

Dominant Lexicon Based Bi-LSTM for Emotion Prediction on a Text

A.Shunmuga Sundari, R. Shenbagavalli

Abstract: *User-generated content and opinionative data has become a massive source of information on World Wide Web in the past few decades. Through social media people can share more conveniently their opinions, views, feelings and attitude about a product, person or event at anytime and anywhere as daily basis. This ever-growing subjective data makes enormous amount of unstructured data in web. Analyzing emotion in this raw unstructured data gives a very fruitful information for any kind of decision making process taken by both government and industries. Sentiment or emotion analysis is a field of Natural Language Processing (NLP), is used to identify the emotion depicted (by) in the form of text. Computation of emotion and emotion intensity depicted by a text is a very difficult task. Feature extraction from the text for vector representation is a difficult step of emotion analysis because it defines the emotion accuracy of the prediction. In this paper, a selective lexicon based BI-LSTM technique has been proposed. This technique uses only the most affected lexicon and its features for final vector representation. This method is a combination of features collected from the convolutional Neural Network (CNN), Long Short Term Memory (Conv - LSTM) and Bidirectional Long Short Term Memory (BI-LSTM). As a result the proposed model Selective Lexicon Based BI-LSTM (SL + BI-LSTM) outperforms all the models with high accuracy.*

Keywords - *Sentiment analysis, Natural Language Processing, CNN, LSTM, BI-LSTM*

I. INTRODUCTION

The Expeditious development of social media services has facilitated the communication of opinions through news portals, forum discussions, reviews, messages, blogs, and microblogs/tweets. Nowadays Microblogging has become a very vogueish communication tool among internet users. Millions of users share opinions on a different aspect of life every day. Therefore these microblogging websites are rich sources of data for opinion mining and sentiment analysis. Sentiment computing is the most salient branch of Natural Language Processing. It deals with the text classification in order to determine the intention and emotion of the author of the text. Text in social media is short, fast evolving and informal which presents challenges to sentiment analysis.

Emotions sculpt a very indispensable and elementary aspect of people existence. Whatever people do and say,

somehow does reflect some of their emotions, though may not be directly. The emotional analysis organizes a radical part of the field called affective computing. Emotion stands for “Affect” and measure or calculate stands for “computing” implies the term “Affect Computing”. Affective computing is all that takes us to design the devices or systems that process, recognize, interpret and simulate the human affects, thus making it possible for us to analyze the human and machine interactions.

Accuracy of predicting emotion from the text based on affective text faces issues based on features have been selected for evaluation. In our proposed work we use the technique called “selective lexica” that is it selects most affected word from a text among all others and then extract the features for that affective word and build the final vector representation. This final vector is given as an input to neural network models CNN [24], LSTM [25], BI-LSTM and Selective Lexicon based BI-LSTM to predict the emotion. This paper analyzes four emotions such as joy, fear, anger and sad. As a result the proposed model comes up with a high average than all other models in calculating the emotion.

II. RELATED WORK

Temporal Sentiment analysis [1] that analyses both trends of sentiments and topics from a text archive which has timestamps of the occurrence of a period of the data. The main objective of this paper is to analyze the tendency of people’s sentiment along with its timeline. The method proposed in this paper has taken the text along with its timestamp from text archives like weblog and news articles etc., as its input and it creates two types of graph that is topic graph and sentiment graph. To do this analyze the former method used are Mood views [15], Topic Detection and Tracking [16], and Theme River [17]. The method proposed in this paper produces both graphs using sentiment phrases. In the future, they tend to use qualitative and quantitative evaluation, automatic construction of a database of sentiment phrases and to apply the proposed method to weblog articles.

Momentarily there is an enormous growth in the extent of research in the field of sentiment analysis[2]; mostly on high subjective text types such as movies, reviews etc., the main distinction between these texts with news articles is that their target is clearly defined and the opinion mining in news is different from these subjective text types. There are three dimensions of views in news articles.

Revised Manuscript Received on September 22, 2019.

A.Shunmuga Sundari, Part Time Ph.D Research Scholar, Rani Anna Government College for Women, Tirunelveli-627412, India.

Dr.R. Shenbagavalli, Assistant Professor, Rani Anna Government College for Women, Tirunelveli-627412, India.

They are author, reader and text. And also the three subtasks that need to be addressed are definition of the target, separation of the good and bad news content from the good and bad sentiment expressed on the target and analysis of clearly marked opinion that is expressed explicitly not needing interpretation or use of world knowledge.

The core purpose of this paper is to create a positive environment by affording a platform for serving only good news [3]. This can be achieved by analyzing the sentiments of the news article and then filtering out the articles which contains only the positive emotion. To make this vision into action at first they did a classification and then feature extraction. For classification they used Naïve Bayes [12], Support Vector Machine [11], and Maximum Entropy [13] algorithm. Among these algorithms Support Vector Machine is considered to be a popularly used classification algorithm. After the classification step then they built a model using training data. Once it is developed which is then used to classify the new dataset based on built model.

For concept level sentiment analysis SenticNet is a publicly available semantic and affective resource. SenticNet 3[4] procreate the use of 'energy flows' to merge various parts of elongated common and common- sense knowledge representation of one and another instead of using graph-mining and dimensionality-reduction techniques. SenticNet 3 is scheduled to link the conceptual and affective gap between word-level natural language data and the concept- level opinions and fetch the sentiments conveyed by them.

The conceptual framework of the News Sentiment Analysis [5] consists of four modules. They are Crawling and extraction module, Data preprocessing and Feature extraction module, Sentiment identification, scoring and classifier training module and finally Sentiment aggregation module. The main tasks identified for news opinion mining consists of extracting sentences from online published news articles that mentions company news, and identifying positive and negative sentiment that exist in that article and further summarizing the article polarity. A large number of companies use news analysis to help them make better business decisions so in this paper they did a sentiment analysis on news article related to company.

To monitor the public's opinion towards their brand and business organizations and individuals use a popular method called 'Sentiment analysis over Twitter data' now-a-days. The Noisy nature of data generated by Twitter is a major challenge of sentiment analysis methods. A famous Method to reduce the noisy nature is to remove stop words [6] by using pre-compiled lists of stop-words or by using dynamic stop-word identification by using more sophisticated methods. This paper investigates whether removing stop-words helps or hampers the effectiveness of sentiment classification.

News events which are an eloquent component of the social media big data on the web. Therefore analyzing these news events gives rich valuable information for both corporation and government. Sentiment computing of the news events [7] helps to discover the people's opinion about

the particular event from the text they posted. This paper finds out the multidimensional emotions like joy, anger, sorrow, love, surprise, and fear with the help of the semantics of the text written. For that it uses word emotion association network which is of like association link network (ALN) [18]

This Paper describes the UWaterloo affect prediction system [8] developed for EMOINT 2017. They tend to find out affect intensity, affect presence, sentiment intensity and sentiment presence, lexical alongside pre-trained word embedding's, which are utilized to extract emotion intensity signals from tweets in an ensemble learning approach. This System utilizes gradient boosted regression as the primary learning technique to predict the final emotion intensities.

Due to the broad applications in both, commercial and public sectors Sentiment analysis on Twitter has magnetized much fascination in neoteric times. This paper proposed a novel semantic based sentiment representation of words called SentiCircle [9], a lexicon-based approach which is able to assign context specific sentiment orientation to words. It takes the co-occurrence patterns of words in various contexts in tweets to seizure their semantics and also updates their pre-assigned strength and polarity in sentiment lexicons accordingly.

This paper [10] proposes and scrutinizes a new problem called social affective text mining, which courses to explore and model the association between online documents and user-originated social emotions. In terms to predict the sentiment from a text they intent to find the correlation between social emotions and affective terms. The techniques involved in this paper are the emotion term model, topic model and, emotion topic model.

This paper [19] proposed a hybrid system to depict the emotion from the medium. There are many ways that Human can express their emotions such as speech, image, facial expression and so forth. Due to the small requirements of textual data this paper focus only on text medium. This paper uses two methods, keyword based and machine learning based method to identify the emotion by the text.

In Paper [20] they presented the model for sentiment analysis of language learning using Naive Bayes Classifier. This paper utilizes Facebook status as data for the sample. The proposed system predicts whether the depicted emotion is positive, negative or neutral by using sentence level classification.

In paper [21] the author proposed a novel approach for emotion estimation from the text entered by the user on social networking sites. The author developed a visual image generation approach that generates images according to emotion in text.

This paper [22] proposed a system that automatically identifies the emotional state in the text which can be used to render facial expressions. In this, the author used a corpus of children's stories. This study used a supervised machine learning technique to classify children stories into one of the predefined emotion classes.

Generally, Text [23] can be written in two writing styles that are formal and informal. This paper analysis sentiment classification in both formal and informal text pieces. For emotion classification, they use different machine learning based methods SVM (support vector machine), NB (Naïve Bayes), Decision Tree.

III. METHODOLOGY

3.1 DATA PRE-PROCESSING

Pre-processing which is a very indispensable key step for sentiment analysis on a text. Microblogging text such as Tweets in prevail are not always linguistically well-structured and also the language used in tweets doesn't always precisely adhere to grammatical rules. Data cleansing is necessary because of the noisy nature of raw text in tweets such as special characters, using of Abbreviations, alphanumeric strings, words in short form etc., so to improve the accuracy of the model the tweets has to be cleaned. Here to normalize all the text in a tweet are standardized by converting the text in to lowercase. We remove stop words that project very low or no notable value to the process of sentiment analysis by using NLTK.

Stemming is performed to turn around the word to its root word. Then Tokenize the document by 1, 2 or n words depends upon the need by splitting the text of a document into a series of tokens in order to identify all words that affect the sentiment in a given document.

3.2 FEATURE EXTRACTION

We used two primary methods for feature extraction from the tweets namely annotated lexicons and pre-trained word embeddings [26].

3.2.1 Annotated lexicons

- **NRC affect intensity Lexicon:**

This lexicon assigns distinct emotion labels to unigrams and provide the intensity at which the emotion is expressed.

- **NRC Emotion Lexicon:**

It contains 3 fields namely term, affect category and association flag. Term is the word or token then affect category contains one of eight emotions (anger, fear, anticipation, trust, surprise, sad, joy or disgust) and the association flag has one of two possible values either 1 (indicates the term associated with the affect category) or 0 (the term is not associated with the affect category).

- **NRC Hashtag Emotion Lexicon:**

This lexicon contains 3 fields affect category, term and score. Affect category contains one of eight emotions (anger, fear, anticipation, trust, surprise, sad, joy or disgust), term is the word for which the emotion is given, and score specifies the value which indicates the strength of association between the terms and affect category. Higher the score then stronger the association.

- **NRC Emoticon Lexicon and NRC Hashtag sentiment**

Lexicon, NRC Emoticon Affirmative Context Lexicon and NRC Emoticon Negated Context Lexicon & NRC Hashtag affirmative context sentiment Lexicon and NRC Hashtag Negated Context Sentiment Lexicon:

All of these lexicons contains 4 fields such as term, score, Npos, Nneg. The terms in these lexicons could be unigrams, bi-grams or a pair of unigrams and bigrams. Score holds the real value sentiment score. Npos and Nneg indicates how many times the term appears in the positive and negative class.

- **SentiWordNet:**

It is one of the most popular opinion mining resource by NLTK. In this net each word is associated with the sentiment Sentiwordnet = wordnet + sentiment information

For each word (w) in sentiwordnet contains the following information

- o Positive score Pos(w)
 - o Negative score Neg(w)
 - o Objective score Obj(w)
- $$\text{Pos}(w) + \text{Neg}(w) + \text{Obj}(w) = 1$$

Here the positive and negative score of each word is taken as feature.

- **Emoji Valence:**

This lexicon contains score within the range of -5 (negative) to +5(positive) for each Unicode emoji's.

- **Depeche Mood:**

This lexicon comprised of more than 37,000 terms with emotion scores for emotions such as afraid, amused, angry, annoyed, don't care, happy, inspired and sad.

3.2.2. Word embeddings

Pre-trained Word embedding also takes part of vector representation of each tweet in addition to the features extracted from the annotated lexica.

Here it uses two word embedding schemes

- **Google News**

This is a word to vector model trained by the google news corpus contains more than 3 million distinct words and phrases of 300 dimensional embeddings.

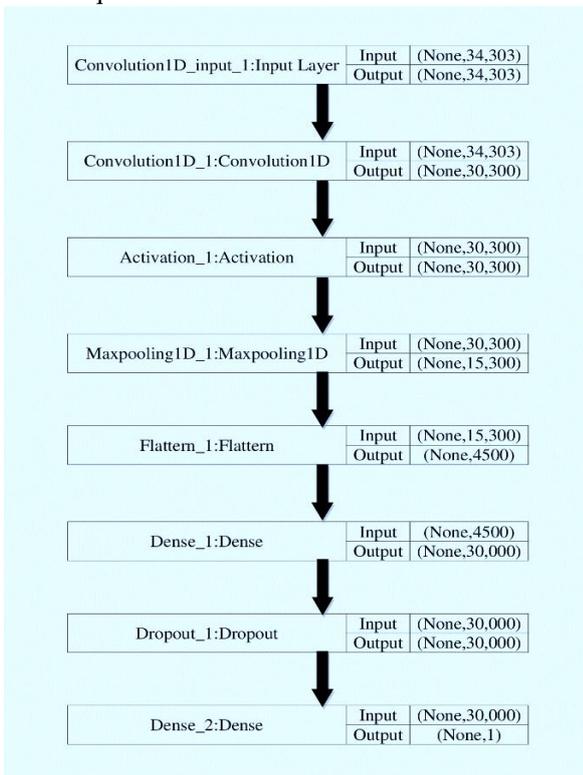
- **Glove Model:**

It is also very similar to word2vec model rather before training it builds a word co-occurrence matrix. Glove Model contains more than 6 billion words and phrases of 300 dimensional.

3 Model Training

3.3.1 CNN

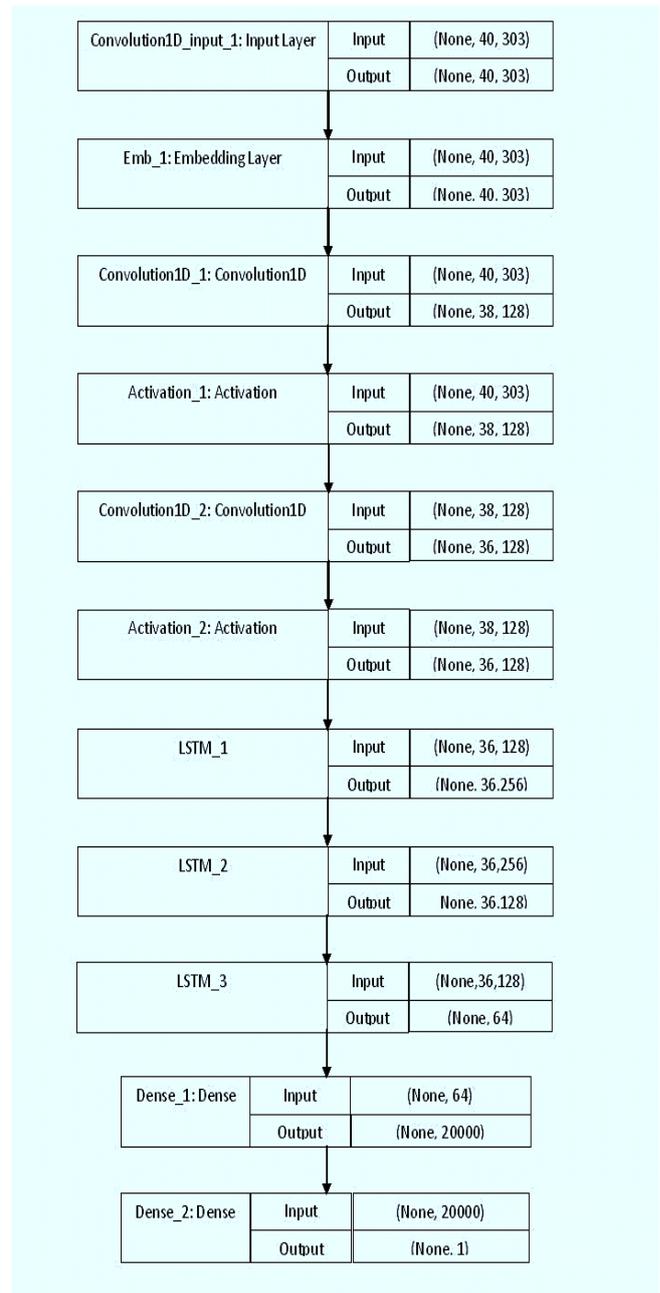
The input to the convolutional layer is of size (no of tweets*no of maximum occurred feature).First input layer of CNN is modeled with 300 filter each of length 5. ReLu is the activation technique used here. Followed by the input layer the next layer is of maximum pooling layer of size 2. Then the input is given to flatten and dense layer of size 30000. To avoid the over fitting issues of neurons we applied a dropout layer with rate of 0.5 to predict the emotion in more accurate way. Then the input is given to the dense layer of size 1 and the sigmoid activation technique is used. Finally the model is compiled with stochastic gradient descent (SGD) optimizer and mean square error loss function.



[Fig-1: CNN Layer specification]

3.3.2 LSTM

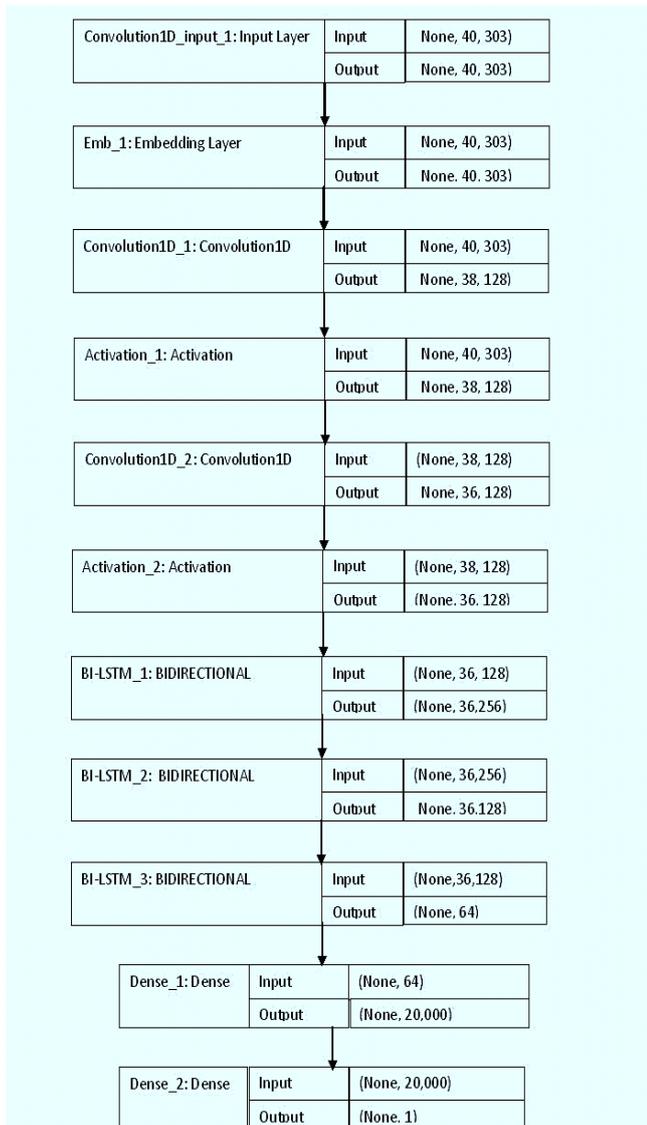
Initially one embedding layer is used then the input to the convolutional layer is of size (no of tweets*no of maximum occurred feature).Here 2 convolutional layer is modeled with 128 filters, each of length 3 and ReLu is the activation technique used. Followed by 3 LSTM layer of filter sizes 256, 128, 64 with 0.2 dropout and 0.2 recurrent. Next the input is given to the dense layer of units 20,000 with ReLu activation. Which is next followed by another one dense layer of units 1 with sigmoid activation. At last the model is compiled with stochastic gradient descent (SGD) optimizer and mean square error loss function. Finally kerasregressor is used to predict the results.



[Fig-2: LSTM Layer specification]

3.3.3BI-LSTM

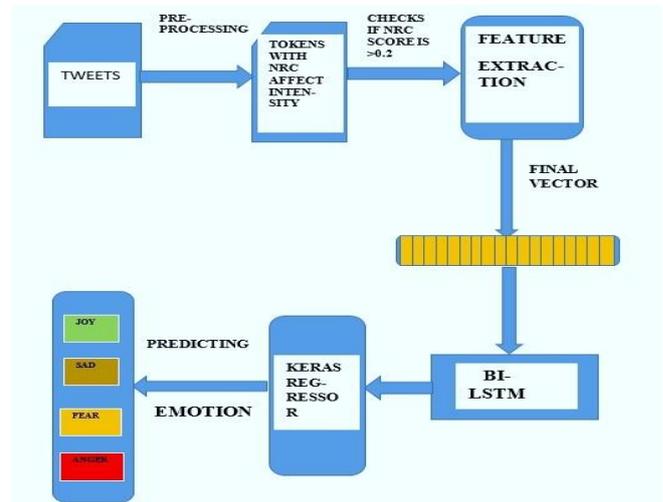
At premier one embedding layer is used then the input to the convolutional layer is of size (no of tweets*no of maximum occurred feature).Here 2 convolutional layer is modeled with 128 filters, each of length 3 and ReLu is the activation technique used. Followed by 3 BI-LSTM layer of filter sizes 128, 256, and 128 with 0.2 dropout and 0.2 recurrent. Next the input is given to the dense layer of units 20,000 with ReLu activation. Which is next followed by another one dense layer of units 1 with sigmoid activation. At last the model is compiled with stochastic gradient descent (SGD) optimizer and mean square error loss function. Finally kerasregressor is used to predict the results.



[Fig-3: BI-LSTM Layer specification]

3.3.4 Selective Lexicon based BI-LSTM (SL+BI-LSTM)[PROPOSED MODEL]

This is a proposed model to predict the emotion more accurately. Here at first for each lexicon we take the feature named NRC affect intensity. If the unigram scores the value which is greater than 0.2 in NRC affect intensity that word is selected for the process. Else all the feature extraction and word embeddings performed only for that selected lexicon. Then this final vector is given input for Bi-LSTM. Bi-LSTM has one embedding layer then 2 convolutional layer is modeled with 128 filters, each of length 3 and ReLu is the activation technique used. Followed by 3 BI-LSTM layer of filter sizes 128, 256, 128 with 0.2 dropout and 0.2 recurrent. Next the input is given to the dense layer of units 20,000 with ReLu activation. Which is next followed by another one dense layer of units 1 with sigmoid activation. At last the model is compiled with stochastic gradient descent (SGD) optimizer and mean square error loss function. Finally kerasregressor is used to predict the results.



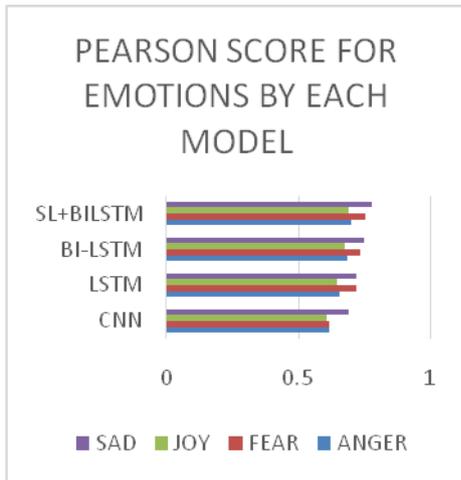
[Fig-4: SL+BI-LSTM MODEL WORK FLOW]

IV. RESULT

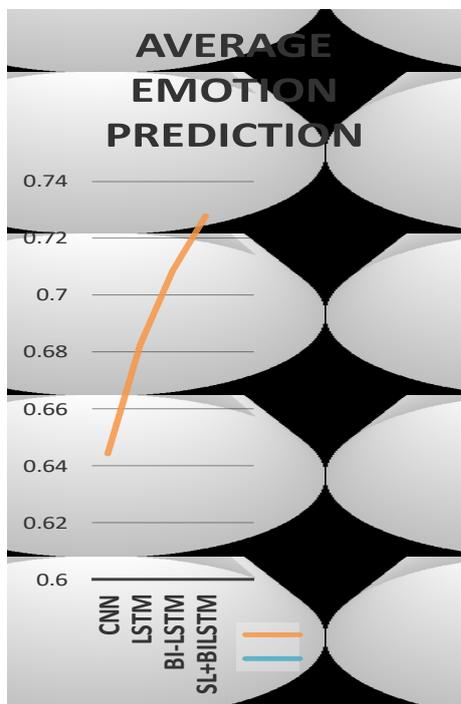
We used Pearson correlation as a analyze technique to check the accuracy of the result. The Pearson correlation generates a coefficient called the Pearson correlation coefficient, denoted as r. It is a measure of the strength and direction of association that exists between two continuous variables. In Pearson correlation calculation for the emotion anger CNN scores 0.6127, LSTM scores 0.653, BI-LSTM scores 0.685 and SL+BI-LSTM scores 0.698. For emotion fear CNN, LSTM, BI-LSTM, SL+BI-LSTM scores 0.6124, 0.715, 0.732, and 0.752 respectively. For emotion Joy CNN scores 0.6054, LSTM scores 0.646, BI-LSTM scores 0.672 and SL+BI-LSTM scores 0.687. For emotion Sad CNN scores 0.6127, LSTM scores 0.653, BI-LSTM scores 0.685 and SL+BI-LSTM scores 0.698. Here among all the models our proposed model SL+Bi-LSTM scores much better results for each emotion and get higher average score. As a result the proposed model Selective lexicon based BI-LSTM (SL+BI-LSTM) comes with a high average score (0.7278) in predicting emotions.

[TABLE-1: Pearson score for the models]

S.NO	MODELS	PEARSON SCORE				
		ANGER	FEAR	JOY	SAD	AVG
1.	CNN	0.6127	0.6724	0.6054	0.6865	0.64425
1.	LSTM	0.653	0.715	0.646	0.716	0.6825
2.	BI-LSTM	0.685	0.732	0.672	0.745	0.7085
3.	SL+BI-LSTM	0.698	0.752	0.687	0.774	0.7278



[Graph 1: Representation of Pearson scores of all models computed by Pearson calculation]



[Graph 2: Representation of Average scores of all models computed by Pearson calculation]

In this paper, we have presented a model called Selective Lexicon based Bi-LSTM for predicting intensity level of an emotion accurately. Preprocessing techniques such as data cleansing, stop words removal and stemming were applied to remove noisiness of tweets and normalize the tweets. The proposed model also takes only the lexicons which have high affect intensity on particular emotion and then all the other feature extraction and word embeddings are done only for that selective lexicon. This approach avoids and minimize the efforts taken to analyze the low affected lexicons for an emotion and also eliminates the redundancy tasks performed by the model to predict the emotion. This finalized vector is only given as an input for the Bi-LSTM model. The proposed approach attains significant improvement over the system. And the proposed model gets better emotion prediction for all the four emotion datasets i.e. anger, fear, joy and sadness.

REFERENCES

1. Fukuhara, T., Nakagawa, H., Nishida, T., 2007. Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events. In: ICWSM.
2. Alexandra Balahur, Ralf Steinberger, MijailKabadjov, VanniZavarella, Erik Van Der Goot, MatinaHalkia, Bruno Pouliquen, and JenyaBelyaeva. Sentiment analysis in the news. arXiv preprint arXiv:1309.6202, 2013.
3. Doddi, K.S. Haribhakta, M.Y.V. Kulkarni, P. (2014)“Sentiment Classification of News Articles”, KiranShriniwasDoddi et al, / (IJSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , pp. 4621-4623
4. E.Cambria, D.Olsher ,D.Rajagopal “Senticnet 3: A Common and common-sense knowledge base for cognition-driven sentiment analysis”, AAAI Conference on Artificial Intelligence (2014).
5. 1Ubale Swati,2Chilekar Pranali, 3Sonkamble Pragati “Sentiment Analysis Of News Articles Using Machine Learning Approach” ,International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-2, Issue-4, April-2015
6. H. Saif, Y. He, H. Alani, and M. Fernandez, “On stopwords, filtering and data sparsity for sentiment analysis of twitter,” in The International Conference on Language Resources and Evaluation, 2014.
7. D. Jiang, X. Luo, J. Xuan, and Z. Xu, “Sentiment Computing for the News Event Based on the Social Media Big Data”, IEEE Access (2017)
8. VineetJohn , Olga Vechtomova, “UWat-Emote at EmoInt-2017: Emotion Intensity Detection using Affect Clues, Sentiment Polarity and Word Embeddings”
9. H. Saif, Y. He, M. Fernandez, and H. Alani, “Contextual semantics for sentiment analysis of twitter,” Information Processing & Management, vol. 52, no. 1, pp. 5–19, 2015
10. K.Thendral, S.Chitrakala , “Emotion Recognition System for Affective Text” ,ISSN (Online): 2347 - 2812, Volume-2, Issue -11,12 2014
11. S.M.Namburu,H.Tu,J.Luo,andK.R.Pattipati,“ExperimentsonSupervisedLearningAlgorithmsforTextCategorization,”AerospaceConference.IEEE,BigSky,MT,pp.1–8,2004.
12. A.Scime,WebMining:ApplicationsandTechniques.IGIGlobal,2005.p.442.
13. E.Boiy,P.Hens,K.Deschacht,andM.Moens,“AutomaticSentimentAnalysisinOn-lineText,”ELPUB,no.June,2007.
14. Review of k-nearest neighbor text categorization method, http://www.cs.fucdavis.edu/~liaoy/research/text_ss02_html/node4.html
15. Mishne, G. and de Rijke, M., MoodViews: Tools for Blog Mood Analysis, AAAI 2006 Spring Symposium on Computational Approaches to analyzing Weblogs (AAAI-CAAW2006), 2006.
16. Wayne, C.L., Multilingual Topic Detection and Tracking: Successful Research Enabled by Corpora and Evaluation. Language Resources and Evaluation Conference (LREC), pp.1487-1494, 2000.
17. Havre, S., Hetzler, E., Whitney, P., and Nowell, L., ThemeRiver: Visualizing Thematic Changes in Large Document Collections, IEEE Transactions on Visualization and Computer Graphics, Vol.8, No.1, pp. 9-20, 2002.
18. X. Luo, Z. Xu, J. Yu, and X. Chen, “Building association link network for semantic link on web resources,” Automation Science and Engineering, IEEE Transactions on, vol. 8, no. 3, pp. 482-494, July 2011.
19. Yong-sooseol and Dong-jookim, “Emotion recognition from text using Knowledge-based ANN”, ITC-CSCC 2008
20. Christos Troussas, Maria Virvou, “Sentiment analysis of facebook statuses using Naive Bayes classifier for language learning”, 2012.
21. Tejasvinipatil and SachinPatil, “Automatic generation of emotions for social networking websites using text mining” in proceedings of IEEE-2012
22. R. A. Calix, S. A. Mallepudi, B. C. B. Chen, and G. M. Knapp. Emotion recognition in text for 3-d facial expression rendering, 2010.
23. J. Kaur and J. R. Saini, “Emotion Detection and Sentiment Analysis in Text Corpus: A Differential Study with Informal and Formal Writing Styles”, 2014.
24. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, <http://www.deeplearningbook.org>.
25. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, Nov 1997, pp. 1735–1780.
26. VineetJohn, OlgaVechtomova, “UWat-EmoteatEmoInt-2017: Emotion Intensity Detection using Affect Clues, Sentiment Polarity and Word Embeddings”.
27. Saif M. Mohammad and Felipe Bravo-Marquez. 2017. WASSA-2017SharedTaskonEmotionIntensity. In Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA). Copenhagen, Denmark.