A Hybrid Based Model on LSTM-CNN to Multi-class Emotion Analysis on Social Networking Dataset

by

Ibtehaz Ali Chowdhury
16101302
Eshad Dewan
16101226
TM Nafiz Mahmood Jishan
16101006
Samiul Kabir
16301041
Joyasish Mazumder
16101292

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Eshad Dewan

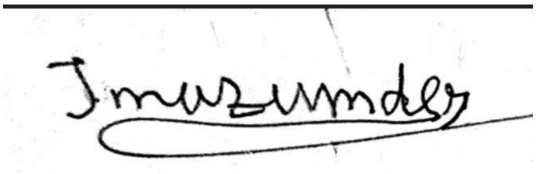
Ibtehaz Ali Chowdhury
16101302

Eshad Dewan
16101226

Samiul Kabir

Samiul Kabir
16301041

TM Nafiz Mahmood Jishan
16101006



Joyasish Mazumder
16101292

Approval

The thesis/project titled "A Hybrid Based Model on LSTM-CNN to Multi-class Emotion Analysis on Social Networking Datase" submitted by

1. Ibtehaz Ali Chowdhury (16101302)
2. Eshad Dewan (16101226)
3. TM Nafiz Mahmood Jishan (16101006)
4. Samiul Kabir (16301041)
5. Joyasish Mazumder (16101292)

Of Fall.2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 11, 2021.

Examining Committee:

Supervisor:

Dr. Jia Uddin
Assistant Professor(Research Track)
Department of Technology Studies
Endicott College, Woosong University
& Associate Professor(On leave)
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:

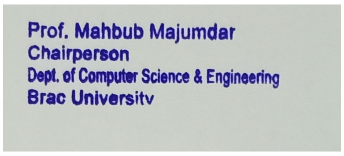
faisal

Faisal Bin Ashraf
Lecturer
Computer Science and Engineering Department
BRAC University

Thesis Coordinator:

Dr. Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)



Professor Mahbub Alam Majumder, PhD
Chairperson
Department of Computer Science and Engineering
BRAC University

Abstract

Sentiment analysis from texts has been a major research field in NLP. However, most of the studies are on binary (positive and negative) classification of the texts. While researching, we found that the accuracy of multi-class text classification according to emotions is very low when compared to binary classifications, as understanding and quantifying emotions is a very difficult task. We studied the two commonly used deep learning models used for text classifications: Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). We found that the greatest accuracy was achieved when the CNN model is used combined with a LSTM. In our paper, we proposed an LSTM-CNN hybrid model to classify texts according to five emotion classes and achieve an accuracy of 65%. We further studied Support Vector Machine (SVM) and Naive-Bayes classifiers. The experimental results show that the LSTM-CNN model had an improved accuracy.

Keywords: LSTM, CNN, NLP, Text classification.

Acknowledgement

Firstly, all praise to the Almighty for whom our thesis has reached this far.
Secondly, to our supervisor Dr. Jia Uddin for his kind support and advice in the progress. He associated us whenever possible.
Thirdly, to our co.supervisor Faisal Bin Ashraf who constantly guided us at every step.
Additionally, to the department of Computer Science and Engineering for giving us the opportunity to work on this thesis.
And finally to our family without their throughout support it may not be possible.
With their kind support and prayer, we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

CNN Convolutional Neural Network

$LSTM$ Long Short Term Memory

NLP Natural Language Process

SVM Support Vector Machine

Chapter 1

Introduction

1.1 Background

Artificial intelligence has contributed significantly in the field of natural language processing (NLP) since its birth in 1950. One of the most prominent research options in the field of NLP is Emotional analysis which refers to finding out the attitude or emotion of the writer by the use of natural language processing, text analysis and computational linguistics. Although emotional analysis started back in the 1950s, now-a-days it is widely used to mine information from the internet.Today, social media like face-book and twitter is used by almost all the people on the earth, where they share their opinions and their feelings. Twitter alone has almost 340 million users where 186 millions of them are active daily.The total number of tweets sent per day is over 500 million. People from different places of different ages, talk about their life events and how they feel about a specific issue or a topic.Since, the users of these sites are huge, the data they produce is enormous and the data is also more and more accessible. Therefore, this data can be used for finding out about the mental states of the individuals such as if they are suffering from anxiety or not, if they are happy or sad, if they are angry or excited et cetera. Due to the massive amounts of data on these social networking sites, emotional analysis has been a great research area in the last decade [8]. In spite of the fact that this analysis at present is really promising, most of the analysis is actually based on positive or negative emotions whereas sentiment can be classified into many categories like happiness, sadness, anger, joy, excitement et cetera.

At present, the analysis of emotion is mostly used in the commercial sectors. It helps the business grow by understanding their customer or getting knowledge about their sentiment for a certain product. People express their feelings and opinions more openly than ever on social networking sites. Thus, the brands can now get the feedback from people more easily now. Furthermore, they can also understand the customer requirements and tailor their product or business according to that information.However, the analysis of the emotion is mostly restricted into just positive, negative and neutral. It is like scratching on the surface where there is valuable information about people's emotions hiding in those social network data.

Nevertheless, with the advancement of deep learning, the algorithms ability has tremendously increased for analysing text from social sites.It helps researchers do in-depth research about emotional analysis .Everyday, huge amounts of data is created through the social network sites and most of them are unorganized. It is almost impossible to analyze the emotion manually in an effective and timely manner. There comes the emotional analysis using deep learning and natural language processing where the data can be processed and analyzed in a minimal amount of time and human inputs. Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) are such network models which can be used to analyze the emotion data and find the expected result. Convolutional neural network or CNN is a type of artificial neural network that uses a machine learning unit algorithm called perceptron for supervised learning and analyzing data. On the other hand, Long Short Term Memory or LSTM is a modified version of Recurrent Neural Networks (RNN). It helps resolve the vanishing gradient issue in RNN as remembering past data from memory is made easier by LSTM.

The applications of these multi-class classifications of emotion would be many if we can get to a higher accurate result. Primarily, sentiment analysis could be helpful for public health issues, safety, emergency response and urban planning [18]. Moreover, emotional analysis could be a great help in preventing suicide and measuring a community's well being [1]. On the other hand, It allows companies to evaluate the satisfaction levels and the expectations of their consumers [2].

1.2 Motivation

Mondher and Tomoaki [17], talked in their work about the importance of multi-class emotion classification. Their works describe that emotions such as happiness and love are both of positive polarity yet they mean completely different things. Therefore, it is not enough to classify sentiment in positive and negative.

We believe that multi-class classification of emotions from posts on the internet can provide large datasets for a lot of problems for this generation which will be further be discussed. Lots of techniques are being used to mine or classify emotions like parsing technique [14], or use R language to mine or analyse data [9] and many machine learning algorithms [13] are used however, binary classification has already achieved a very high accuracy where as for multi-class emotion classification it still has not fully developed yet and is still maturing and has room for further study and improvement which can provide a great incentive for researchers to work further in this field and not only that understanding human emotions from texts has become really popular and also in demand and in many different languages and in various fields [6], [7]. This Motivated us to look into the researchers of scholars and try to improve the accuracy and work in the field of NLP.

1.3 Problem Statement

The two major sectors of sentiment analysis are opinion mining and emotion mining [12]. Research on mining opinion from texts on websites like twitter, amazon etc has been a focus of the last two decades. There is much state-of-the-art work done on classifying opinions into 'positive' and 'negative'. The data collected from online social networking sites, at present, can easily be classified into positive and negative. However, the classification of emotions is inefficient if it is classified as binary. Humans show many emotions, a positive post can mean both happy and sad at the same time. For example the text - "I am glad to live long enough to see and play with my great grandchild, yet I cant stop these years." The text implies a positive meaning yet the speaker is clearly sad. There are thousands of posts like these that can be found over the internet which can be extracted and trained and mined for better understanding of human emotions which can further help us to solve many day to day problems or find things that attract people more and grab the marketplace more efficiently. For instance:

- Understanding the emotions of customers are very important for the business to flourish and grab the marketplace with preferences to the customers' choice. Customers often give lots of reviews for products, some are most of the time very straightforward, bad or good however in many cases the review gives a mixed feeling. Therefore knowing the response towards a product is crucial, as many brands can run survey to know the feelings of their customers towards their products and take appropriate actions.

- In recent times, E-Learning has become the main focus as a medium to study, as a result efficient systems are required to know and understand how the students are feeling, what they need and want from the faculty to provide. Many platform like coursera provides recorded videos where faculties are presenting live for which it has become a necessity to develop such systems.

- In many systems, AI systems are placed which keeps track of our activity and suggests us movies, music and places to visit with respect to our preferences and these might seem small and irrelevant however it makes our life a lot simpler and easy.

- Read diaries/notes or posts in social network sites to understand the emotional or mental state of a person which might help to avoid any disasters [1].

1.4 Goals and Solutions

In present times, the binary classification of emotions have advanced a lot, however as mentioned earlier emotion is not bound only to two emotions but rather there are many more emotions involved with humans which gives meaning to our life and day to day communication. As a result, advancement in classifying more than two emotions is very much important and has become a need with advancement of technology to make our life easier. So, our main focus or aim is to improve the accuracy of multi-class emotion, as the number of emotions increase the accuracy for the emotions falls down drastically. In our research we try to reach our goal by applying deep learning algorithms such as LSTM and CNN. Our model uses the deep learning algorithms to classify 40000 tweets from twitter into 5 emotions.

1.5 Research Contribution

Research about emotion analysis from social networking sites using deep learning is becoming more and more admired as people are being involved on social sites regularly now more than ever. Moreover, the applications of this are also enormous. At first, emotional analysis was only used in commercial sectors. However, researchers today realised what is underneath the surface of emotional analysis. This can be used for human welfare to societal development. Hence, we have taken the opportunity to research emotional analysis from social networking sites using deep learning and tried to do our best to come up with a system where the emotion we extracted from social media would not be binary but multi-class. In our research, we tried to figure out a way which will help us classify the social media data and find out the emotions people tried to express on their twitter posts. Using the resources that we had, we attempted to come up with a process where the accuracy will be much higher. Therefore, the emotions we get from the data will be much more accurate. In the following sections we have described briefly about the functions of our system. It showed the process of how we took a twitter data set and processed it so that we can classify the emotions from the data. This will be very crucial in further study regarding multi-class emotional analysis using deep learning.

1.6 Chapter Orientation

In the later part of the paper, the following chapters are presented:
Chapter 2 consists of Related Works, it consists of all the works that has been done previously on the classification of multiple emotions.
Chapter 3 comprises Methodology. This chapter starts with how the data has been processed and then moves on to explaining the different models used to classify the data to achieve the results.
Chapter 4 includes Results. This chapter gives an evaluation of the results of the different models and a comparison between the results received from different models.
Lastly, chapter 5 contains the conclusion and future work. It gives a summary of what we did and what we can achieve more with time in the future.

Chapter 2

Related Works

Ali, Ameneh, and Osmar (2017) [12] in their article have talked about the current state of sentiment analysis from text. There are two types of sentiment analysis: opinion mining and emotion mining. Most of the research since the 1900s has been done on opinion mining which is a ternary classification (positive, negative and neutral). In the paper, they have given data of the previous work that has been done on opinion mining - detection, classification and summarization. They have described that the methods can also be used for emotion mining with certain modification. The datasets used are of two types, annotated datasets and unannotated datasets. The use of convolutional neural networks provided the best result for the classification. Also some of the more common ways described are pattern-based and frequency-based, where each word is given a vector according to its index in a dictionary and the number of times they occur. Another approach was breaking the text in unigram (single word) and parts of speech-tagging - it was seen that adjectives gave us most knowledge about emotions. The work on unannotated data is much more difficult. One approach is to use lexicons. Lexicons are a dictionary of words that are given a value based on their polarity. The data are then used to train a machine. This provided good results. They described four emotion models proposed by the four theorists - Ekman (1972), Plutchik (1986), Shaver (1987) and Lovheim (2011). The Ekman model is the most widely used with its sets containing the emotions: anger, disgust, fear, joy, sadness and surprise. With surprise being the emotion that has both positive and negative polarity.

The paper presented by Mondher Bouazizi and Tomoaki Ohtsuki [17] has worked on to create a classification of data collected from twitter to sort them into seven different emotion classes: Fun, Anger, Happiness, Sadness, Love, Hate, Neutral. At first they classified them with two at a time using the opposites for example 'Love' and 'Hate', this resulted in an accuracy higher than 80%. They gradually increased the number of classes and when they used seven different emotions the accuracy dropped to 60%. They used Random Forest Classifier and 4 Key performance Indicators to make the classification. After they obtained the results they verified and proposed some models to help minimise the errors. They presented an innovative way to represent the words into an emotion space. They described the emotions as clouds with different depths. Words that implied emotion were deeper into the cloud and words with less intensity were in lower depths. In certain cases like the emotion fun and happy have many common words so their clouds intersect and so the words can be represented in an intersecting region. The correlation of the words were found by the vector difference between the words from the centre of the emotion cloud. In conclusion they stated that it is important to classify the texts based on emotions, it will be even more helpful if the words extracted be given separate value for different emotions as they can be describing more than one emotion.

Mansur, Okan and Adil in their paper [19], approached the issue of analysis of emotions from tweets from twitter from the Turkish citizen by using deep learning techniques CNN , RNN with LSTM, and also used ANN however the result was very poor in comparison to the predecessors. For the purpose of preprocessing, initially they removed the 'http' which is often seen to be in use in tweets and serves no purpose for understanding human emotions, later the eliminated all the punctuations numbers and extra spaces and used stemmer approaches (F5 and Snowball Stemmer) to preprocess the datasets and lastly used NLTK to remove stop-words. For the analysis of emotion they used three deep learning algorithms. In ANN architecture there are three layers input layer, hidden layer and lastly output layer where the layers consisted of 1000 neurons , 128 neurons and 6 neurons accordingly. Then for the architecture of convolutional neural network, from the framework of KERAS they produced an embedding layer consisting of 100 input sequence length and a vector space consisting of 8 dimensions followed by 32 filters of convolutional layer with 8 as the kernel size later they max pooled and brought down the convolutional layer to half of what it was originally. The output 2D max pooled layer is then mapped to one dimension so that it can be fed into the fully connected sequential dense neural network. Whereas for the last variant of the three techniques used by the author, they used three different layers where initially it was embedded later used RNN algorithms with its variation LSTM and lastly used fully connected sequential dense neural network. For all three techniques they used a drop rate of 20% with Relu and Softmax as the activation function. And for the optimizer they utilized Adamax and Categorical CrossEntropy functions. For most cases they realized best results using the CNN Architecture where as for LSTM in few cases they outmatched CNN however for ANN it received poor results for all the emotions other than 'fear'.

In the paper, Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu1 categorizes the text-data into emotion classes, according to the emotion the messages represent by developing a machine learning system which could be supervised offline [18]. In this paper, they illustrated their approach which includes collecting a huge chunk of preprocessed data for training the classification models of emotion and then utilized those models for classifying the tweets. The messages in the dataset were labeled with their respective emotions and the classified tweets were being classified simultaneously from the live-stream messages. They have generated a system called Emotex for classifying purposes and for organizing the data by allocating the messages to their particular classes. Since they needed a chunk of emotion-labeled messages, they have used the hashtags from the text for annotating the data. People often use hashtags on twitter for sharing their feelings. Therefore, it is easier for them to use it for automatic annotation. After creating the labeled text-data, the authors trained the classifier model from those data. The authors used Naïve Bayes and SVM classifiers. At first, they identify the tweets which are not part of any class of emotion. A binary classifier is used for doing this job. Finally, they perform the main classification of emotion by implementing their system and classify the tweets explicitly.

In the paper, Li and Xu approached the classification of emotions not by traditional methods of just creating binary classification but used to classify emotions by training them based on the extracted emotion cause events [4]. Li and Xu initially gained ideas about how the emotions of humans work by studying the fields of Sociology and later analyzed the comments or posts of many users on Weibo to identify the relationships between the emotions and the cause events. They developed an emotion cause extraction subsystem to identify and extract the cause events and mined deep-level emotional information and used these cause events for classifying the emotions which is not a very common method of emotion classifying technique. For a human to experience any kind of emotion there is usually a cause behind it and this cause can give us a measure of how strong a certain emotion is and the reason for which the emotion occurred is known as emotion cause event which the authors LI and Xu tried to identify as it would gave a better view or understanding of the emotions however the posts in Weibo had some unnecessary things in it like geotags and web links which had to be removed, therefore after these unwanted links or tags been removed, ICTCLAS tool-kit was used to parse, segment and tag the post with respective parts of speech (POS) tags. At first Li and Xu tested using a different dataset of around 1000 random emotion based inputs for development purposes to realize the extent to which it would be helpful to identify the amount of emotion cause events however the posts in Weibo are very informal and had some casual words in it which made it difficult for them to run algorithms as a result they themselves looked at the data and replaced a lot of words which had less relation to cause events to new one. They used a list of emotion words which had a total of 1845 words and short phrases that were being used for expressing emotions or opinions. It also consists of synonyms. From these they generalized a group of semantic patterns to find out emotion causes of internet users which is (Noun, Verb, and Noun). The system that the authors developed finds out the structure of (Noun, Verb, and Noun) as the root frame however since the users use many casual words, therefore even if any one of the three is recognized it has been taken as a cause event. Later they removed all the other words and only took this cause event as a substring since the rest of the words did not have any contribution to emotions or feelings for instance, " "The Voice" surprised me. This song really made me excited and unable to sleep. " Here the cause events are "The Voice" and "this song really made" whereas the emotions are "surprised" and "excited". The authors faced an issue which was, there were some posts where no cause event was used but still reflected emotions therefore to tackle these kinds of issues they resorted to another approach of feature selection for these data which is Chi-Squared test. The two types of features realized through Chi-Squared test and emotion cause extraction are merged together into a single feature set. Li and Xu used Support Vector Regression (SVR) for classifying the emotions which is a version of Support Vector Machine (SVM). They chose SVR because the data taken are short and informal and not balanced and SVR works better for imbalanced data sets. The data set that the authors used are more than 20000 posts which are randomly taken from Weibo but used only 16485 posts since there were a lot of duplicate posts. All the data were labelled into a corpus with emotion type and cause of emotion where present. They used 75% for training sets and 25% of the data for testing. Their system had improved precision with similar recall however their disgust class faced decreased precision without cause event whereas with cause event it increased a lot. By evaluating F1-score, the precision and recall of their system was better. The f-score increased by 1.6% for happiness, 2.0% for anger, 9.2 or disgust and 0.8% for surprise however for sadness it was a little less they assumed that it was probably because the users are not very open about their sadness fully

This paper by Jonathan Herzig, Michal Shmueli-Scheuer and David Konopnicki, improves the accuracy of emotion detection of manually annotated datasets which have high accuracy and small sized datasets quantity using five datasets namely; ISEAR, Semval, Fairy Tales, Blog posts, Twitter dialogues [11]. They have included a table (Table 1) detailing the characteristics of these datasets. Since generation of high density vector representation of word embedding was not possible they used pre-trained representations; word2vec and Glovec using ensemble methods they improved the accuracy of detecting emotions in the datasets.They employed one vs rest binary machine learning classifiers for each emotion. Depending on the datasets, they can be multi class or multi labeled problems when representing a classification of the datasets. Hence, they used one vs rest SVM classifier for emotions. First they used an SVM classifier with a linear kernel which represented every document as a BOW, After lemmatization they extracted various features and used TFIDF weights for these features. These scores are compared with previous results and displayed in a BOW baseline performance table(Table 2). Using word embedded based vectors they experimented with CBOW, TFIDF weights and classifier weights according to their combined word vector notations. Here they used SVM classifier similar to BOW classifier but used RBF kernel which is non-linear, hence this classifier provided better results than linear kernel. This classifier's results are then compared with the datasets previous results and included in a table (Table 3). They used an ensemble approach which combines both a linear model based on BOW, and a non-linear model based on the pre-trained word vectors from word2vec and Glovec, and transformed the classifier's results into probabilities using their ensemble method notation. They reported their results using precision based weighting method which they found to be the best performance (Table 4). Using the BOW, Word embedding based classifiers and implementing the ensemble method, their results outperformed the previous results of every document representation and the best result, which is also significantly better for each dataset, is of EN-CLASS model that achieved an average relative improvement of 11:6% in F1-score over all datasets.

We then expanded our research to study deep learning techniques for text classification. Initially we studied the best models for text classification. Weighted-RNN model produced the highest accuracy compared to other deep learning models [20]. Multi-tasking Recurrent Neural Network model presented in the study by Liu, Qiu and Huang [2016] [10]. The proposed three models (uniform-layered architecture, coupled-layered architecture and shared-layered architecture) to classify text into d-dimensional vectors.

Another study showed that convolutional neural networks also presented high accuracy when applied to text classification by sentiments [3]. Two channel CNN can be used to extract patterns in the text and classify them with the softmax function. The sentences are first converted into vectors using an embedded layer and then fed to the CNN as an input. This model produced very decent results when compared to the machine learning models discussed above. Finally we studied LSTM models which are considered the best when used to classify texts. Word2Vec combined with LSTM model provided a theoretical accuracy of 93% whereas Word2Vec combined with CNN model gave an accuracy of 81% [15].

Chapter 3

Methodology

3.1 Data Processing

The dataset contains 40000 tweets from the social media site Twitter. During the pre-processing, the unnecessary words, i.e “a”, “the” etc. known as stop words are removed. We used the list of stop words in the nltk library for python. We further replaced the hashtags commonly used in Twitter for tagging by removing the “#” from the texts.

The text has been categorized into 13 emotions: ”empty”, ”sadness”, ”enthusiasm”, ”neutral”, ”worry”, ”surprise”, ”love”, ”fun”, ”hate”, ”happiness”, ”boredom”, ”relief”, ”anger”. The emotions like enthusiasm and fun are related to happiness so we decide to pre- process the 13 categories into 5 different categories, so that the emotions are more distinguishable: “Neutral” = “0”, “Happy” = “1”, “Sad” = “2”, “Hate” = “3” and “Anger” = “4”. These emotions were chosen because they are the most common replies on a social media site.

1956967341	empty	xoshayzers	@tiffanylue i know i was listenin to bad habit earlier and i started freakin at his part =[
1956967666	sadness	wannamama	Layin n bed with a headache ughhhh...waitin on your call...
1956967696	sadness	coolfunky	Funeral ceremony...gloomy friday...
1956967789	enthusiasm	czareaquino	wants to hang out with friends SOON!
1956968416	neutral	xkilljoyx	@dannycastillo We want to trade with someone who has Houston tickets, but no one will.
1956968477	worry	xxxPEACHESxxx	Re-pinging @ghostridah14: why didn't you go to prom? BC my bf didn't like my friends
1956968487	sadness	ShansBee	I should be sleep, but im not! thinking about an old friend who I want. but he's married now. damn, & he wants me 2! scandalous!
1956968636	worry	mcsleazy	Hmmm. http://www.djhero.com/ is down
1956969035	sadness	nic0lepaula	@charviray Charlene my love. I miss you
1956969172	sadness	Ingenue_Em	@kelcouch I'm sorry at least it's Friday?
1956969456	neutral	feinyheiny	cant fall asleep
1956969531	worry	dudeitsmanda	Choked on her retainers
1956970047	sadness	Danied32	Ugh! I have to beat this stupid song to get to the next rude!
1956970424	sadness	Samm_xo	@BrodyJenner if u watch the hills in london u will realise what tourture it is because were weeks and weeks late i just watch itonlinelol
1956970860	surprise	okiepeanut93	Got the news
1956971077	sadness	Sim_34	The storm is here and the electricity is gone
1956971170	love	poppygallico	@annarosekerr agreed
1956971206	sadness	brokenangel1982	So sleepy again and it's not even that late. I fail once again.
1956971473	worry	LCJ82	@PerezHilton lady gaga tweeted about not being impressed by her video leaking just so you know

Figure 3.1: Raw Dataset collected from twitter

i know i was listenin to bad habit earlier and i started freakin at his part	0
layin n bed with a headache ughhhh waitin on your call	2
funeral ceremony gloomy friday	2
wants to hang out with friends soon	1
we want to trade with someone who has houston tickets but no one will	0
re pinging why didn t you go to prom bc my bf didn t like my friends	2
i should be sleep but im not thinking about an old friend who i want but he s married now dam	2
hmmm is down	0
charlene my love i miss you	2
i m sorry at least it s friday	2
cant fall asleep	0
choked on her retainers	2
ugh i have to beat this stupid song to get to the next rude	2
if u watch the hills in london u will realise what tourture it is because were weeks and weeks la	2
got the news	2
the storm is here and the electricity is gone	2
agreed	1
so sleepy again and it s not even that late i fail once again	2
lady gaga tweeted about not being impressed by her video leaking just so you know	2
how are you convinced that i have always wanted you what signals did i give off damn i think i	2

Figure 3.2: Pre Processed data set collected from Twitter

Fig 3.1 and Fig 3.2 shows an example of the pre-processing which we applied. For the sentence, the emoticons and the twitter tag was removed, and the emotion was changed to 0 which has the value as neutral. For sentences like "got the news" it was difficult to process it into one class since the emotion surprise can be both "happy", "sad" or "anger", so we decided to classify surprise as sad because most of the sentences in the "surprise" group were related to sadness.

3.2 Word Embedding

In all natural language processing tasks, the variable length words are first needed to be converted into fixed length vectors. The neural models consist of a projection layer that maps the words in the text, known as n-grams, into vectors based on their meaning. For our project, we have used the GloVe model, developed by the CSE Department of Stanford [5]. This unsupervised model has been the most useful to us because it was pre trained over 2 billion tweets from twitter. This allowed us to minimize our pre-processing of many words that is widely used on social media sites that are a bit different from actual english vocabulary.

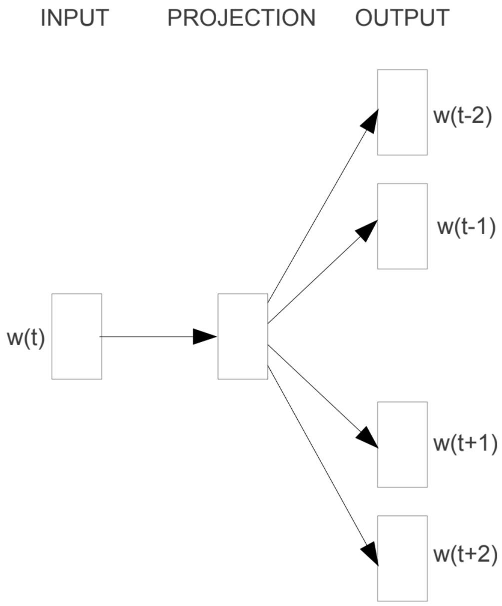


Figure 3.3: Skip gram

3.3 Models Used

3.3.1 LSTM

Recurrent Neural Network (RNN), is widely used for natural language processing sequences. The problem with long sequences arises from the vanishing gradient problem which results in information loss in RNN. Long Short-Term Memory(LSTM) is now used to counter the problems with RNN. The LSTM neural network solves the information problem by adding a memory cell. The memory cell is used to store the previous outputs in the layer, therefore like an actual human brain, the LSTM can use previous information to make a future prediction.

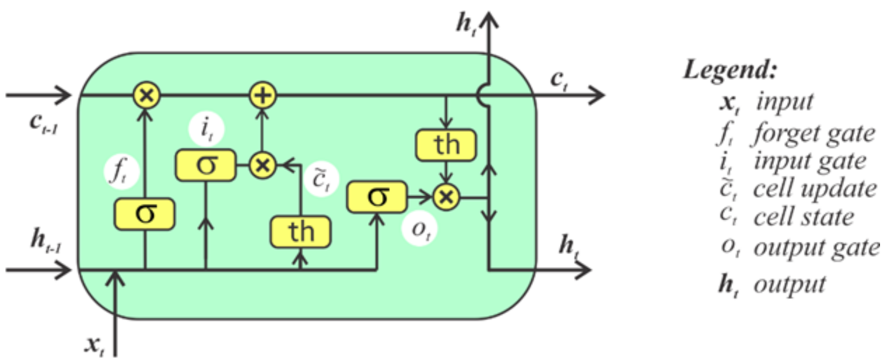


Figure 3.4: LSTM Structure

The history information in the memory cell is updated by using the three gates shown in the figure 3.4. The input gate determines how the incoming vectors change the memory cell, the output manages what goes on to the next layer from the memory cell as an output, the forget cell keeps in check whether the memory cell is deleted or kept as it is.

Let the input be X= [x1, x2, x3, . . . , xt], ‘xt’ is a d-length word and the amount hidden layer be H, the output of the three gates are as follows:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{3.1}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{3.2}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{3.3}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{3.4}$$

$$h_t = o_t \otimes tanh\left(c_t\right) \tag{3.5}$$

The state of the memory state at time step is $'c_t'$ and $'h_t'$ is the final output of the LSTM, with a dimension equal to the hidden dimension layer H of the LSTM. LSTMs are very useful for detecting emotions in written language. The LSTM will remember what it has previously read, and store it in memory to give a better reading for sentences with changing sentiment. For example, “The mother was happy her son was travelling abroad for studies, but inside she was dying of loneliness” sentences like this have very little accuracy when classified with traditional classifications like SVM.

3.3.2 CNN

Convolutional Neural Network (CNN) is the most popular for image processing but they are widely used for many tasks. CNNs are designed to detect local features or patterns in a multidimensional array. For example, CNN can easily detect between different animals in photos. In CNN, the input vector data is passed through a convolutional layer made up of different filters. The output is then pooled and passed onto a connected layer.

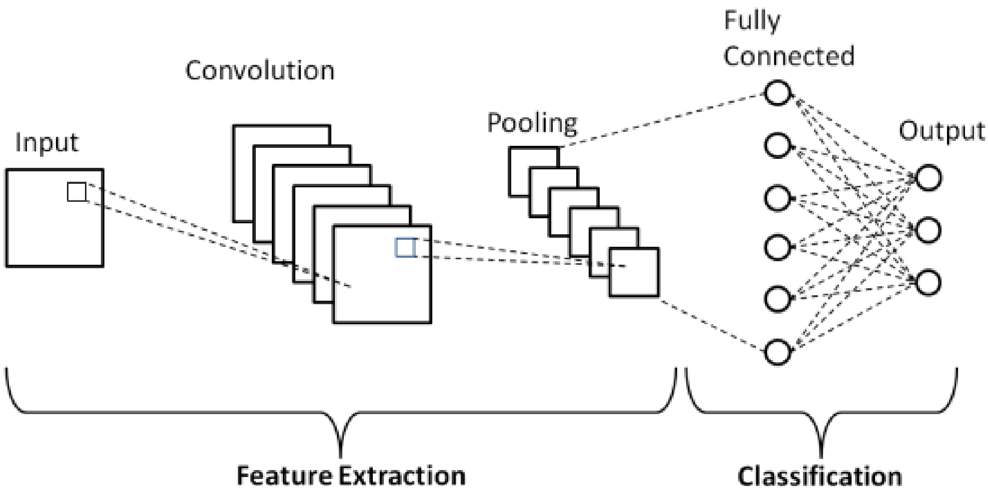


Figure 3.5: Convolutional Neural Network Structure

Here, Fig 3.5 shows the structure of a CNN. In CNN, the input layer takes an n-dimensional array, and sends it to the convolution layer for feature extraction. The convolution layer extracts various different patterns, which are accumulated in the pooling layer. From the pooling layer it is passed into the fully connected layer, where the classification occurs according to probability calculated by the soft-max function.

The input X= [x1, x2, x3, . . . , xt] where xt is a d-length word, the n - length sentence is represented by x1:n=x1 x2 ... xn. A convolutional filter, $w\epsilon \mathbb{R}^{hk}$ is used on the 'h' number of words to extract a new feature.

$$c_i = f\left(w.x_{i:i+h-1} + b\right) \tag{3.6}$$

Where $b \in \mathbb{R}$ is a bias term and f is non-linear. The filter is then applied in all possible words till to produce a feature map $c = c_1, c_2, ..., c_{n-h+1}$. Max over-time-pooling operation is used to generate the max value c =max{c} as the feature corresponding to this particular filter. This process for one filter, various filters are used to extract all the features which are passed to the softmax layer that calculates the probability.

The idea of using CNN is to find patterns in sentences like the use of phrases. For example: "I am feeling down" classifies as a sad sentiment whereas "He is a down to earth person" has a positive meaning. These patterns are best classified with CNNs, as LSTM has a low accuracy during the classification of these forms of texts using the feature map.

3.3.3 LSTM-CNN Hybrid model

In this paper, we are using a hybrid model combining both lstm-cnn neural networks for categorizing the emotions from basic text. Both the LSTM and CNN models have their drawbacks, when used for classification of emotions from text. Our research allowed us to study a model which takes the best from both the models. We used the model proposed for text classification by Jiarui Zhang [2018] [16] and classify our data into five different emotion classes.

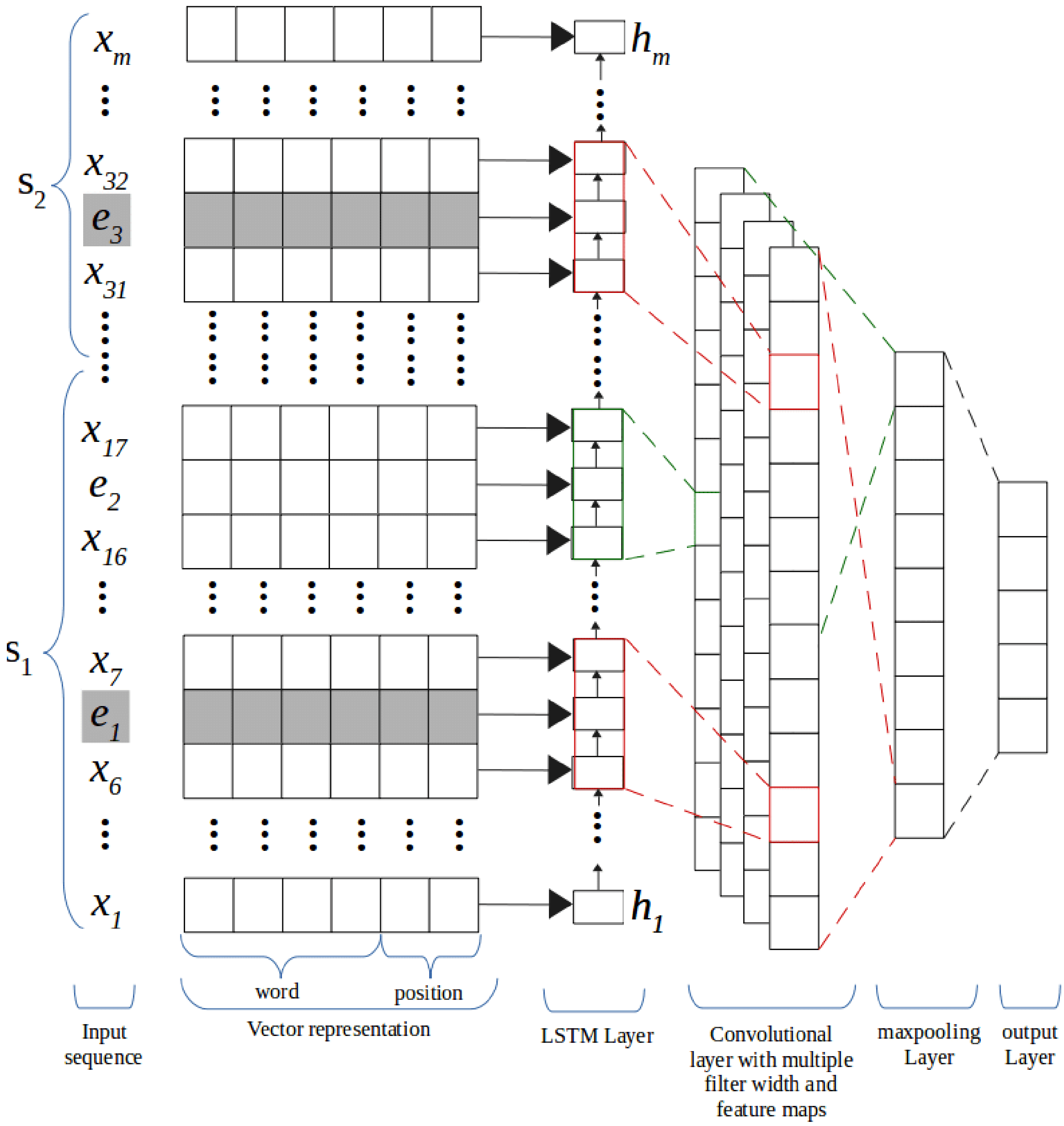


Figure 3.6: Long-Short Term and Convolutional Neural Network Hybrid Model

The tweets are passed into the LSTM model as input vectors after vector embedding. The model classifies the texts based on previous equations, and passes the output $h = [h_1, h_2 ... h_t]^T$ in the convolutional layer as the length of this vector is equal to the number of hidden layers of the LSTM.

The convolutional layer then applies the filter $w \epsilon \mathbb{R}^{jk}$ to extract the feature c, where 'j' is the number of words and 'k' is the dimension of the word embedding vector. A feature map is then generated as $c = [c_1, c_2 ... c_m]$ by applying the filter to all possible inputs. An output is generated at time stem 't' as

$$O_{w_t} = ReLU\left[\left(\sum_{i=0}^{m} h_{t+1}^{T} w_i\right) + b\right] \tag{3.7}$$

We used the ReLU activation function and as explained in equation 3.7 'w' and 'b' are parameters of the filter function.

The max-pooling layer is generated after all the possible features are extracted.

The most important feature is then selected according c =max{c}. The process is repeated by using multiple convolutional filters to extract multiple different features. The output is then passed on to the final fully connected layer where it is labelled according to the probability that is calculated using the soft max function:

$$P\left(y^{(i)} = j|x^{(i)};\theta\right) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{k=0}^{K} e^{\theta_j^T x^{(i)}}} \tag{3.8}$$

Chapter 4

Results and Analysis

After Embedding words in Glove, we trained the data with LSTM-CNN hybrid algorithm, which has their own specific roles to play in analyzing emotion from text as mentioned in chapter 3. Finally, after training our LSTM-CNN hybrid model, we measured our model's accuracy by cross referencing the results with our dataset. We compared our LSTM-CNN hybrid model with a SVM and a Naive Bayes model in 4.4. The results of our model's accuracy are shown and described in 4.1 and 4.2.

We have used the 4-key performance indicator to evaluate the effectiveness of the models after classification : Precision, Recall, F1-score and Support.

- **Precision** is the ratio of the tweets correctly classified to the total number of tweets classified belonging to that particular emotion class.
- **Recall** is the ratio of the tweets that were correctly classified to the total number tweets that actually belonged to that emotion class.
- **F1-score** is the harmonic mean of both the precision and recall.

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.1}$$

- **Support** is the number true response that lie in the class.

4.1 Model Parameters

The parameters set for this model are based on the implementations of CNN, LSTM on our dataset, after several tweaking, tuning and testing the model, we decided the parameters given below works best for the given model and dataset.

Table 4.1: Model parameters

Model Parameters	
epoch	200
batch size	128
pool size	4
filters	24
dropout	0.3

4.2 Experimental Results

Table 4.2: 4 Key Performance Indicators for LSTM-CNN hybrid

	Precision	**Recall**	**f1-score**	**support**
neutral	0.53	0.39	0.45	1938
happy	0.66	0.74	0.70	3257
sad	0.65	0.70	0.67	3212
hate	0.84	0.72	0.78	840
anger	0.93	0.73	0.82	210
avg/total	0.65	0.66	0.65	9457

From training our LSTM-CNN hybrid model, we ran tests on each emotion against a number of texts from the dataset and cross validated it to the results in our datasets. According to our results in table 4.2, we have an average accuracy of 65% and have acquired more than 50% of precision on each emotion. We have the lowest F1-score of 43% in neutral, and highest score of 82% in anger, in average of all emotions, we got 65% F1-score.

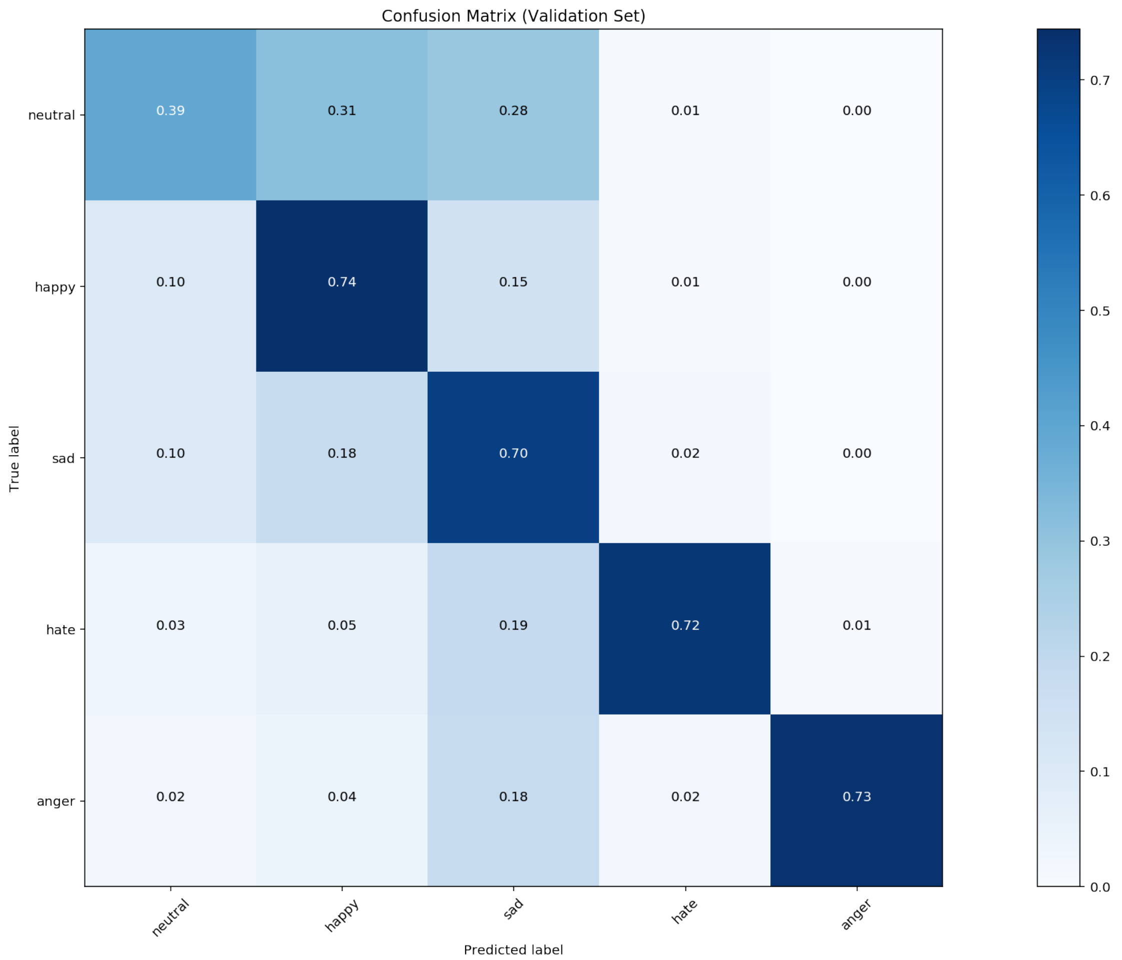


Figure 4.1: Confusion Matrix

The confusion matrix for each individual emotion is cross validated with each of the five emotions, and resulted in their precision and accuracy with 0 being lowest and 1 being highest, the graph above is illustrated with ranges of color from light blue to deep blue respectively. In the Confusion Matrix of our model, Fig 4.1 we achieved good results on happy, sad, hate and anger. However, performed poorly in neutral compared to previous papers, hate and anger were sometimes misclassified as sadness, and happiness could not get as high, as we combined love, fun and happiness.

4.3 Result Analysis

Based on our results, we came to three conclusions; One is the use of short words like “gr8” for great, “rn” for right now, etc. Two is the ability of this model to detect sarcasm, our machine did not learn sarcasm, hence miss-classifies texts which are sarcastic. Three is the reason for sentences carrying multiple emotions, sentences can carry multiple emotions, but here, this LSTM-CNN hybrid model favors the highest number of words corresponding to a single emotion in a sentence.

4.4 Comparing Models

Here, we show and explain about the results of our emotion classification analysis, and comparing them to different algorithms apart from our LSTM-CNN hybrid system.

Table 4.3: 4 Key Performance Indicators for Naive Bayes

	Precision	**Recall**	**f1-score**	**support**
neutral	0.44	0.46	0.45	500
happy	0.07	0.38	0.12	24
sad	0.34	0.50	0.41	262
hate	0.59	0.50	0.54	1021
anger	0.56	0.48	0.52	611
avg/total	0.52	0.48	0.50	2418

Table 4.4: 4 Key Performance Indicators for SVM

	Precision	**Recall**	**f1-score**	**support**
neutral	0.31	0.51	0.38	315
happy	0.17	0.46	0.24	48
sad	0.41	0.58	0.48	269
hate	0.81	0.49	0.61	1424
anger	0.41	0.59	0.48	362
avg/total	0.63	0.52	0.54	2418

We tried Support Vector Machine, and Naive Bayes models to compare our models. Both of these models are used in machine learning mostly for binary classifications. Since, our goal was a multi-class approach both the models presented a low accuracy as shown in the tables above. Using Naive Bayes gave us an average F1-score score of 50% compared to the model we used. Where as, using SVM gave us an average F1-score of 54% compared to the model we used. Both the models Naive Bayes and SVM performed very well for “hate” and “anger” and reached an F1-score of 54% and 61% respectively for Naive Bayes and achieved a F1-score of 52% and 48% for anger respectively. This is mainly because most of the tweets are negative in the dataset. The models behaved very poorly for the “happy” as the word embedding method is not optimum for the models. We believe it was confused between neutral, and happy comments. This problem could be solved if we took a pattern-based approach, i.e parts-of-speech tagging.

Between the models we observed, SVM outperformed Naive-Bayes in all the classes. Naive-Bayes treats each feature extracted as independent, whereas SVM looks at the interaction between the features to some extent. Thus for text classification, we found SVM to be superior to Naive-Bayes.

4.5 Discussion

We used deep learning for classification with the softmax function, that is why the accuracy is much higher than the other models. The models of the Naive Bayes, and the SVM will acquire a better accuracy if our models were pattern based.

In the paper published by Mondher Bouazizi and Tomoaki Ohtsuki [2019], they used a pattern-based approach to classify texts into different classes using Random Forest Classifier. Their model reached an accuracy of 61.8% when they classified the emotions into 5 different classes.

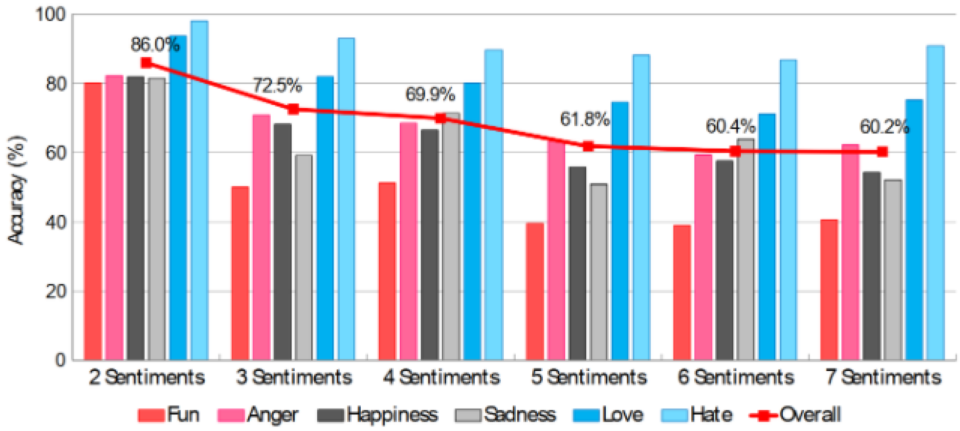


Figure 4.2: Overall Classification Accuracy

The LSTM-CNN model reached an accuracy of around 65%. Using a deep learning model provided a better accuracy, while compared to a pattern-based model.

Chapter 5

Conclusion and Future Works

5.1 Conclusion

Our research on deep learning approach to Multi-class emotion analysis on social networking sites will help to give a broader perspective on emotions involving humans giving a meaning to our life rather than binary emotions only since there are many more emotions in our day to day communication. For our research, we used a hybrid model combining both lstm-cnn for categorizing the emotions from basic text.

During Data Preprocessing we collected a data set containing 40000 tweets from Twitter. For our Word Embedding model, we have used the GloVe mode, developed by the CSE department of Stanford[5]. This unsupervised model has been the most useful to us because it was pre trained over 2 billion tweets from twitter.

During our research, we realised that both the LSTM and CNN models have their drawbacks when used for classification of emotions from text. We studied a model proposed for text classification by Jiarui Zhang which takes the best from both the models and tried to apply it for our project. The model uses the features extracted by LSTM by retaining the historical information and the local features extracted by CNN. We hope that our research work will have a great effectiveness in emotion mining. The perspective of having multi class emotion rather than binary will make it more realistic.

5.2 Future Work

Our work is only based on 5 emotion classes. However, human emotions are more than just 5 emotions. We have seen the accuracy to drop significantly with the addition of more classes. We faced two major problems. The first challenge was detecting sarcasm. Our model could not classify between sarcasm. Sadly, even with further research we could not fix the issues caused by sarcasm detection. The second problem was tweets with two emotions together. This problem could be solved by using a multi-labelled tag like sad and angry together. We believe our model could be further improved if we can combine the pattern-based machine learning approach and this model. In future, we hope to achieve better results for the problems we discussed.

Reference

[1] B. Desmet and V. Hoste, "Emotion detection in suicide notes," *Expert Systems with Applications*, vol. 40, no. 16, pp. 6351–6358, 2013.

[2] M. A. Cabanlit and K. J. Espinosa, "Optimizing n-gram based text feature selection in sentiment analysis for commercial products in twitter through polarity lexicons," in *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*, IEEE, 2014, pp. 94–97.

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

[4] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1742–1749, 2014.

[5] 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.

[6] B. Sun and V. T. Ng, "Analyzing sentimental influence of posts on social networks," in *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, IEEE, 2014, pp. 546–551.

[7] M. Al-Ayyoub, S. B. Essa, and I. Alsmadi, "Lexicon-based sentiment analysis of arabic tweets," *International Journal of Social Network Mining*, vol. 2, no. 2, pp. 101–114, 2015.

[8] A. Giachanou and F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," *ACM Computing Surveys (CSUR)*, vol. 49, no. 2, pp. 1–41, 2016.

[9] S. Kumar, P. Singh, and S. Rani, "Sentimental analysis of social media using r language and hadoop: Rhadoop," in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, IEEE, 2016, pp. 207–213.

[10] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," *arXiv preprint arXiv:1605.05101*, 2016.

[11] J. Herzig, M. Shmueli-Scheuer, and D. Konopnicki, "Emotion detection from text via ensemble classification using word embeddings," in *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, 2017, pp. 269–272.

[12] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, pp. 1–33, 2017.

[13] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment analysis of tweets using machine learning approach," in *2018 Eleventh international conference on contemporary computing (IC3)*, IEEE, 2018, pp. 1–3.

[14] C. Sowmya, S. Deena, and S. Anbuchelian, "Performance of sentimental analysis by studying and mining social media using parsing technique," in *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*, IEEE, 2018, pp. 294–299.

[15] L. Xiao, G. Wang, and Y. Zuo, "Research on patent text classification based on word2vec and lstm," in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, IEEE, vol. 1, 2018, pp. 71–74.

[16] J. Zhang, Y. Li, J. Tian, and T. Li, "Lstm-cnn hybrid model for text classification," in *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, IEEE, 2018, pp. 1675–1680.

[17] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on twitter: Classification performance and challenges," *Big Data Mining and Analytics*, vol. 2, no. 3, pp. 181–194, 2019.

[18] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing twitter data," *International Journal of Data Science and Analytics*, vol. 7, no. 1, pp. 35–51, 2019.

[19] M. A. Tocoglu, O. Ozturkmenoglu, and A. A. Alpkocak, "Emotion analysis from turkish tweets using deep neural networks," *IEEE Access*, vol. 7, pp. 183 061–183 069, 2019.

[20] D. Wang, J. Gong, and Y. Song, "W-rnn: News text classification based on a weighted rnn," *arXiv preprint arXiv:1909.13077*, 2019.