Human Emotion Recognition System using Deep Learning Techniques

Sreenidhi M J[1], Surabi Sri Dhanya S[2], Sahithi R[3]
[1,2,3]UG Scholar,
Department of Information Technology,
Sri Venkateswara College of Engineering, Pennalur,
Sriperumbudur, Tamil Nadu, India

Ms. Sharon Femi P[4]
[4]Assistant Professor,
Department of Information Technology,
Sri Venkateswara College of Engineering, Pennalur,
Sriperumbudur, Tamil Nadu, India

Abstract—**Using AI to help humans with handling their emotions and identifying their stress levels in the current stressful lifestyle will greatly help them manage their lifestyle. Using the deep learning techniques, it can be made possible by creating a virtual bot to observe and understand human emotions. The virtual chatbot helps to, understand the behavior of people suffering from depression, helps to monitor children's activity without their knowledge and used in literature to fid genre of various books by identifying the words. In this paper, the comments from Reddit are used, preprocessed and trained using Deep Neural Network to learn the emotions of the user. The inference engine module, which is a hybrid network consisting of convolutional neural network and recurrent neural network, is also interfaced. The model provides a high accuracy of response.**

Keywords— Machine learning, artificial intelligence, deep learning, neural network, chatbot.

I. INTRODUCTION

With the boom of artificial intelligence in this modern world, we can imagine future where humans and machine will build the society hand in hand. This is evident in today's world. In such a case, imagine designing a bot to understand the most complex feature of human, emotion. Over the years many researches have been done to decode the complexity of human emotions. While many scientists, poets, artists have always been puzzled and wondered about our most common feature of mankind, why not use a machine to decode the mystery? Using the machine learning techniques will open up a new world of understanding human behaviors and will bridge the gap between the computers and human understandings. While the AI's power has been unleashing to a great extent in the recent years, implementation of cognitive intelligence in bot systems and development of humanoids has got into the play to meet the requirements of the ever developing society. So making the machines to literally understand humans is very essential in the near future.

Also this will be a starting point for an actual sentiment analyzer which can really understand human emotions through words and even by actions by image processing mechanisms. This can act as a sentiment analyzer for social media and business marketing strategies, in order to maintain the product's quality. Hence in this project we are developing a bot which can understand the depth of a human emotion through words, which can produces results to a great extent of accuracy.

The purpose of this project is to give a detailed description about EMOT project. It will illustrate the purpose and complete declaration for the development of system. It will also explain how the system extracts the emotions from the datasets. This document is primarily intended anyone who wants to get an overview of how EMOT works, its outcomes and possible usages in future. EMOT takes input from 1.7 billion Reddit comments which includes texts, emojis, images etc. and then classifies the data based upon their emotions. The classification is done by deep learning mechanism. The classified data forms as a training module for the EMOT. A response model is constructed using Inference engine which is a combination of Convolutional Neural Network (CNN) and recurrent neural Network (RNN) in form of an interactive chatbot. The paper consists of five chapters, Introduction, where a brief introduction about the paper is defined, Literature Review, Structural Architecture of the module and the results of the module.

II. LITERATURE REVIEW

Zhao Jianqiang et al.[1] proposed that emotion-aware mobile applications have been increasing due to their smart features and user acceptability. To realize such an application, an emotion recognition system should be in real time and highly accurate. In this paper, emotion recognition with high performance for mobile applications is proposed. In the proposed system, facial video is captured by an embedded camera of a smart phone. Some representative frames are extracted from the video, and a face detection module is applied to extract the face regions in the frames. The dominant bins are then fed into a Gaussian mixture model-based classifier to classify the emotion. Experimental results show that the proposed system achieves high recognition accuracy in a reasonable time. The merits of this model include enhanced recognition features, easier implementation and quick response. The demerits of this model include less speed, unsuitable for large datasets and increased overhead.

M. Shamin Hossain et al.[2] proposed that the Twitter sentiment analysis technology provides the methods to survey public emotion about the events or products related to them. Most of the current researches are focusing on obtaining sentiment features by analyzing lexical and syntactic features. These features are expressed explicitly through sentiment words, emoticons, exclamation marks, and so on. In this paper, a word embeddings method is used which is obtained

by unsupervised learning based on large twitter corpora, this method using latent contextual semantic relationships and co-occurrence statistical characteristics between words in tweets. The project had made provisions to analyze real time data. The implementation of the model is practical and more realistic. The project also classifies emotions based on emojis. The project involves complex interpretations, increased overhead. The response model is not direct and involves collective responses.

Jian Guo et al.[3] proposed that emotion recognition has a key role in affective computing. A compound facial emotion includes dominant and complementary emotions (e.g., happily-disgusted and sadly-fearful), which is more detailed than the seven classical facial emotions (e.g., happy, disgust, and soon). To address these problems, the iCV-MEFED dataset is released, which includes 50 classes of compound emotions and labels assessed by psychologists. The task is challenging due to high similarities of compound facial emotions from different categories. However, the proposed data set can help to pave the way for further research on compound facial emotion recognition. The model produces accurate feature extraction results. The processing and computing speed were high. There was no real time analysis of data involved in the project. More stable algorithm is needed since the model may be prone to change.

Mondher Bouazizi et al.[4] proposed that with the rapid growth of online social media content, and the impact these have made on people's behavior, many researchers have been interested in studying these media platforms. A major part of their work focused on sentiment analysis and opinion mining. These refer to the automatic identification of opinions of people toward specific topics by analyzing their posts and publications. The dataset was manually labeled and the results of the automatic analysis were checked against the human annotation. The experiments show the feasibility of this task and reach an F1 score equal to 45.9%. The model classifies wide range of data. It provides enhanced real time sentiment analysis. Variety of data has been analyzed. The project module has hierarchical dependency of different algorithms making it complex and less susceptible to maintenance.

Guixian Xu et al.[5] proposed that with the rapid development of Internet technology and social networks, a large number of comment texts are generated on the Web. In the era of big data, mining the emotional tendency of comments through artificial intelligence technology is helpful for the timely understanding of network public opinion. The technology of sentiment analysis is a part of artificial intelligence, and its research is very meaningful for obtaining the sentiment trend of the comments. The essence of sentiment analysis is the text classification task, and different words have different contributions to classification. In the current sentiment analysis studies, distributed word representation is mostly used. The experimental results show that the proposed sentiment analysis method has higher precision, recall, and F1 score. The method is proved to be effective with high accuracy on comments. The

computational power is high yielding accurate results. The model has been designed using simple algorithms. The model classification is more focused on texts rather than other elements such as images, emojis etc.

Xin Kang et al.[6] proposed that understanding people's emotions through natural language is a challenging task for intelligent systems based on Internet of Things (IoT). The major difficulty is caused by the lack of basic knowledge in emotion expressions with respect to a variety of real world contexts. In this paper, a Bayesian inference method is proposed to explore the latent semantic dimensions as contextual information in natural language and to learn the knowledge of emotion expressions based on these semantic dimensions. The Bayesian inference results enable us to visualize the connection between words and emotions with respect to different semantic dimensions. And by further incorporating a corpus-level hierarchy in the document emotion distribution assumption, we could balance the document emotion recognition results and achieve even better word and document emotion predictions. The model is simple to deploy and real time data analysis was made possible. The model can only synthesize texts.

Wanliang Tan et al.[7] proposed that sentiment analysis of product reviews, an application problem, has recently become very popular in text mining and computational linguistics research. Here, the correlation between the Amazon product reviews and the rating of the products given by the customers need to be studied. Both traditional machine learning algorithms including Naive Bayes analysis, Support Vector Machines, K-nearest neighbor method and deep neural networks such as Recurrent Neural Network (RNN), Recurrent Neural Network (RNN) are used. By comparing these results, better understanding of these algorithms can be obtained. They could also act as a supplement to other fraud scoring detection methods. The project involves feedback oriented analysis, making it more realistic and best fit for practical applications. Data contains only customer reviews, because of which model is trained in a unidirectional path.

Tzuu-Heseng S.Li et al.[8] proposed that Facial expression recognition (FER) is a significant task for the machines to understand the emotional changes in human beings. The result of final recognition is calculated using softmax classification. Fine-tuning is effective to FER tasks with a well pre-trained model if sufficient samples cannot be collected. The model contains well-constructed algorithm and is highly stable. It is really slow in computation wise and requires large processing power.

D. Yangm Abeer Alsadoon et al.[9] proposed that Robots must be able to recognize human emotions to improve the human–robot interaction (HRI). This study proposes an emotion recognition system for a humanoid robot. The robot is equipped with a camera to capture users' facial images, and it uses this system to recognize users' emotions and responds appropriately. The emotion recognition system, based on a deep neural network, learns six basic emotions: happiness, anger, disgust, fear, sadness, and surprise. First, a

convolutional neural network(CNN) is used to extract visual features by learning on a large number of static images. Second, a long short-term memory (LSTM) recurrent neural network is used to determine the relationship between the transformation official expressions in image sequences and the six basic emotions. The system is applied to a humanoid robot to demonstrate its practicability for improving the HRI. It was developed as high PDA level model and it was easier to operate. It requires hardware and cross platform supports.

Carlos Busso et al.[10] proposed a study to introduce a method based on facial recognition to identify students' understanding of the entire distance learning process. This study proposes a learning emotion recognition model, which consists of three stages: Feature extraction, subset feature and emotion classifier. Experimental results show that the model proposed in this paper is consistent with the expressions from the learning situation of students in virtual learning environments. This paper demonstrates that emotion recognition based on facial expressions is feasible in distance education, permitting identification of a student's learning status in real time. Therefore, it can help teachers to change teaching strategies in virtual learning environments according to the student's emotions. The model provides virtual environment support. It is highly stable and has high learning rate. The model has high hardware and time complexity and integration is difficult.

III.   PROPOSED ARCHITECTURE OF EMOT
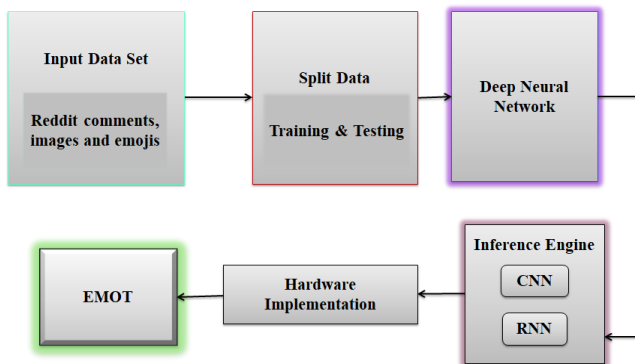
Figure 3.1 Structural architecture of EMOT

The overall architecture of the response model is given in the figure 3.1. The input dataset (about 1.7 billion Reddit comments data) is first split into test and train set. Using Tensorflow model, the classification parameters and parameter tuning is done using Deep Neural Network. The classified data is paired and store in SQLite Database. The data is again fetched for training. The training steps are partitioned by epoch count of 5000 per step. This will improve the accuracy of the response model. Once the model is trained, the inference model, which is a combination of Recurrent Neural Network and Convolutional Neural Network along with Deep Neural framework create a response model. The model provided by these specifications will yield high accuracy of response. Finally the response model with the help of inference framework is deployed as an interactive chatbot in Command Line Interface configuration

(CLI). Further this model will be deployed in an exclusive environment along with hardware integration.

IV.   MODULE DESCRIPTION OF EMOT

The module includes the following components:
i.   Data extraction.
ii.   Preprocessing.
iii.   Response Model Construction.
iv.   Performance Evaluation

A. DATA EXTRACTION

The data was collected from Reddit community website, which contains comments and reply. The dataset consists of 1.7 billion Reddit data, consisting of a parent comments and their corresponding reply. The data was collected by web scraping mechanisms and is converted to ".json" format for better integrity.
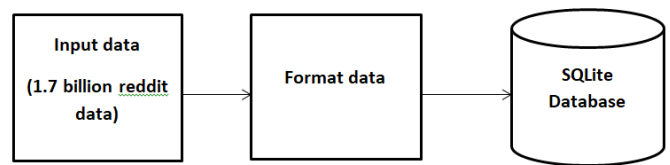

Figure 4.1 Storing data in SQLite database

B. PREPROCESSING

The data collected from Reddit community site were raw and unstructured. Such data would be difficult to handle, store and process. So, the data is first formatted by pairing up the child replies with their corresponding parent comment. Redundant data and null characters are also removed. The formatted data is then stored in the SQLite database which acts as the back-end data store of our model. Only the parent comment data, with more than one response is extracted and dumped in to SQLite database. This count is indicated by the score column. The higher the score, the greater that data can yield. The deep neural network framework is initialized by specifying weights and constraints. The parameters are tuned in such a way that the classification parameter is set in-tact and hidden layers are tuned accordingly. The variation and test parameters are set for initial running. Once the network is set, it is then interfaced with the inference model.
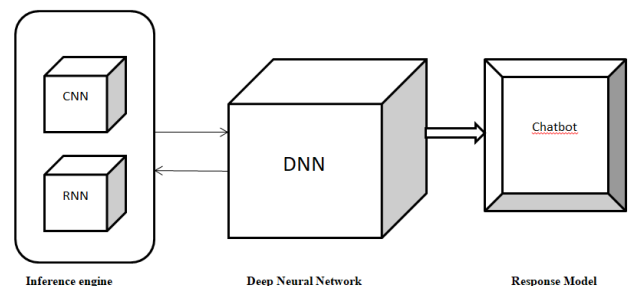
C. RESPONSE MODEL CONSTRUCTION

Figure 4.2 Response model generation by interfacing inference engine with Deep Neural Network

The inference model is hybrid machine learning model which consists of a combination of Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN). It is highly helpful for yielding more accurate results. It is interfaced with the tuned deep learning framework.

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, CNN have the ability to learn these filters/characteristics. The architecture of a CNN is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

Recurrent Neural Network (RNN) is a type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence. RNN have a "memory" which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks. RNN converts the independent activations into dependent activations by providing the same weights and biases to all the layers, thus reducing the complexity of increasing parameters and memorizing each previous outputs by giving each output as input to the next hidden layer. Hence these three layers can be joined together such that the weights and bias of all the hidden layers is the same, into a single recurrent layer.
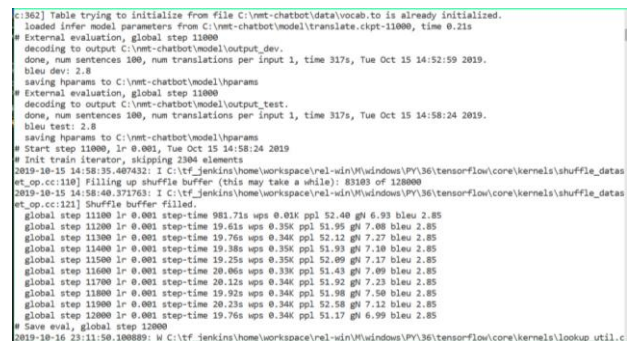
V. RESULT

The acquired unstructured data from Reddit must be formatted and stored in a database before processing the data and splitting it up. The code for formatting and preparing the data. The output of the will display the total number of rows formatted and dumped into the SQLite database as given in figure 5.1.

Figure 5.1 Preprocessed Data

The SQLite database is provided as the back-end support for storing and retrieving the data for formatting and to make handling of data easier. Later these data is fetched into the model via training process while each training step gets completed. The inference and the deep neural network parameters are tuned. The module is trained for 5000 epoch for each training global step buffering 1, 00,000 of data. The training step process is displayed in diagram 5.2.

Figure 5.2 Training steps

The output of the module is given in the figure 5.3 and figure 5.4.

```
_np_qint32 = np.dtype([("qint32", np.int32, 1)])
C:\ProgramData\Anaconda3\envs\tensorflow_env\lib\site-packages
 Passing (type, 1) or '1type' as a synonym of type is deprecat
(type, (1,)) / '(1,)type'.
  np_resource = np.dtype([("resource", np.ubyte, 1)])


Starting interactive mode (first response will take a while):

> Hi
- Hi
- Hi,
- Hi!
- Hi <unk>
- hi
- <unk>
- Hi.
- http:
- Hello
- Hello,

> I am good
- I'm not good at
- <unk>
- I'm sorry, I
- I'm good at <unk>
- I am good at how good
- <unk>
- <unk>
- <unk>
```

Figure 5.3 Test Output console

```
Hey I am emote! Ready to hear you out and help you anytime. Call me your buddy
me :)
Ready...
Your last command couldn't be heard
Ready...
You said: i wish i had someone to talk

Hey, you have me. Your buddy here, why fear? ;)
Ready...
You said: i had a great day

Great to see you happy!
Ready...
Your last command couldn't be heard
Ready...
You said: i am feeling sad

That makes me concerned, I wish I can help you out :)
Ready...
```

Figure 5.4 Final Output

VI. CONCLUSION

We utilize a depth convolution neural network for sentiment classification based on Reddit comments in this work. Our approach concatenates the pre-trained word embedding feature generated using inference word sentiment polarity features based sentiment lexicons and engrams features as the sentiment features vector of the Reddit comments and inputs the feature sets to a deep convolutional neural network. Our model captures contextual information with the recurrent structure and constructs the representation of the text using a convolution neural network. We finally conclude that deep convolution neural network utilizing pre-trained word vectors has good performance in the task of emotion classification.

The implemented chatbot will be further developed into a complete module by implementing Virtual Reality features consisting of facial and speech emotion recognition and sentiment analysis through interactive chat. Cross platform support will also be given during the course of development and cloud features will also be added. The end product will be a multi operating, cross platform device which can identify emotion through text, audio and images.

REFERENCES

[1] Zhao Jianqiang, Gui Xiaolin and Zhang Xuejun- School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, "Deep Convolution Neural Network on Twitter Sentiment Analysis", IEEE Conference on Development of personal Humanoids for better human understanding (2017)

[2] M. Shamin Hossain- Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, Ghulam Muhammad - Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, "An Emotion Recognition System for Mobile Applications", IEEE Special Section of emotion-aware mobile computing(2017)

[3] Jian Guo, Zhen Lei, Jun Wan, Eglis Avots – Dept. of Electrical and Electronic Engineering, Hasan Kalyponeu University, Turkey, "Dominant and Complementary Emotion Recognition from Still Images of Faces", IEEE Special Section on Visual Surveillance And Biometrics (2018)

[4] Mondher Bouazizi and Jomoaki Ohtsuki – Graduate School of Science and Technology, Keio University, Yokohoma, Japan, "Multi-Class Sentiment Analysis in Twitter", IEEE special Section on Emerging Trends, Issues and Challenges of Array signals (2018)

[5] Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu and Xu Wu – College of Information Engineering, Minzu University of China, Beijing, China, "Sentiment Analysis of Comment Texts Based on BiLSTM", IEEE Special Section on Artificial Intelligence and Cognitive Computing for Communication and Network (2019)

[6] Xin Kang, Fuji Ren, Yunong Wu, IEEE Members, "Exploring Latent Semantic Information for Textual Emotion Recognition in Blog Articles", IEEE Special conference on Automation (2018)

[7] Wanliang Tan, Xinyu Wang, Xinyu Xu – Department of Computer Engineering, Stanford University, USA, "SentimentAnalysis for Amazon Reviews", International Conference on Human and AI interaction (2018)

[8] Tzuu-Heseng S.Li, Ping-Huan Kuo, Ting-Nan Tsai and Po-Chien Luan- Department of Electrical Engineering, National Cheng Kung University, Pingtung, Taiwan, "CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot", IEEE conference on Human and AI interaction (2019)

[9] D. Yangm Abeer Alsadoon, P.W.C. Prasad, School of Computing and Mathematics, Charles Sturt University, Sydney, Australia, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment", International Conference on Smart Computing and Communications (2017)

[10] Carlos Busso, Zhigang Deng, Serdar Yildirim, Viterbi School of Engineering, University of Southern California, Los Angeles, Analysis of "Emotion Recognition using Facial Expressions, Speech and Multimodal Information", IEEE conference on AI For Tomorrow (2019)