International Journal of Scientific Research in Computer Science and Engineering **Research Paper**
Vol.8, Issue.1, pp.01-06, February (2020) **E-ISSN:** 2320-7639
DOI: https://doi.org/10.26438/ijsrcse/v8i1.16

Sentimental Analysis of Twitter Using Long Short-Term Memory and Gate Recurrent Unit

S. Suriya[1*], K. Sindhu Meena[2]

[1]Dept. of Computer Science and Engineering, PSG College of Technology, Coimbatore, India
[2]Dept. of Information Technology, Sri krishna College of Technology, Coimbatore, India

[*]*Corresponding Author: suriyas84@gmail.com, ss.cse@psgetch.ac.in*

Available online at: www.isroset.org

Received: 28/Dec/2019, Accepted: 10/Feb/2020, Online: 28/Feb/2020

***Abstract*—** Sentimental analysis is the process of discovering and categorizing the opinions in the texts/reviews. It is to determine the attitude of the speaker and also to identify the topic, product whether it is positive, negative or neutral. It is used in the process on analysing the texts, computing linguistics and biometrics for identifying systematically extracting, quantifying the information. This paper aims to classify the reviews and apply the word embedding techniques and to find which is more suitable word embedding for LSTM and GRU.LSTM and GRU are the solution for the vanishing gradient problem. LSTM more suits for classifying, processing and predicting the time series data. It is composed of 3 states input, output and forget state. The main thing about GRU is it can train to keep a information from long ago without wasting it through time or remove irrelevant information to be predicted. GRU is composed of update and reset gates. The performance are evaluated by using the accuracy, precision, recall and f1score.The anaconda navigator is used as a tool and python is the language used.

***Keywords*—** LSTM, GRU, sentimental analysis, word embedding techniques

I. INTRODUCTION

Deep learning is also called as a structured learning or hierarchical learning it is a part of the machine learning techniques depends on the layers used in the artificial neural networks. The learning may be of supervised, unsupervised, semi supervised. Deep neural networks, deep belief networks, recurrent neural networks and convolution neural network are used and applied in the field of computer vision, recognization of speech, natural language processing, bio informatics, filtering of social networks. Sentimental analysis is mining the context of a text and identifies the subjective information. It understands the sentiment of the product, topic while conversing through online. It identifies whether the given text is positive, negative or neutral. Sentimental analysis is also known as opinion mining. It tracks the mood of a public about a particular product. It collects and categorize the opinion about a product. Opinion mining uses machine learning. It analyses the text and predicts whether the text is positive, negative or neutral. It identifies whether it relates about opinion, news, marketing, suggestion, query. It helps for detecting the emotions like angry, sad, happy. The sentimental analysis is classifying the polarity of a given text at the document, sentence is positive, negative or neutral. It is for predicting for various aspects of the given restaurants. In our paper it helps to classify the reviews of a twitter and to predict whether the user reviews is positive or negative or neutral. We will ignore the neutral because either the review will be considered as a positive or negative so will drop the neutral reviews. The word embedding is a word or phrase of a paragraph is mapped into a vector space by using real value numbers. Word embedding is representing a word that act as a bridge the understanding the language by the human to that of machine. It is representing the text in n-dimensional space. It solves many Natural Language Processing problems. Word embedding helps to predict the meaning of texts and similarity of words. The words with similar meaning are represented similarly. Some of the word embedding techniques are word2vec, doc2vec, glove, Elmo, fasttext.Here the word embedding techniques like word2vec, doc2vec, glove, fastext and elmo are taken LSTM and GRU is applied to it and find which word embedding techniques works efficiently with LSTM and GRU. It is used in deep learning such as sentimental analysis, parsing the syntactic, name entity recognization. Word embedding makes easier to learn the machine learning on larger inputs like sparse vectors representing the words or paragraphs. It helps to extract the meaning for a text to enable the natural language processing.

II. LITERATURE SURVEY

In Paper [1] proposes a method based on the sentiment analysis technique and the intuitionistic fuzzy set theory to rank the products through online reviews. An algorithm based on sentiment dictionaries is developed to identify the

positive, neutral or negative sentiment orientation on the alternative product concerning the product feature in each review. High Performance is achieved. Machine Learning technique is used. Paper [2], reviews the various current approaches and tools used for multilingual sentiment analysis, identifies challenges along this line of research, and provides several recommendations including a framework that is particularly applicable for dealing with scarce resource languages. They recommended possible remedies as well as a hybrid framework for developing sentiment analysis resources particularly for languages with limited electronic resources. Artificial Intelligence is used. In Paper [3], Unsupervised approach is being used for the automatic identification of sentiment for tweets acquired from Twitter public domain. the supervised approach where we combine unigram, bigram and Part-of-Speech as feature is efficient in finding emotion and sentiment of unstructured data. For short message services, using the unigram feature with MNB classifier allows us to achieve an accuracy of 67%. It achieves an accuracy of 80.68%. Multinomial Naive Bayes (MNB), Maximum Entropy and Support Vector Machines are used. In Paper [4], they give a collection of review texts, the goal is to detect the individual product aspects comments by reviewers and to decide whether the comments are rather positive or negative. Better performance is achieved. UnSupervised approach is used. In Paper [5], aims is to promote research in sentimental analysis of tweets by providing annotated tweets for training, development and testing. The objective of the system is to label the sentiment of each tweet as "positive", "negative", "neutral". It maximizes the positive and negative precision and recall, the rule-based classifier used to correct or verify the neutral SVM predictions. In Paper [6], is the study of the people's opinions, attitudes and emotions towards an entity. Opinion helps to collect information about the positive and negative aspects of a particular topic. Finally, the positive and highly scored opinions obtained about a particular product are recommended to the user. The positive and highly scored opinions obtained about a particular product are recommended to the user. Several challenges in Sentimental analysis. The first is an opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in same way. Paper [7], examines customer reviews. It is used to extract most important aspects of an item and to predict the orientation of each aspect from the item reviews. The people cannot analyse exact information in the document and sentence level opinion mining on customer reviews. Aspect level opinion mining is one of the solutions to problem. This gives fine detail information in aspect level. Benefit of using Naïve Bayes is to identify whether sentence is positive or negative opinion and also identifies the number of positive and negative opinion of each extracted aspect. The number of positive and negative opinions in review sentences is estimated. Sentiment orientation gives a good accuracy

III. RESEARCH APPROACH

3.1 Data Collection

Dataset is a collection of the related sets of information that contains separate attributes. It is a collection of data. It is the contents of a database each column has a variable and the row contains the members of the dataset. The dataset contains 1.6 million entries and no null entries in the sentimental column. Here the 50% of data is negative and another 50% is positive. It contains of id, reviews and sentiment where the id contains the id's assigned to users it is of unique number, Review contains the tweets by users and sentiment has the details about whether the user tweet is positive or negative. If it is positive the value will be of 1 and if it is negative it will be of 0. It is a tweets which are taken from the different cities. The dataset is of csv format. It is named as clean_tweet.csv. It was downloaded from the github. The sentiment is of binary classification whether zero or one.

3.2 Pre-processing - Data Pre-processing is to convert the raw data into a efficient and useful data. Our raw data in dataset will contain some noise, missing values and inconsistencies. In order to improve the quality of the data and to improve the efficiency of the raw data. It deals with the preparation and transformation of the initial dataset. Data Pre-processing are divided into following categories: Data Cleaning, Data Integration, Data Transformation, Data Reduction. Data Cleaning is If the data in the dataset is incomplete (lacking attribute values), noisy data (error values or outlier values) and inconsistencies. The steps for each process are explained below: Missing values is If the dataset contains many tuples which doesn't have values for several attributes, then those values can be filled in for the attribute by ignoring the tuple, Fill in the missing values, Use a global constant, Use the attribute mean to for all samples belonging to the same class as the given tuple, Use the most probable values to fill in the missing values. Noisy Data is a random error or variance in a measured variable. The data are smoothened to remove the noise. To smoothen the data some techniques are used: Binning, Clustering, Combined Computer and Human inspection, Regression. Data Integration is It combines the data from multiple source into a coherent data and store n data warehouse. The sources may be of database, data cubes or files. Several problems may occur during data Integration. Redundancy is the main issue in the data integration. Data transformation is the data are transformed into appropriate forms.

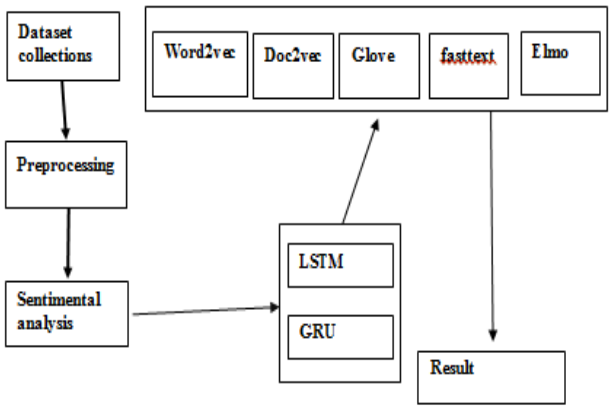

Figure 1 Proposed System Architecture

Data integration can involve following steps: Normalization, Aggregation, generalization, Normalization is the attributes are scakled so that it can be in specified range. Smoothing removes the noises from data using binning, clustering and regression, Aggregation is based on the aggregated data the results are derived. Generalization is the low level of raw data are replaced by the higher levels through the hierarchies. Data Reduction is Mining on the huge data is very complex and its infeasible. It helps to analyse the reduced representation of dataset without altering the original data. It is either reducing the data or dimension. Steps involved in Data Reduction are: Data Cube Aggregation, Dimension reduction, Data Compression, Numerosity reduction, Discretization. Data Cube Aggregation is applied in the construction of data cube. Dimension reduction- irrelevant, redundant attributes or dimensions are detected and removed. Numerosity reduction is the data replaced by alternative. Discretization is the raw data values are replaced by the specific range. Hence what we done in our paper is: Tokenization Loaded the dataset, splited it by the white spaces, Select words (regex model and split the documents into words), Split by Whitespace and remove punctuation, Normalizing(All case are converted to lower case).

3.3 Word Embedding Technique - Word embedding are the type of word representation that allows words with a similar meaning and represented similarly. They are distributed representation for text to improve the performance of deep learning methods on natural language processing problems. It is mainly used in document vocabulary. It takes the context of words, semantic and syntactic similarity. It is a word or phrase in the paragraph are mapped into a space vectors which is a real value number. It is mapping a low dimensional space into high dimensional space vectors. It takes a large input as a sparse vector denoting words. It literally means a word mapped into a vector. While applying the one hot encoding we may have many zeros sparse vectors of high dimensionality. For larger dataset it will have performance issue. So words like aircraft and aeroplane refers to the same meaning instead of marking in a different vector, the similar words are mapped into a same vector. We work with twitter sentimental analysis where it contains positive and negative. The idea behind word embedding technique is converting the word to numbers. Another way is using the one hot encoder. Each tweets are represented in a vectors with a dimension equal to the words in the documents. They are mapped into multidimensional space. The words with similar meaning or equal words will be closer by distance the words which are entirely different are larger by distance. Some of the word embedding techniques we used a word2vec, doc2vec, glove, fasttext, Elmo.

3.3.1 Word2vec - Word2vec is a word embedding techniques. They are known as shallow, two-layer neural network they are trained to reconstruct the linguistic context of words. Word2vec takes large corpus of text as a input and produce it in a space vector, each unique word are mapped into a new vector in a space. The words which share the common context are positioned into a similar space and the words with a similar context are located closer proximity in the space. Word2vec contains skip-grams and continuous bag of words(CBOW). The skip gram is used to predict the context words for the centre words by taking the centre word and window of the context words or neighbour within the context words. Whereas the continuous bag of words is opposite to the skip gram and it predict the centre word from the given context by summing the vectors of surrounding words. The Wikipedia pre trained model is downloaded and accuracy we obtained is 69% with LSTM and 67% with GRU. So word2vec efficiently work with LSTM. It uses genism package.

3.3.2 Doc2vec - It generates the vectors for paragraph and documents. In word2vec it projects the words into latent d-dimensional space whereas doc2vec is a projecting a document into a latent d-dimensional space. It finds the similarity between documents. The concept of Mikilov and Le was used they used the word2vec model but added another vector to it(paragraph ID). It is to create the numeric representation for the documents. Instead of using word for predicting the next words we used the unique document id. So, while training the word vectors the document is also trained and at last the document is represented numerically. It uses the Distributed Memory Version of Paragraph Vector which acts as a memory which remembers what is missing from the given context or topic of paragraph. It uses the Wikipedia document is downloaded and used as a pre trained model. The accuracy we obtained while working with LSTM is 74% and 76% with GRU. So doc2vec efficiently works with GRU.

3.3.3 Glove - Glove denotes Global vectors for word representation. It aggregates the word-word co-occurrence matrix from the corpus. The result of glove is linear substructure of the words in a vector space. The global matrix factorization is used. It is the process of using a method from linear algebra to perform the rank reduction on larger frequency matrix. They are represented by the term document frequencies. The rows denote words and

columns denotes document. It is applied to the term document frequency called as Latent Semantic analysis. Glove is how frequently a word appear together in a larger text corpora. It is a count-based model. It learns the vectors by performing the dimensionality reduction on a co-occurrence count. Construct a large matrix of co-occurrence information which has the information about how frequently each word occur, then they are factorize this matrix to get a lower dimensional matrix of words and features. It minimizes the reconstruction loss. It uses the Wikipedia document is downloaded and used as a pre trained model. The accuracy we obtained while working with LSTM is 80% and 83% with GRU. So, glove efficiently works with GRU.

3.3.4 Fast text - Fast text is a library which is developed by the facebook which mainly helps to learn the word vectors and classify the texts. Instead of considering the words being independent of each other where fast text takes all the character sequences while computing a representation for the words. It produces better vectors for the rare words that is if we don't have adequate context for the words because it presents very few time in the corpus ,so we will be using the related words context.And it also helps in computing the vectors the unseen words.It represents the sentences with bag of words and bag of n-grams and it make use of the sub word information and it share the information for the hidden representation class also. It is used to speedup the computation. But it is slow to train and test the dataset. So fast text helps to solve this problem. It use a hierarchical classifier instead of a flat structure. It reduces the time complexity of training and testing text classifiers. It represent the text in a low dimensional vectors.It allow the information about the words of one category used by the other category known as bag of words.Here we use Wikipedia as a pretrained model.The accuracy obtained while working with LSTM is 68% and with GRU is 65%. So fast text works efficiently with LSTM.

3.3.5 Elmo - Elmo stands for Embedding from Language Models it was developed by Allen NLP. ELMo is a way of representing a word in a vector or embedding. It helps in achieving state of the art for different NLP tasks. Elmo uses two layered bidirectional language model .Each layer consists of two passes backward pass and forward pass. Initially it represents the words of a string into a raw word vector. It act as a input for the first layer of biLM. The forward pass contains the information about a certain words and context before the word whereas the backward pass contains the information about the word and context after it. The information from the backward pass and forward pass forms the intermediate word vectors. These intermediate words vector is fed into the next layer of biLM. At last the weighted sum of raw word vectors and 2 intermediate vectors. Here we use Wikipedia as a pretrained model. The accuracy obtained while working with LSTM is 63%.

IV. DEVELOPMENT PRACTICES

The LSTM and GRU are the models used. LSTM is type of a recurrent neural network (RNN) architecture which remembers the values over arbitrary intervals. LSTM is applied to classify, process and predict the time series of unknown duration. The RNN structure is similar to hidden Markov model. Advantage of LSTM is insensitivity to gap length. The long-term memory is called as the cell state. The looping arrows indicate recursive nature of the cell. This allows information from previous to be stored with in the LSTM cell. Cell state is modified by the forget gate placed below the cell state and also adjust by the input modulation gate. From equation, the previous cell state forgets by multiply with the forget gate and adds new information through the output of the input gates. The remember vector is also called as the forget gate. The output of the forget gate says to the cell state which information should be forgot by multiplying 0 to a position in the matrix. If the output of the forget gate is 1, the information is kept in the cell state. From equation, sigmoid function is applied to the weighted input/observation and previous hidden state. The save vector is usually called the input gate. These gates determine which information should enter the cell state / long-term memory. GRU (Gated Recurrent Unit) tends to solve the vanishing gradient problem. GRU can also be considered as a variation on the LSTM because both are designed similarly and, in some cases, produce equally excellent results. To solve the vanishing gradient problem of a standard RNN, GRU uses, so called, update gate and reset gate. Basically, these are two vectors which decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction. The update gate helps the model to determine how much of the past information (from previous time steps) needs to be passed along to the future. this gate is used from the model to decide how much of the past information to forget.

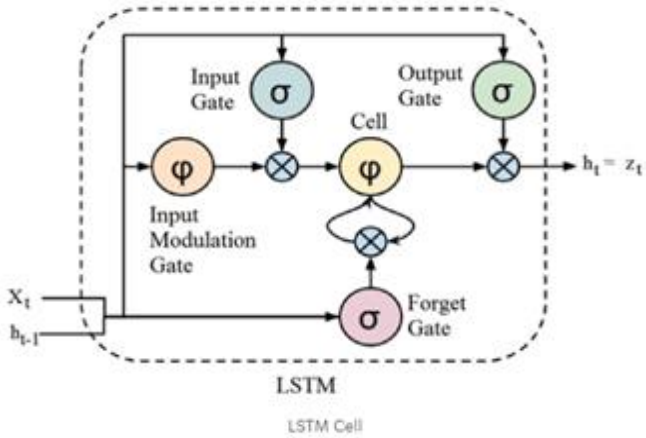

Figure 2 System Architecture of LSTM

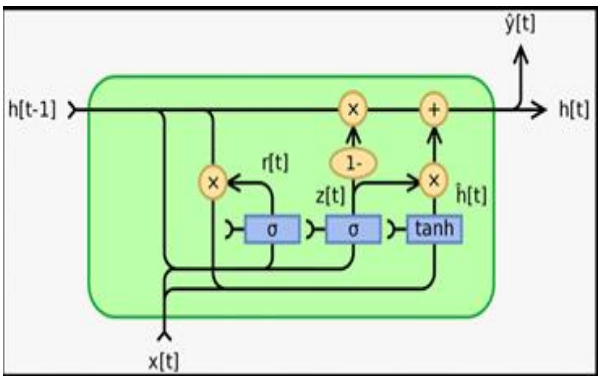

Figure 3 System Architecture of GRU

Training the Model - The dataset is preprocessed and the word embedding technique is selected, the efficient word embedding technique is selected by checking which word embedding techniques works well with LSTM and GRU the accuracy is checked. Here the glove has the highest accuracy with LSTM and GRU. The sentimental analysis is performed. The keras toolkit is used for implementation. We split the dataset into 80% for training and 20% for testing. The input vocabulary is set to 50K and maximum length of the sentence to 15 and maximum number of words in a sentence to 100.The glove pretrained vector is used for the embedding with 100 dimension and trained with our dataset. We trained the models for 10 epochs and used early stopping to determine the stopping condition for the iteration. When loss decrease and accuracy increases the epoch stops. Early stopping is val_loss increases at some epoch and remains same for particular time then the training will be stopped. Adam optimizer is used. The categorical cross entrophy is used for the output of probability.

V. RESULT AND ANALYSIS

We implemented word embedding technique with LSTM and GRU. The evaluation is done for the twitter sentimental dataset and compared the result of word embedding techniques like word2vec, doc2vec,glove,fast text and Elmo. The Glove efficiently works well with LSTM and GRU. The accuracy for glove is 80% and 83%.

Table 1 Result of Twitter Sentimental Analysis Dataset

Approach	Accuracy
Word2vec with LSTM	0.69
Word2vec with GRU	0.67
Doc2vec with LSTM	0.74
Doc2vec with GRU	0.76
Glove with LSTM	0.80
Glove with GRU	0.83
Fast text with LSTM	0.68
Fast text with GRU	0.65
ELMo with LSTM	0.63

VI. CONCLUSIONS AND FUTURE WORK

Text classification is an important task in Natural Language Processing to classify whether the given tweet is positive or negative. If it is positive it is denoted as 1 if it is negative it is denoted by 0. The major challenge in text classification is predict the categories accurately. The proposed approach classifies the tweets properly and comparatively the performance also increases with the existing approach. In future the ULM fit and other embeddings can be used and also try to improve the performance.

REFERENCES

[1] Chikersal, P., Poria, S., & Cambria, E., "SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning", In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), **pp. 647-651, 2015.**

[2] Zemmal, N., Azizi, N., Dey, N., & Sellami, M., "Adaptive semi supervised support vector machine semi supervised learning with features cooperation for breast cancer classification", Journal of Medical Imaging and Health Informatics, **Vol.6, Issue.1**, pp.53-62, **2016.**

[3] Vanitha, C. D. A., Devaraj, D., & Venkatesulu, M. (2015). Gene expression data classification using support vector machine and mutual information-based gene selection. procedia computer science, 47, 13-21.

[4] Joshi, A., Dangra, J., & Rawat, M., "A decision tree based classification technique for accurate heart disease classification and prediction", International Journal of Technology in Resource Management, **Vol. 3,** pp.1-4**, 2016.**

[5] Gai, K., Qiu, M., & Elnagdy, S. A., "Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data." , In IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), **pp. 197-202, 2016**

[6] Sarmah, J., & Sarma, S. K. (2016). Decision tree based supervised word sense disambiguation for Assamese. Int. J. Comput. Appl, 141(1).

[7] Jeyapriya, A., & Selvi, C. K., "Extracting aspects and mining opinions in product reviews using supervised learning algorithm", In 2015 2nd International Conference on Electronics and Communication Systems (ICECS), **pp. 548-552, 2015.**

[8] Feizollah, A., Ainin, S., Anuar, N. B., Abdullah, N. A. B., & Hazim, M., "Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms"., IEEE Access**, Vol.7**, pp.83354-83362, **2019**

[9] Wen, S., Wei, H., Yang, Y., Guo, Z., Zeng, Z., Huang, T., & Chen, Y., "Memristive LSTM Network for Sentiment Analysis", IEEE Transactions on Systems, Man, and Cybernetics: Systems, **2019.**

[10] Gaikar, D., Sapare, G., Vishwakarma, A., & Parkar, A., "Twitter Sentimental Analysis for Predicting Election Result using LSTM Neural Network", International Research Journal of Engineering and Technology, **Vol.6, Issue.4, 2019.**

[11] Ge, S., Qi, T., Wu, C., & Huang, Y. "Dialog Emotion Classification using Attentional LSTM-CNN", In Proceedings of the 13th International Workshop on Semantic Evaluation, **pp. 340-344, 2019.**

[12] Minaee, S., Azimi, E., & Abdolrashidi, A., "Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM

Models", Computation and Language, Cornell University, ref: arXiv preprint arXiv:1904.04206, **2019.**

[13] Yu, Q., Zhao, H., & Wang, Z., "Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis", In Proceedings of the 2nd ACM International Conference on Artificial Intelligence and Pattern Recognition, **pp. 116-119, 2019.**

[14] Li, C., Li, C., & Liu, P., "Sentiment Analysis Based on LSTM Architecture with Emoticon Attention", In Pacific-Asia Springer Conference on Knowledge Discovery and Data Mining, **pp. 232-242, 2019.**

[15] Rosenthal, S., Farra, N., & Nakov, P., "Sentiment analysis in Twitter", Computation and Language, Cornell University, ref: arXiv preprint arXiv:1912.00741, **2019**

[16] Han, H., Li, X., Zhi, S., & Wang, H., "Multi-Attention Network for Aspect Sentiment Analysis", In Proceedings of the 8th ACM International Conference on Software and Computer Applications, **pp. 22-26, 2019.**

[17] Reddy, A. V. M., Dinesh, Y., Krishna, V., & Miranam, S., "Stock market prediction using RNN and sentiment analysis", International Journal of Advance Research, Ideas and Innovations in Technology, **Vol.5, Issue.3,** pp.5-9, **2019.**

[18] Kwaik, K. A., Saad, M., Chatzikyriakidis, S., & Dobnik, S., "LSTM-CNN Deep Learning Model for Sentiment Analysis of Dialectal Arabic", In Proceedings of Springer International Conference on Arabic Language Processing, **pp. 108-121, 2019.**

[19] Abulaish, M., Rahimi, M. B., Ebrahemi, H., & Sah, A. K., "SentiLangN: A Language-Neutral Graph-Based Approach for Sentiment Analysis in Microblogging Data", In proceedings IEEE/WIC/ACM International Conference on Web Intelligence, **pp. 461-465, 2019.**

