International Journal of Future Generation Communication and Networking
Vol. 13, No. 1s, (2020), pp. 321-- 330

Emotion Detection through Speech using Bidirectional LSTM and Attention Mechanism

Megha Agarwal[1], Pranav Bhatt[2], Aditi Gupta[3], Kavita Bathe[4]
[1,2,3,4]*Department of Computer Engineering, K.J. Somaiya Institute of Engineeing and IT, Sion, Mumbai, Maharashtra, India*

Abstract
Artificial Intelligence is growing and developing at an exceptional rate. Artificial machines and robots are incorporated with various ways to handle different scenarios and come up with accurate solutions through artificial intelligence. However, when it comes to taking some decisions based on emotions and including emotional quotient in the decision-making process, artificial machines face some issues. Apart from this, embedding emotions into Artificial intelligence just widens the scope for various further researches. To work on improving the emotional aspect in artificial intelligence systems, we need to first tackle the issue of detecting emotions with least possible errors. In this paper the aim is to find ways to improve upon accuracy in emotion detection through deep learning. Deep learning methods work by processing a vast database gathered from a number of sources. The analysis initiates by vectorizing each word in the input given by the user and deriving the meaning of the words in both, forward and backward direction. Upon understanding the meaning, attention mechanism defines the weights to be assigned to the words based on the importance they carry. This results in a maximum pooling of the highest weight vectors. The vectors then proceed to be classified in one of the six major emotions.
Keywords: *Bidirectional LSTM, attention mechanism, word vectorization, word embedding, CNN.*

1. Introduction

In a world where machines are taking over humans and emotional quotient is given equal importance as that of intelligence quotient, it has become necessary for an individual to recognize their own emotions. Emotion recognition refers to the technique of classifying the emotional state possessed by each input sequence taken in the form of either verbal features or facial aspects in order to understand the purpose, thoughts and mental state of a human. The analysis to determine the fraction of different emotions in a desired input sequence requires processing in more than one modes, i.e. analyzing the features of human emotion in the form of textual input, speech input or visual data input. To classify various emotions, the detection is done by incorporating information from various fields of input such as posture changes, gestures made, facial alterations, verbal data and text. Machines can be trained in such a way that they can help us detect emotion and how one is feeling. Currently, artificial systems lack only in terms of emotions, so, in order to make it better, and to help advance in the artificial system technology, this paper intends to design an efficient system-independent algorithm. Learning that there is a lack of emotions in machines, which is a growing technology, this paper proposes a system which detects emotion of an individual using the data provided by them by describing the events that took place throughout their day. Every user that makes use of this system has their unique way of behaving in different situations, which may even be different from the response that is generally expected, however, the use of personalised model is a practice that possesses great importance in making the user experience and accuracy very high. The goal is to develop an efficient and optimised algorithm with high accuracy using a combination of various existing algorithms taking input in various formats such as speech and text. The system takes a speech input which is then converted to text and then emotion is detected and categorizes emotion as 'Joy', 'Sad', 'Angry', Surprise, 'Fear', 'Love'. Using collaborative learning, data from various users can be gathered for more efficient and accurate detection. This algorithm can then be implemented in various walks of everyday life where emotion detection can help in enhancing the user experience.

2. Literature Review

The study on method proposed by Lalitha et al [8] explained the processing of seven emotions classified with the use of the high or low tone of the input audio sequence and the manner of uttering

ISSN: 2233-7857 IJFGCN

each syllable. This method used Support Vector Machine(SVM) classifiers to direct towards the final output and database used for this task is Berlin database which resulted in an accuracy of 81%. The classifier used in this method is SVM as it can be used on less quantity of dataset. This method [8] uses Berlin EMODB which comprises 5 males and 5 females consisting of 535 speech files. The method is divided into two parts called as extracting the analytical features and the second one is the module performing the resulting classification. The features that were extracted for analysis were the pitch of the audio input, the prosody in it and the quality possessed by the voice and for classifiers, SVM was chosen amongst HMM (Hidden Markov Model), GMM(Gaussian Mixture Model), ANN(Artificial Neural Networks) because of its usage in low quantity of datasets. SVM was used along with the Radial Basis function to implement this method. However, disgust had a substantially lower identification rate as it is marginally complex in nature to be detected even by a human. The efficiency of the proposed system could also be enhanced even further by using different classifiers instead of SVM [8]. Further, as suggested by Tarunika et al[6], application of Deep Neural Network (DNN) and k- nearest neighbour (k-NN) is done for recognising emotion from audio input data, especially the state where person is scared. In this paper [6], the audible sound signals are converted to be represented as a wave where the level of audibility of the words uttered is extracted, the already stored databases are checked for those features and such analytical processing is performed. Initially, the raw input audio sequence is transformed to be represented in a wave form after being cleaned to get rid of the undesirable features contained in the audio sequence taken as input. This is then processed statistically to understand the different features and their representations which are later labelled as per the detected emotion using the training model selected resulting in the user being able to understand the fraction of emotions contained in the input. The feature extraction is based on Mel spectrogram, Harmonic percussive, Chromagram, Mel frequency cepstral, Beat tracking, Beat-synchronous features aggregation. In order to achieve better results, a combination of audio and video is used as the input which gives 95% accuracy that can be seen as an exceptional result. On the contrary, this paper[6] demands the usage of video analysis along with speech analysis without which the accuracy of the system proposed in this paper [6] drops to 75% and additional measures have to be considered to conduct a video analysis which is beyond the study of this paper.

On further study of the paper by Shahin et al[4], it was understood that it focuses on identifying emotions for a system that does not focus on the speaker or the text using an innovative classifier which is a congregation of two various classifiers, i.e. Gaussian Mixture Model and Deep Neural Networks. The model [4] has been experimented on “Emirati Speech Database” to identify the accuracy and working mechanism of the system resulting in a output divided in six different emotions. The hybrid GMM-DNN classifier [4] when compared with Support Vector Machine(SVM) and multilayer perceptron(MLP) highlighted the difference in accuracy which was learned to be 83.9% for the hybrid model and 80.33% and 69.78% for the other two using SVM and MLP, respectively. The results throw light on how using a hybrid classifier helps increase the detection accuracy as opposed to using just one classifier. Initially feature extraction takes in this method which is followed by tagging of the data into one of the six emotions. Training model is obtained which along with hybrid GMM-DNN is used to deduce emotions from the speaker. The feature extraction phase looks for MFCC which is namely Mel Frequency Cepstral Coefficient. This can be seen as a progression of paper [4]. Similarly enough, this method also fails to recognize the emotion of disgust with utmost accuracy. Proceeding forward to look at the study by Prerna Mishra et al[2], the use of Hidden Markov Model to detect emotions is well explained. This system [2] analyses the emotion on input in the form of video to result in one of the six major emotions. The system makes use of the already existing methods and proposes to use a model of HMMs for extracting the various features of facial expression and splitting them as per differences. This system [2] works by initially detecting the target’s face in the video input sequence where the face is captured by detecting the skin colour and then the facial features of a human such as eyes, nose, etc. The image background can be easily eliminated to focus on the important features of the video. The next step is to perform normalisation with the aim of getting rid of unnecessary noise and modify the face to balance its luminosity and the position of pixels. In the third step of processing, features of importance are highlighted and those inappropriate are removed. The last step, all important gestures captured in the video are labelled as one of the six resulting emotions

ISSN: 2233-7857 IJFGCN

i.e. joy, anger, fear, sorrow, astonish, antipathy. However, this paper [2] gives an overview of the implementation of emotion identification model using HMM in order to have emotions detected at real time from video sequence. Another technique to detect emotions by Reza Lotfian et al.[1], is by creating a reference model, that is neutral in nature, from synthetic speech as opposed to the speech signal. This system [1] determines whether the synthesized speech is any different from the predefined reference template of neutral emotion speech content using feature based processing and interpretability evaluation. It [1] works by designing a synthetic speech sequence to be used as reference, which is then compared to the input audio sequence at each analytical frame. Many speech modifications are obtained using various synthesis approaches and possessing different voices. This is beneficial as it suppresses those aspects of speech that do not contribute towards emotional classification. The keywords spotting technique is also used which converts a set of sentences into a set of various keywords. Although in lexical Dependent Emotion detection, the proposed technique [1] works on the assumption that the lexical data in the sentence is already known by the system. This assumption is valid only for those processing scenarios in which the transcriptions are readily available (e.g., analysis of jury trial). In other cases, the lexical information has to be inferred from speech by using automatic speech recognition [1].

Finally, as mentioned by Ploignano et al[10], the method used for emotion detection is Bidirectional LSTM along with Self-Attention mechanism. The method [10] begins by using an Embedding layer where each word in the sentence is converted to a vector format in order to be easily understood by the computer. The output of this layer is transferred to BiLSTM layer which focuses on the sequential relationship between words in a sentence where the meaning of each word depends on the terms that occur before it. It assigns the weight to each word only on the basis of the weights of the words that have been used in the sentence before it. Once these weights have been assigned, the output is forwarded to the self-attention layer. In this layer, the model is provided with the ability to weigh the separate words of the sentence differently depending on the weight of the tokens in its neighbourhood. The concurrent neural network layer is used next. This layer works to make the obtained input dense and smaller to converge towards the output. The final layer which classifies the input into one of the categories, also known as emotion classification layer, takes the output from CNN layer. It applies a dropout on the output of maximum pooling taking place in the CNN layer and determines the percentage of each emotion in each input [10].

3. Methodology

Upon understanding the drawbacks and advantages of all the systems studied, the system used to detect emotion according to this study is a compilation of two distinct deep learning classifications, long-short term memory networks (LSTM) [10] which can be categorized into two variations unidirectional and bi-directional. The one used in this paper is bi-directional. Besides using LSTM, convolutional neural networks (CNN) [10] are used too which contribute by providing the max pooling approach to make the vectorized and weighted input text denser to classify. Additionally, self-attention layer is also added after the LSTM layer to identify the link between words by recognizing the weights carried by each word depending upon the importance of its contribution to the system.

The knowledge used is divided into 2 sections, the input sequence and the desired emotion on it. This paper intends to design system to give a more efficient and accurate output having optimized algorithms. Thus, the following method is proposed that will take the input from a targeted user in the form of either text or speech. The input is taken in the form of speech or text as desired by the user, in case of speech input, it is first converted to the text format and then used for processing.

The system initially focuses on retrieving speech data. This can be done by acquiring data from a person on call, the speech produced when mindlessly driving, talking to someone in order to put forth their point and similar methods. The acquired input is then converted to text with respect to the spoken words in the speech, like “Hey! I don’t like how my day went”. Each word from the input is segregated and semantic and syntactic analysis is performed on the input.

The input first goes through the embedding layer. Here, the input is first pre-processed to extract only the data of importance and neglecting factors such as hashtags, numbers, etc. On processing through the embedding layer, the output obtained is a vectorized format of the input words in order to make

ISSN: 2233-7857 IJFGCN

them easily understandable by the computer. The output is then padded to ensure equality in the length of all vectors obtained to make the analysis uniform and unbiased. After this, the BiLSTM layer processes the input. The result obtained from the previous layer is first processed through a dropout mechanism to prevent overfitting. The application of BiLSTM follows next, which focuses on assigning weights to the terms in the input sentence depending on the previous terms that occurred in the sentence, because it believes in the fact that the meaning of each word banks on the meanings of all the previous words. This step grants to the layer a memory for figuring the affinity and relations with the previous elements in the input. The bi-directional variant considers the relations among inputs by both the directions.
This can be explained using (Eq. 1).

$$x_i = \overleftarrow{x_i} || \overrightarrow{x_i} \quad x_i \in R^{2d} \tag{1}$$

where || is the operator of concatenation and d is the dimension of the LSTM in terms of hidden units.
The weighted vectors are then passed to the attention mechanism layer where the weights on the words are assigned based on the neighboring vectors. The level of attention can be used to get a jist of what features the network is looking at most during learning and subsequent classification. In this layer, the dense vectors are first distributed based on time and weight and reshaped according to their groups. This helps us understand the weight hierarchy. The next step attention layer is to perform softmax activation. Softmax, and activation function, that converts number logits into probabilities which converge to one. Softmax gives a vector that displays the probability distributions of a list of potential outcomes. The final stage is classification, here the probability outputs from the above layers are presented to the user in a graphical format denoting the fraction of the emotion in the given input text.

ISSN: 2233-7857 IJFGCN

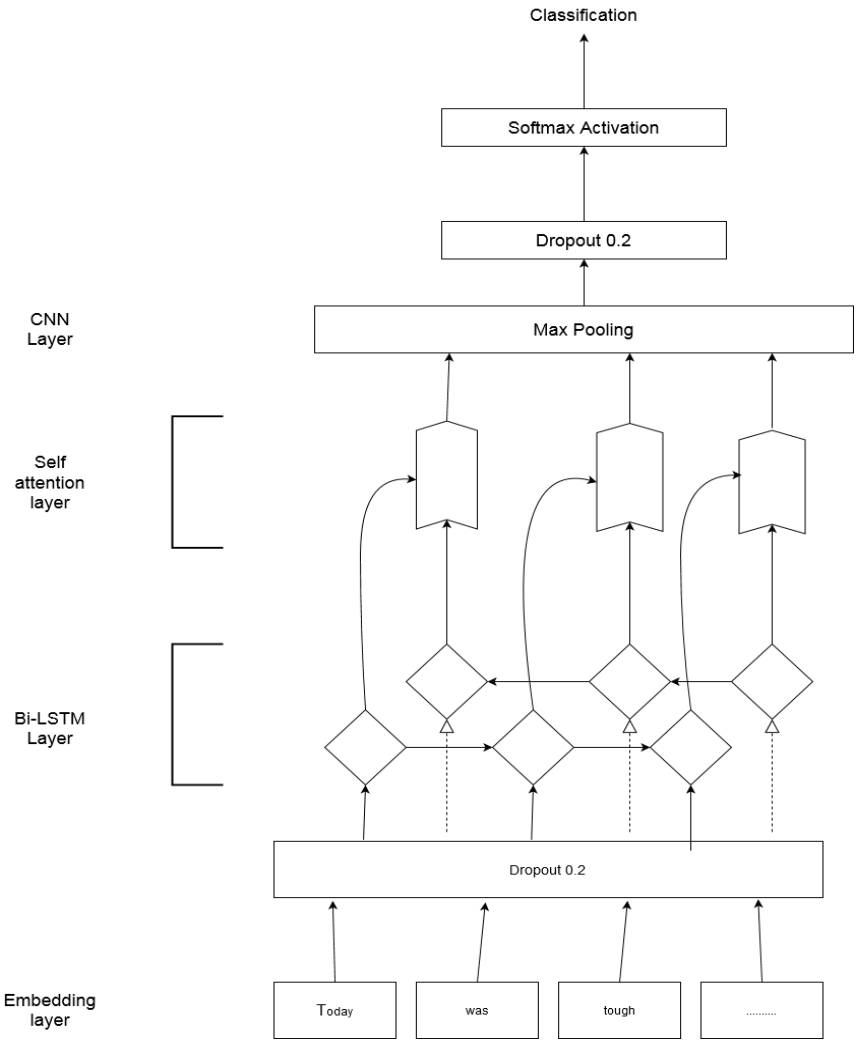

Figure 1. Flowchart of the system

Embedding Layer: This is the initial layer of the system which serves the purpose of converting each word in the input sentences to a vector form in order t make it easily computable by the processing machine. Due to the varying length of each input sentence, the maximum number of terms is considered as max_terms, which is the definite length to which each input sentence is to be converted by adding pad bits as necessary in order to make the processing justified and result in a matrix of correct dimensions to multiply with the vector matrix. On the other hand, for sentences with terms exceeding the number of max_terms, only the initial max_terms terms are considered, rest are discarded. Here, the method used for embedding is a very common method where words are embedded into vector spaces on pre-calculated domains. This enables the user to reach over a vast variety of terms as it reduces the computation cost. To successfully gain the word embeddings, a series of data pre-processing methods are to be applied to convert all data in a standard format easy to process.

Bi-LSTM Layer: This layer takes into consideration the sequential internal relationship amongst each word in the input sentence which means that the meaning of each term depends on the meaning of its previous terms. However, in bi-directional LSTM, the process runs both ways, i.e. first the meanings are derived in a left to right manner such that the meaning of the words lean on the meanings of the words preceding it and then in a right to left manner, such that the meaning of each word depends on the meanings of words after it. For example, consider the sentence “The water in the sea is clear”.

ISSN: 2233-7857 IJFGCN

Here, on encountering the word 'water' when moving in the forward direction only so much can be understood that a noun is being spoken about, but when the same word 'water' is encountered in the reverse approach it is clear that the water of the sea is the topic of focus here. Thus, this step is said to work as a memory for the system. However, as the system uses long short term memory, words upto a limit of 3-5 words in each direction can be related and used to derive meanings. LSTM uses a set of hidden neurons to calculate the weights of each word dynamically.

Self-Attention Layer: This layer is placed right after the LSTM layer. The mechanism at this layer provides the system with an ability to weigh the vectors of each word not independently but based on the weights assigned to the vectors of neighboring terms. It helps get a clear idea of what is the focus of the network at when it is under learning mode and when it is classifying for the output. an additive self-attention context-aware equal to the whole set of words in input is considered (Eq. 2)

$$g_{t,t'} = tanh(W_g h_t + W'_g h'_t + b_g) \quad (2)$$

$$\alpha_{t,t'} = \sigma(W_\alpha g_{t,t'} + b_\alpha)$$

$$a_{t,t'} = softmax(\alpha_{t,t'})$$

$$l_t = \sum_{t'=1}^{n} a_{t,t'} h'_t$$

where, σ is the element-wise sigmoid function, W_g and W'_g are the weight matrices corresponding to the hidden states h_t and h'_t; Wα is the weight matrix corresponding to their non-linear combination; bα and bα are the bias vectors. The attention-focused hidden state representation l_t of a token at timestamp t is given by the weighted summation of the hidden state representation h'_t of all other tokens at timesteps t.

Emotion classification: After the output from max pooling layer is obtained, where the vectors are made denser in order to get rid of the unimportant vectors as decided by the attention layer, a dropout function is applied to dampen the number of connections inside the network and hence, reducing the risk of over-fitting. Dropout is regulatory method that, for a defined value of q, converts the q fraction of units to 0 after each iteration during the training phase. After this, a merge is performed between the outputs until this point to the output obtained from the Bi-LSTM in the previous output. Once the concatenation is performed, relu activation is used to flatten the curve. This is done by making all the negative values as zero and considering the positive values as it is. This is done because negative processing is not what the scope of this paper demands. To end the processing in this layer, a softmax activation function is used which estimates the probability for each output class.

4. Results and Analysis

Table 1. Comparison of the various techniques

Sr No.	Technique	Dataset	Results	Remarks
1	Support Vector machine classifier extracting pitch and prosody of input audio	Berlin Emotional Database which comprises 5 males and 5 females consisting of 535 speech files.	81% accuracy achieved	SVM classifier used, could be further improved
2	Using audio and video to extract features and process input using k-nearest neighbour	Raw acoustic voice dataset	75% accuracy achieved using audio data	Can be further improved to be up to 95% accurate by analyzing based on video data too

ISSN: 2233-7857 IJFGCN

	and deep neural network			
3	The hybrid of Gaussian mixture model and deep neural network is used as a classifier for feature extraction	Emirati speech database (Arabic United Arab Emirates Database)	GMM-DNN classifier provides an accuracy rate of 83.97%, SVM gives about 80.33% and MLP gives 69.78% accuracy	Using a combination of models, the accuracy was higher
4	The image is converted to gray scale and then feature extraction is done using Principle Component Analysis (PCA)	Images are captured from the still video	78% accuracy achieved	Facial data can be used to increase the accuracy
5	Bi-directional LSTM along with attention mechanism and CNN	Text dataset comprising of input sequence and desired emotion	93.9% accuracy	Input in the form of speech and text and analysis using a combination of models

As can be seen from the above table, when support machine classifiers were used on speech files [8], they classified the data as male speech input and female speech input after which the classification of emotion was done. However the accuracy obtained was only 81%. Although this is good, achieving a higher accuracy is not impossible. Further, when k-nearest neighbor and deep neural networks were used for processing [6], on data comprising of speech and facial inputs, an accuracy of 95% was observed. However, this was only possible when the processing was multimodal, i.e. on more than one input features. Thus, it can be understood that as input features increase, the accuracy of detection increases, but this only increases the load on the processing system as learning may take up to a few days and hence, classification may be delayed. Moving ahead, it was learned that using a mixture of models [4] such as Gaussian mixture model and deep neural networks provide a better result as compared to using just one classification model. The next method [2] helps understand that using facial features to detect the emotion helped accurately analyze the emotion of the user as besides speech another analytical feature is used that is the facial configuration. Finally, looking at the method vividly explained and used in this paper, it can be observed that using two mechanisms, namely, bi-directional LSTM and attention mechanism gave the highest accuracy on a speech based input. The aim of this paper is hence achieved, to design a high accuracy emotion detector. This paper comprises the advantages extracted from the reviewed methods, i.e. it uses a combination of mechanisms to process the input to give the classified emotion [4] and it takes as input more than one analytical feature, speech and text, to increase the accuracy of the detection system [2].

ISSN: 2233-7857 IJFGCN

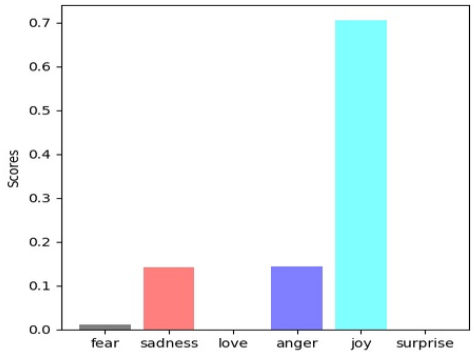

Figure 2. Implementation graph

Figure 2 demonstrates how the system classifies the input sequence "I finally had a talk with my family". It can be understood that naturally a person feels happy talking to their closed ones. However, it should also not be neglected that such an event may even anger a person in some situations besides making them sad in some other situations too. The result obtained is in the form of probability of a given input sequence possessing a particular emotion.

5. Discussion

The results indicate that on applying word embedding to convert words to vector so that they can be processed using bi-directional LSTM method besides applying Attention mechanism to understand how each word is assigned an importance to classify the input sequence into one of the six major emotions is the most accurate and reliable method by far. Moreover, it is well understood that upon taking input in more than one forms, speech and text further support the accuracy of analysis as it provides processing in one more dimension and thus, higher accuracy is obtained.

This paper highlights how the meanings of words are just not based on the meaning of the current word but the words that appear before and after it participate equally in imparting the meaning to any word. Additionally, it is also comprehended that to understand the context of a sequence of sentences only the understanding of the most important words is enough, other words only help in linking the words without imparting any considerable meaning to the verse.

6. Conclusion and Future Scope

The desire to pursue emotion detection and improve upon its accuracy comes fromthe future need for enhanced human-machine interaction. Currently there is a python model which is based on word database and count of words. It is based on tokenization of words and increasing the counter upon encounter words from each specific database. This method which is used as a text based emotion detection and provides excellent accuracy can be implemented in speech emotion recognition as well by implementing a module that converts speech to text. Using word vectorization to extract meanings and relations between words of sentence help analyse input more accurately. Due to the complexity of the project we have been through a lot of literature survey and came across various ways and techniques to implement the emotion detection algorithm. Upon careful analysis we decide to work on aggregating different types of systems and their outputs. Upon careful evaluation we concluded that the text based implementation and speech based implementation if combined could produce improved results.

ISSN: 2233-7857 IJFGCN

However, the system is limited by the size of the training dataset. If a huge dataset containing up to ten lakhs data instances is considered, the accuracy achieved can be higher, but such a vast dataset is not available. Moreover, in case of the absence of negation words in the dataset, the accuracy is hampered because the input sequence from the user might contain such words and if the system does not consider the negation the classified emotion for an user input sequence might be incorrect. Additionally, analyzing the emotion using facial expressions such as smile, growl, etc. or non-verbal aspects such as pauses, sighs, laughter, etc. could help increase the accuracy of the overall system.

References

[1] Reza Lotfian And Carlos Busso, "Lexical Dependent Emotion Detection Using Synthetic Speech Reference" (2019)Received January 14, 2019, accepted January 29, 2019, date of publication February 8, 2019, date of current version March 1, 2019.

[2] Prof. Prerna Mishra, Prof. Saurabh Ratnaparkhi "HMM Based Emotion Detection in Games" (2018) 3rd International Conference for Convergence in Technology (I2CT)The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018

[3] Maruf Hassan, Md. Sakib Bin Alam, Tanveer Ahsan "Emotion Detection from Text Using Skip-thought Vectors" (2018) 2nd Int. Conf. on Innovations in Science,Engineering and Technology (ICISET),27-28 October 2018, Chittagong, Bangladesh.

[4] Ismail Shahin, Ali Bou Nassif and Shibani Hamsa"Emotion Recognition Using Hybrid GaussianMixture Model and Deep Neural Network" (2019)Received January 29, 2019, accepted February 18, 2019, date of publication February 25, 2019, date of current version March 12, 2019.

[5] Imtinan Attili, Mohammad Azzeh and Khaled Shaalan "Speech Recognition Using Deep Neural Networks: A Systematic Review" (2019)Received January 1, 2019, accepted January 24, 2019, date of publication February 1, 2019, date of current version February 22, 2019.

[6] K.Tarunika, R.B Pradeeba, P.Aruna"Applying Machine Learning Techniques for Speech Emotion Recognition"(2018)9th ICCCNT 2018, July 10-12, IISc, Bengaluru, Bengaluru, India

[7] Kun-Yi Huang, Chung-Hsien Wu, Qian-Bei Hong, Ming-Hsiang Su and Yi-Hsuan Chen "Speech Emotion Recognition Using Deep Neural Network considering Verbal and Nonverbal Speech Sounds"(2018)Department of Computer Science and Information Engineering,National Cheng Kung University, Tainan, Taiwan

[8]S.Lalitha, Abhishek Madhavan, Bharath Bhushan, Srinivas Saketh "Speech Emotion Recognition"(2014)International Conference on Advances in Electronics, Computers and Communications (ICAECC).

[9]Oduwa Edo-Osagie, Beatriz De La Iglesia, Iain Lakeand Obaghe Edeghere "Deep Learning for Relevance Filtering in Syndromic Surveillance: A Case Study in Asthma/Difficulty Breathing" (2019)School of Computing Sciences, University of East Anglia, Norwich, United Kingdom, Conference Paper · February 2019.

[10]Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro "A comparison of Word-Embeddings in Emotion Detection from Text using BiLSTM, CNN and Self-Attention" (2019) University of Bari "Aldo Moro", Dept. of Computer Science, UMAP'19, Cyprus

[11] DzmitryBahdanau, Kyunghyun Cho, and YoshuaBengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).

ISSN: 2233-7857 IJFGCN

[12] Alexandra Balahur, Jesus M Hermida, and Andres Montoyo. 2012. Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. IEEE Transactions on Affective Computing 3, 1 (2012), 88–101.

[13] Pierpaolo Basile, Valerio Basile, Danilo Croce, and Marco Polignano. 2018. Overview of the EVALITA 2018 Aspect-based Sentiment Analysis task (ABSITA). Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITAâĂŹ18), Turin, Italy. CEUR. org (2018).

[14] YoshuaBengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. A neural probabilistic language model. Journal of machine learning research 3, Feb (2003), 1137–1155.

[15] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146.

[16] Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. 2019. SemEval-2019 Task 3: EmoContext: Contextual Emotion Detection in Text. In Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval-2019). Minneapolis, Minnesota.

[17] Jianpeng Cheng, Li Dong, and Mirella Lapata. 2016. Long short-term memorynetworks for machine reading. arXiv preprint arXiv:1601.06733 (2016).

[18] Kyunghyun Cho, Bart Van Merriënboer, DzmitryBahdanau, and YoshuaBengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014).

[19] François Chollet et al. 2015. Keras.

[20] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning. ACM, 160–167.

ISSN: 2233-7857 IJFGCN