Experiment

During LSTM Training we documented every epoch of the model as a checkpoint for precision and loss development. We set the maximum number of the epoch to 2 The Check's precision and loss outputs of the LSTM-points are noted.

```
Train on 375128 samples, validate on 41681 samples
Epoch 1/2
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:190: The na

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:197: The na

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:207: The na

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:216: The na

WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorflow_backend.py:223: The na

375128/375128 [==============================] - 4472s 12ms/step - loss: 0.2037 - acc: 0.9010 - val_loss: 0.09
Epoch 2/2
375128/375128 [==============================] - 4471s 12ms/step - loss: 0.0907 - acc: 0.9415 - val_loss: 0.09
<keras.callbacks.History at 0x7f1dfd41cf60>
```

Fig. 3. Loss Outputs and accuracy of LSTM during Training.

B. Testing Result

For the LSTM process Table III ,Table V, Table VII provides a matrix with uncertain properties. In this table, 1 is for Rage; 2 is for Worry; 3 is for Happiness; 4 is for Affection; 5 is for sadness; 6 is for astonishment and 7 is for kindness. The LSTM achieves overall accuracy of 94.15 %.

Table IV shows efficiency, recall and f1-score of the LSTM process. The efficiency of the LSTM method and its overall average for each class is calculated on the basis of the matrix of uncertainties. The LSTM method achieves an estimated 94.7 percent efficiency and an average 94.2 percent sensitivity (recall), thus an average f1 score of 94.1 percent.

TABLE. III. LSTM UNCERTAINITY MATRIX

		Estimated						
		1	2	3	4	5	6	7
Current	1	11536	13	53	20	103	1	11
	2	39	5650	45	6	35	0	4
	3	56	46	51648	65	55	1	16
	4	37	10	56	14604	30	0	5
	5	236	47	543	88	63648	9	13
	6	6	3	8	4	5	1564	4
	7	8	6	44	15	18	1	4657

TABLE. IV. EFFICIENCY, RECALL, AND F1-SCORE OF LSTM

Group	Efficiency	Sensitivity	F1-Ranking
Rage	95.9%	94.8%	95.1%
Worry	95.48%	95.33%	95%
Happiness	90.4%	91.24%	90.13%
Affection	96.36%	95.62%	95.4%
Sadness	96.5%	95.78%	96.50%
Astonishment	94%	93.25%	94%
Kindness	94.8%	93.41%	93%
Average	**94.7%**	**94.2%**	**94.1%**

TABLE. V. NESTED LSTM UNCERTAINITY MATRIX

		Estimated						
		1	2	3	4	5	6	7
Current	1	42506	11	42	19	255	4	3
	2	32	4631	54	14	43	7	2
	3	45	14	31274	67	19	3	8
	4	15	7	28	2467	66	7	4
	5	122	96	136	87	36234	5	1
	6	5	9	41	2	61	5382	3
	7	32	6	13	51	17	4	2538

TABLE. VI. EFFICIENCY, RECALL AND F1-SCORE OF NESTED LSTM

Group	Efficiency	Sensitivity	F1-Ranking
Rage	93.23%	92.5%	93.2%
Worry	93.26%	93.42%	92.54%
Happiness	90.30%	90.2%	90.7%
Affection	91.62%	91.5%	90 %
Sadness	92.27%	92.64%	91.6%
Astonishment	96.52%	95.4%	96.26%
Kindness	90.2%	90.8%	91.3%
Average	**92.4%**	**92.3%**	**92.2%**

Table V displays an uncertainty matrix for Model Nested LSTM. Nested LSTM model is also evaluated using emotion classification dataset. The Nested LSTM reaches a 92.34 percent overall accuracy.

Table VI displays efficiency, sensitivity and f1-score ofNested LSTM systems. The efficiency of the Nested LSTM method on each classification and its overall average was estimated using the uncertainity matrix. The Nested LSTM approach produces an average 92.4 percent efficiency and 92.3 percentage sensitivity and an average f1 score of 92.2 percent.

For the SVM method an uncertainty matrix is shown in Table VII. Also, the SVM model is validated utilizing dataset from emotion classification. The average performance of the SVM is 93.1 percent. The SVM method thus generates the low accuracy with respect to the LSTM model.
Table VIII demonstrates efficiency, sensitivity, and f1-score of SVM processes. On the basis of the uncertainty matrix the performance of the SVM method for each class and its exceptional performance are calculated. The SVM method produces an average 93 percent efficiency and an average 93.03 percent sensitivity, hence an overall 92.9 percent f1 ranking.

Table IX demonstrates measurements of efficiency, accuracy, sensitivity and f1score between two versions. LSTM obtains better test grades in spite of efficiency, accuracy, sensitivity and f1-score compared to the SVM.

TABLE. VII. SVM UNCERTAINITY MATRIX

		Estimated						
		I	II	III	IV	V	VI	VII
Current	I	48496	5	34	8	32	7	2
	II	6	2671	54	4	64	3	5
	III	23	74	13624	624	863	87	86
	IV	35	68	124	8624	753	93	29
	V	74	93	83	77	13257	2	8
	VI	6	3	67	4	9	3557	5
	VII	3	1	76	28	64	6	3483

TABLE. VIII. EFFICIENCY, RECALL AND F1-SCORE OF SVM

Group	Efficiency	Sensitivity	F1-Ranking
Rage	97.5%	96.4%	97.5%
Worry	95.2%	94%	94.9%
Happiness	95.63%	95.46%	95%
Affection	93.1%	92.6%	93.35%
Sadness	90.7%	91.6%	90%
Astonishment	91.3%	90%	90.13%
Kindness	90.4%	91.16%	90.1%
Average	**93%**	**93.03%**	**92.9%**

TABLE. IX. OVERVIEW OF PERFORMANCE ANALYSIS

Model	Accuracy	Efficiency	Recall	F1-Ranking
LSTM	94.15%	94.7%	94.2%	94.1%
Nested LSTM	92.34%	92.4%	92.3%	92.2%
SVM	93.1%	93%	93.03%	92.9%

CONCLUSION

Throughout this study we discussed our work on the use of LSTM to identify emotions based on text. LSTM, SVM, Nested LSTM methods can be used to identify emotions in multiclass based on the results of the discussion and evaluation conducted in the previous section. LSTM has the best accuracy among accuracy methods. LSTM has the best average performance in terms of efficiency, sensitivity and f1score at 94.7%, 94.2%, and 94.1% respectively.
Experimental results show that our superiority model overtakes others in terms of accuracy. In future works, we expect to test as well as use other more sophisticated deep learning models to find the best approach to emotion- recognition tested in a more difficult dataset.

REFERENCES

[1] I. Hatzilygeroudis and I. Perikos, "Recognizing emotion presence in natural language sentences," in Proceedings of International Conference on Engineering Applications of Neural Networks, pp. 30-39, 2013.

[2] A. Balahur, J. M. Hermida, and A. Montoyo, "Detecting implicit expressions of sentiment in text based on commonsense knowledge," in Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pp. 53-60, 2011.

[3] A. Gill, R. M. French, D. Gergle, and J. Oberlander, "Identifying emotional characteristics from short blog texts," in Proceedings of the 30th Conference on Cognitive Science, BC Love, K. McRae, and VM Sloutsky, eds, pp. 2237-2242, 2008.

[4] F. H. Rachman, R. Sarno, and C. Fatichah, "CBE: Corpus-based of emotion for emotion detection in text document," in Proceedings of the 3rd International Conference on Information Technology, Computer, and Electrical Engineering, pp. 331-335, 2016.

[5] S. Shaheen, W. El-Hajj, H. Hajj, and S. Elbassuoni, "Emotion recognition from text based on automatically generated rules," in Proceedings of International Conference on Data Mining, pp. 383- 392, 2014.

[6] Xiao Sun,Xiaoqi Peng & Shuai Ding, "Emotional Human Machine Conversation Generation based on Long Short- term Memory in Cognitive Computation, Vol. 10,pp 389- 397,December 2017.

[7] R. Balabantaray, M. Mohammad, and N. Sharma, "Multi- class twitter emotion classification: A new approach," International Journal of Applied Information Systems, vol. 4, pp. 48-53, 2012.

[8] Y. Rao, J. Pang, H. Xie, A. Liu, T.-L. Wong, Q. Li, et al., "Supervised Intensive Topic Models for Emotion Detection over Short Text," in Proceedings of International Conference on Database Systems for Advanced Applications, pp. 408-422, 2017.

[9] Bernhard Kratzwald,Suzana Ilić,Mathias Kraus,Stefan Feuerriegel,Helmut Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support" in Decision Support Systems,Elsevier, November 2018.

[10] Hossein Ebrahimpour-Komleh., Soghra Lazemi*, "Multi-Emotion Extraction from Text Using Deep Learning", March 2018.

[11] Ming-Hsiang Su,Chung-Hsien Wu,Kun-Yi Huang,Qian- Bei Hong "LSTM-based Text Emotion Recognition Using Semantic and Emotional Word Vectors" ACIIAsia.2018.8470378 on Affective Computing and Intelligent Interaction (ACII Asia).

[12] "Multi-Modal Emotion recognition on IEMOCAP Dataset using Deep Learning", arXiv:1804.05788v1 [cs.AI] 16 Apr 2018.

[13] Daniel Haryadi1 , Gede Putra Kusuma2 "Emotion Detection in Text using Nested Long Short-Term Memory", 11480 (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[14] Xiao Sun,Xiaoqi Peng & Shuai Ding, "Emotional Human Machine Conversation Generation based on Long Short- term Memory in Cognitive Computation, Vol. 10,pp 389- 397,December 2017.

[15] Navonil Majunder ; Soujanya Poria ; Alexander Gelbukh; Erik Cambria, "Deep Learning based document modelling for Personality Detection from Text" in IEEE Intelligent Systems, Vol. 32,pp 74 – 79, March 2017.

[16] Che-Wen Chen, Shih-Pang Tseng Ta-Wen Kuan "Outpatient text classification using Attention based Bidirectional LSTM for Robot assisted servicing in hospital" February,2020.

[17] Marco Polignano, Pierpaolo Basile, Marco de Gemmis, "A Comparison of Word-Embeddings in Emotion Detection from text using Bi-LSTM, CNN and Self-Attention" in User Modelling,Adaptation and Personalization, pp 63-68,June 2019.

[18] Erdenebileg Batbaatar, Meijing Li, Keun Ho Ryu, "Semantic-Emotion Neural Network for Emotion Recognition from text" , in IEEE Access, Vol. 7,pp 111866-111878,August 2019.

[19] Mohammad Al-Smadi Bashar Talafha Mahmoud Al-Ayyoub & Yaser Jararweh, "Using Long Short-Term Memory deep neural networks for Aspect-based sentiment analysis of Arabic reviews" in International Journal of Machine Learning and Cybernetics,Vol. 10 ,pp 2163-2175, March 2018.

[20] Md Shad Akhtar, Dushyant Singh Chauhan, Deepanway Ghosal, Soujanya Poria, Asif Ekbal, Pushpak Bhattacharyya,and Sentiment Analysis", Computation and Language,arXiv:1905.05812,May 2019.