DEEP BIDIRECTIONAL RECURRENT NEURAL NETWORKS-BASED SENTIMENTAL ANALYSIS OVER BIG DATA

P.KALADEVI

Assistant Professor, Department of Computer Science and Engineering,K.S.Rangasamy College of Technology, KSR Kalvi Nagar, Tiruchengode, Namakkal(Dt)-637215, TamilNadu,India.E-Mail: kaladevi@ksrct.ac.in

Dr.K.THYAGARAJAH

Professor, Department of Electronics and Communication Engineering, K.S.Rangasamy College of Technology, KSR Kalvi Nagar, Tiruchengode, Namakkal(Dt) - 637215, Tamil Nadu, India.E-Mail: drkt52@gmail.com

ABSTRACT:

The significant opportunities and challenges in the research of text mining is realized in the recent days due to the speedy increase in the amount of unstructured textual data with suitable tools for investigating them. This Deep Bidirectional Recurrent Neural Networks-based sentimental analysis Approach is determined to be sentiment polarity, since it is capable of preparing a dataset with sentiment for the objective of training and testing that is potential in extracting unbiased opinions. In this paper, Deep Bidirectional Recurrent Neural Networks-based Sentiment Analysis (DBRNN-SA) Scheme was proposed over Big data for preventing the challenges and investing the vital opportunities in the process of text mining. This proposed DBRNN-SA Scheme in particular is contributed to establish a framework that facilitates opinion mining using sentimental analysis for the case of students' university choice feedback. This proposed DBRNN-SA Scheme is compared with the existing frameworks in order to determine a reliable deep neural network that aids as a suitable classification entity in the process of sentimental analysis.

Keywords: Deep Bidirectional Recurrent Neural Networks, Sentimental analysis, Opinion Mining, Sentiment Polarity, Text mining.

1. INTRODUCTION

The social media and its associated applications facilitate the option of permitting millions of users to express and

disseminate their opinions about a concerned topic in order to express their attribute by supporting or disliking a content [1]. This process of expressing the options of a specific topic incur constantly accumulating actions that in turn generate high value, high volume, high dimension and high velocity data considered as the big social data [2]. In general, this big social data refer to the massive collection of opinions that has the maximum probability to be processed for elucidating the tendency and potentialities of the users in the digital realm [3]. A diversified number of researchers have considered a focused view and concentration in the process of exploiting big data, with the objective of describing, estimating and predicting human perceptions and characteristics over the significant domains of applications [4]. At this juncture, text analysis is considered for the potential for determining the human perceptions and characteristics of million users in the social network. In particular, nearly 82% of the data available over the internet are text and hence the process of text analysis is the key component essential for the process of elucidating public opinion, emotion and sentiments [5]. In this context, sentiment analysis also termed as opinion mining is determined to be suitable in estimating users' sentiment related to a topic by investigating their diversified actions and posts on the social media [6]. This sentiment analysis is potential categorizing the polarity of the posts into various opposite emotions that includes positive feedback, negative feedback and the like [7]. This sentiment analysis is classified into lexicon analysis and machine learning analysis. The former focuses on the objective in estimating the degree of polarity of the specific content depending on the words' semantic orientation perspective or the phrases present in the document. But, this lexicon analysis method of sentiments fails to consider the context under which the investigation is applied [8]. The latter focuses on the process of building models that are gathered from the training dataset that are labeled for estimating the orientation degree of the document (instances of sentences and text are considered). The machine learning analysis method is widely suitable and applicable in the process of gathering public opinion for assessing the satisfying internauts pertaining to a subject incorporated in a big social data. The subjects in which the machine learning analysis method can be enforced includes products, topics, persons, events and services in diversified domains that are not limited to marketing, politics and health [10]. However,

the enforcement of machine learning analysis method results in a highly variable accuracy level that changes between each and every method. Further, these machine learning analysis methods fail due to the challenges of words' semantic orientation that dynamically varies depending on the context in the opinion mining. In this context, deep learning-based opinion mining schemes are considered to be important in appropriate estimation of users; emotions, sentiments, likes and dislikes.

In this paper, the proposed DBRNN-SA scheme is contributed as a reliable machine learning-based sentiment analysis method incorporated for opinion mining about a peculiar topic of interest. This proposed DBRNN-SA scheme concentrates on handling semantic analysis that introduces a potential adaptive approach that purely depends on social media posts and big data framework to investigate internauts' behaviors and emotions towards a subject in a real time scenario. This proposed DBRNN-SA scheme embeds the model of Deep Bidirectional Recurrent Neural Networks for imposing the task of categorizing public opinions into positive or negative sentiments.

2. RELATED WORK

In this section, the most potential sentiment analysis techniques contributed in the big data domain over the recent years are presented with their merits and shortcomings.

Initially, a deep convolution neural network-based twitter sentiment analysis method free from language agnostic translation was proposed for determining the appropriate polarity of tweets that are unique for different languages [11]. This deep convolution neural network-based twitter sentiment analysis method incorporated a few parameters compared to the most accurate deep neural architectures used for sentimental analysis. This deep convolution neural network scheme is potential of learning the latent features that are incorporated for facilitating the process of training. This neural network scheme does not necessitate any process of translation as it inherently embeds the process of investigating the various patterns of the language in a significant manner. The empirical investigation of this deep convolution neural network scheme was determined to be significant compared to the traditional methods used for sentimental analysis. An Integrated Convolutional Neural Network and Recurrent Neural Networks

(ICNN-RNN) - based sentiment analysis of short texts was proposed for utilizing only a limited amount of contextual information [12]. This ICNN-RNN- based sentiment analysis method inherited the merits of local features that are coarse grained through the incorporation of CNN that widely explores the parameters essential for sentiment analysis. The accuracy of this ICNN-RNN-based sentiment analysis method was determined to be 83% superior to the benchmarked corpora schemes used for investigation.

An expressive model for sentiment analysis was contributed for the effective maximum exploration level of influential features that aids in estimating the appropriate polarity for different languages [13]. This expressive model aided in exploring the possible categories of semantics that build up the meaning of linguistic description that could be possibly encountered in the document used for the purpose of classification. A generalized big data framework was developed for the process of sentiment analysis using the benefits of Naive Bayes algorithm [14]. This generalized big data framework ensured high storage capacity, processing potential with the view to improve the effectiveness in handling the data with the utmost maximum speed. The accuracy of this generalized big data framework was determined to be enhanced by 89% superior to the benchmarked corpora schemes used for investigation. A feature and news polarity classification approach for sentiment analysis was proposed based on the benefits of ontology [15]. This ontology-based sentiment analysis method wide opened the option of semantically expressing the associations between the entities in the domain that are highly correlated with the financial news. This ontology-based sentiment analysis method utilized linguistic description of each and every feature by considering the number the words that contains the polarity of the features. The linguistic description of every word is estimated through the methods of all phrases, N-GRAM before, N-GRAM after and N-GRAM around. This ontology-based sentiment analysis method confirmed an F-score of 64.92% and accuracy of 66.73% in the process of feature polarity classification.

Further, a sentiment analysis method for optimal extraction of product features from customer opinion was proposed for text analysis [16]. This sentimental analytical approach is formulated for retrieving the tweets that are highly related to the product that are derived from the data set for

determining the degree of polarity. The investigation of this sentiment analysis method is achieved using Bing Lius Group dataset that contained reviews that are concerned to the products' features of Nokia , Samsung, Iphone and Canon. The accuracy of this sentiment analysis is determined to highly superior compared to the lexicon-based sentiment analysis approaches. Then, a deep learning scheme for facilitating sentiment analysis over Stocktwits was contributed to enhance the performance under opinion mining [17]. This Stocktwits-based sentiment analysis inherited the benefits of convolutional neural network, doc2vc and long short-term memory for mining opinion collection and investigation. The mean absolute error and Root mean square Error of this deep learning method was determined to be comparatively 18% and 15% superior to the ontology-based sentiment analysis method. An adaptive sentimental analysis scheme was proposed for extracting users' opinion through the construction of dynamically less polarized words dictionary [18]. This adaptive sentimental analysis aided in classifying the tweets that correlate with the post which possess a maximum degree of polarity. The precision and recall value of the adaptive sentiment analysis was determined to be comparatively 22% and 31% superior to the ontology-based sentiment analysis method.

Furthermore, a deep learning model for sentiment analysis was contributed for opinion mining about Hindi movie reviews based on multiple linguistic description [19]. This deep learning model for sentiment analysis used a variable number of convolution layers and filters with different epoch value for facilitating the process of training. This deep learning model was tested for its predominance with 50% of dataset for training and the remaining 50% of dataset for testing purpose. This deep learning model is estimated to facilitate 95% of accuracy compared to the ontology-driven sentiment schemes used for investigation. In addition, Another deep neural network-based sentiment analysis scheme was proposed for minimizing the time used for training through the utilization of region-based long short term memory [20]. This deep neural network-based sentiment analysis scheme was estimated to facilitate superior accuracy of 99.82%, precision of 98.54%, recall of 98.21% and kappa score of 97.32% compared the deep neural network-based sentiment analysis approaches used for investigation.

3. PROPOSED WORK

3.1 Deep Bidirectional Recurrent Neural Networks-Based Sentiment Analysis (Dbrnn-Sa)

The proposed Deep Bidirectional Recurrent Neural Networks-based Sentiment Analysis (DBRNN-SA) Scheme comprises of four phases that includes, i) Building process of sentimental words, ii) Classification of sentimental words using Deep Bidirectional Recurrent Neural Networks and iii) Process of balancing the collection of words prior to the prediction algorithm execution and iv) process of prediction as depicted in Figure 1.

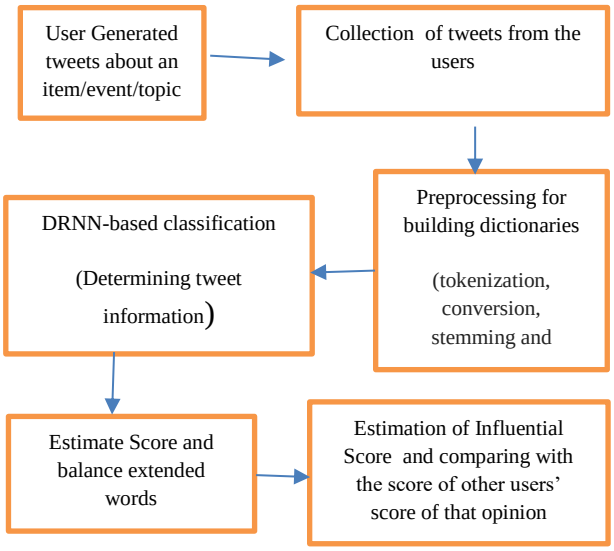

Figure 1: The system model of the proposed DBRNN-SA Scheme

Let $X_i (1 \le i \le n)$ be the collection of products, persons or services that are focused to be compared in a particular context with $T = \{X_1, X_2, ..., X_n\}$ as the concentrated targeted context.

Step 1: Building process of sentimental words

In this initial step, potential dictionaries of words are built with high orientation in words' semantics based on a specific context with the aid of a small collection of hash tags in a predominant manner [21]. The hash tags are generally utilized to identify whether the user post is positive, negative or neutral depending the description of hash tags. The majority of the word dictionary building approaches utilized a comparatively huge collection of hash tags that are manually annotated for integrating with dictionaries with the view to concentrate on the improvement of the accuracy in classification posts by public. However, the utilization of huge collection of hash tags is determined to incur a huge amount of time. Thus, small collection of hash tags is incorporated in the proposed DBRNN-SA scheme for effective and potential building of word dictionaries.

The possible sub steps involved in the process of building sentiment words are listed as follows:

Step 1: At first, the collection of posts associated to X_i are collected and stored as the major objective of this proposed DBRNN-SA scheme focuses on determining the most popular hash tags depending on the degree of high occurrence related to each and every individual X_i. Then, a comparatively small collection of hash tags (two and three in each individual category) is manually classified into positive and negative, such that data related to the positive and negative categories are gathered uniquely. In specific, a upper threshold of hash tags polarity is used for classifying the collected posts in a predominant manner.

Step 2: At this step, the set of data that are determined through classification enabled by preprocessing derived from the hash tags. This preprocessing of social data is essential, since it might be informal and may contain non-textual information and misspelled words. This process of processing includes tokenization, conversion, stemming and filtering processes. In the process of tokenization, Unicode characters, HTML tags, symbol repetition, twitter mentioned hashtags, phone numbers, URLs, nouns, verbs, adverbs, adjectives and common emoticons. Next, the process of conversion is responsible for transforming all the words into lowercase letters (for instance SUPER into super) and the replaces the continuous existence of the letter into a single frequency occurring letter words (For instance, buuunny to bunny). Then, the process of morphological stemming is imposed for eliminating conjugation and plural genders. Finally, the process of filtering is applied to improve the adjectives and verbs, since they are good indicators of positive and negative sentiment analysis. As the result, positive and negative intermediate sentimental words are derived at the end of this preprocessing step.

Step 3: This step is mainly for refining the positive, negative and neutral intermediate sentiment words that aids in building the annotated word dictionary for each of the individual X_i. In this step, the neutral hashtags are eliminated since they infer in various dimensions depending on the context of users' postings. This process of neutral biasing is applied over the word occurrence for all comprehensive set of classes. In specific, an empirical test is employed by testing a collection of values that lies between 0.4 and 0.8 with the view to determine a threshold that permits the option of

classifying the sentiment words with a highly reduced error rate. The threshold value of 0.7 is used in this proposed DBRNN-SA scheme for attaining a highly reduced error rate. Hence, the sentiment words are set to -1, 0, +1 corresponding to negative, neutral and positive sentiment words.

Phase 2: Classification of sentimental words using Deep Bidirectional Recurrent Neural Networks

In this proposed DBRNN-SA scheme, the Deep Bidirectional Recurrent Neural Networks are applied for facilitating learning in a more sequence manner that particularly concentrates on the task of natural language processing. This utilized DRNN comprise of two RNNs, in which the first one is essential for estimating the forward hidden sequences $F_{hs(1)}$ and the latter one is used for estimating the backward hidden sequences $F_{hs(2)}$. The output sequence $(Y_{O(t)})$ of the utilized DRNN corresponding to the input sequence $X_{i(t)}$ is determined at time based on Equation (1).

$$Y_{O(t)} = Act(W_{M(hs1)} * F_{hs(1)} + W_{M(hs2)} * F_{hs(2)} + O_{bias}) \quad (1)$$

Where, $Act(*)$ is an activation function with O_{bias} as the output biases. Further, the process of activations are performed in the forward and backward directions are represented through Equation (2) and (3)

$$F^{n}_{hs(1)} = f(W_{m(n-1)} * F^{n-1}_{hs(t)} + W_{m(n)} * F^{n}_{hs(t-1)} + b^{n}_{hs(1)}) \quad (2)$$

$$F^{n}_{hs(2)} = f(W_{m(n-1)} * F^{n-1}_{hs(t)} + W_{m(n)} * F^{n}_{hs(t+1)} + b^{n}_{hs(1)}) \quad (3)$$

The utilized DRNN consists of three hidden layers that are sufficient enough for facilitating effective performance in the process of learning the intermediate sentiment words.

In this context, it is notable that the language used in the social media is non conventional, since it consists of special words that are intentionally written in upper case letters or possess the repetition of more than two consecutive letters designated as extended words. Further, these extended words are considered to be poorly exploited by the method of sentiment analysis. Hence, an addition step is included in the proposed scheme for balancing and handling extended words during the process of sentimental analysis.

Phase 3: Process of balancing the collection of words prior to the prediction.

In this phase, the degree of polarity associated with each and every opinion is computed by integrating the score values of each related word in a specific instant of time with a specified message length based on Equation (4)

$$POL_{(t)} = \sum_{k=1}^{m} Ind\_Score(word) \quad (4)$$

In this process of balancing the collection of words prior to the prediction algorithm execution, the opinion tweets are classified into strongly positive, moderately positive, weakly positive, neutral, strongly negative, moderately negative and weakly negative. Further, an empirical test is performed for determining the threshold of each of the classes incorporated in the process of sentiment analysis.

Phase 4: Extracting the potential degree of sentiments

In general, most of the researchers have estimated positive, negative and neutral categories for the process of extracting the potential degree of sentiments from the document through the derivation of emotions and words [22]. However, one of the predominant approaches contributed by Kahtua et al [23] based its investigation based on the polarity degree by classifying the opinions into high, moderate, negative and weakly positive. However, they have only used the classes of strongly negative and strongly positive as decision indicators that may not be realistic in many situations. Thus, the proposed DBRNN-SA scheme assumes the dissimilarity of class influence for determining the strength of the public opinion with the rate associated with each concerned domain or topic

In this process of prediction, initially a weight is attributed with value ranging from -3 to +3 to each and every individual class into class i) strongly positive (+3), ii) moderately positive(+2), iii) weakly positive(+1), iv) neutral(0), v) strongly negative(-3) , vi) moderately negative (-2) and vii) weakly negative (-1). Then, the degree of influence attributed by each opinion is measured by including a potential metrics that aids in the balancing the weights of each opinion as defined in Equation (5)

$$OP_{\inf l} = W_{OP}(t) + (N_L(t) + N_{RT}(t)) \quad (5)$$

Where, $W_{OP(t)}$, $N_L(t)$ and $N_{RT}(t)$ is the weights associated with each class, the number of opinions that are liked and the number of opinions that are retweeted.

Finally, the overall rating related to each and every opinion is computed based on the aggregate sum of influence degree and the volume of tweets that exchange between the users about the item of interest through Equation (6)

$$X_i(Rate) = \frac{\sum_{i=1}^{k} OP_{\inf}(i)}{K_{pos-op}} \quad (6)$$

Where, K_{pos-op} is the number of positive likes concerned with an item.

In this context, the maximum rated opinion is considered the maximum likelihood of like appreciated by the internet users over an item of interest of concern.

4. METHODS AND EXPERIENCE

The support for Jallikattu is identified through the social media like Facebook and Twitter as it was widely used during the protest of Jallikatu in order to identify the support and non-support for interacting with people in order determine their opinion about the issue. In this case study, the interaction through Twitter is only considered as a micro-blogging service that permits the users for interacting and posting messages named tweets that are limited to 140 characters. The REST API is estimated to be suitable as it permits information to be widely accessible by the researchers and developers through the grant of permission for gathering published tweets associated about a concerned hashtag.

Implementation process

This implementation of the proposed scheme is facilitated over a cluster of 3 servers since this sentimental analysis-based data analysis system must have sufficient memory, potential for performing parallel processing of activities and bandwidth. The utilized servers possessed two Intel Xeon E5530 Quad core CPU 2.4 GHz processors that execute over 64-bit Linux Ubuntu Environment. Further, the servers are incorporated with 1 TB hard disk and 24 Go DDR3 RAM. Furthermore, apache Kafka is used for the process of data gathering, since it is a distributed streaming platform that utilizes the process of publishing, subscribing and messaging that ensure a replicated and distributed service option. In particular, an integrated stream of API library is utilized for allowing the process of constructing

applications that are suitable for the processing of the stream in an effective manner. In addition, huge amount of gathering data is stored in the Hadoop Distributed File System (HDFS). The Spark is employed for the processing of data that aids in constructing dictionaries and data classification. The employment of Spark permits fault tolerant and scalable processing of streams of data under subsequent real time. In this case, the spark is incorporated over the Hadoop with management through YARN for the processing distribution over the 3 included servers. The implementation of the deep recurrent neural network is facilitated through the benefits of Spark MLlibs for the need to compare the predominance of the proposed scheme with the other existing sentimental approaches. The complete set of algorithm that are used for the process of gathering data and processing is implemented through Python. Finally, the deep recurrent neural network modeled through Spark MLlibs is applied for the retrieval of tweets in real time.

Data Collection and Processing

Phase 1: In this implementation, the only two possible keywords such as support and non- support are used for retrieving the tweets that are categorized into $P = \{x_1, x_2\}$ with $x_1 = \text{sup}\ port$ and $x_2 = non - \text{sup}\ port$. The data are gathered with the use of the Twitter REST API under real time with tweets compulsorily written in English. The selection of most relevant hashtags is determined from the word frequency list in order to estimate the most popular hashtags for the complete gathered data. Then, the classification of tweets are facilitated as follows: **Positive support for Jallikatu**: #tamizhan identity, #tamizhan history, # tamizhan gethu,# tamizhan super, **Negative support for Jallikatu:** #jallikatu notsafe, #risky jallikatu, #lifesucking jallikatu, # dangerous jallikatu. The process of validating the construction process of dynamic dictionary, the data are collected associated to the aforementioned hashtags over the dates of December 21 and 22 2016. The prototype of data comprises of 1,30,000 tweets that are partitioned into 65,000 tweets for positive support for Jallikatu and 65,000 tweets for negative support for Jallikatu.

Phase 2: The prediction about the support to Jallikattu is collected from all the twitter messages that are posted over the dates of 21 and 22 December 2016 that comprises an amount of 32,00,000 tweets for both the

support and non support to the issue of analysis.

Phase 3: The polarity degree associated with each of the tweets is computed for determining the degree of influence and weight in the forthcoming step.

Phase4: This phase utilized the results of the data that are processed in the previous stage with tweets information that includes the weight of the class, the labeled class for which the tweet pertains to the count of likes and retweets.

In the first phase, the positive and negative dictionaries are constructed for the subject of support and non support to Jallikatu. Then, the data are preprocessed by eliminating the duplicated tweets and stop words through the application of filters. In the second phase of classification through DRNN, the extended words are handled with further classification of tweet words determined through the process of training. Then, the degree of polarity is determined in the third phase. Finally, the degree of influence and the overall rating that determines the support and non support to the issue of investigation.

5. RESULTS AND DISCUSSIONS

The simulation experiments and result investigation of the proposed DBRNN-SA Scheme are conducted by setting the parameters like the vector size, vocabulary size and the number of hidden layers, filter counts, size of the filters, drop out, regulariser and activation functions are detailed in Table 1. (The simulation parameters are set based on analyzing the other research works conducted in this area [24-25])

Table 1:Simulation Parameters set for experimenting the proposed DBRNN-SA scheme

Simulation Parameters	Parameter Value
Size of the input vector	100
Size of the vocabulary	13, 398
Size of the filters	3, 4, 7
Number of filters	10,40,70,100,`28,256
Dimension of the output	128
The value of the regularizer	1.3

The value of drop out	0.4
Size of the batch	64
Epochs count	5
Number of hidden layers	4,5
Number of recurrent layers	2,3

In the first part of the investigation, the predominance of the proposed DBRNN-SA Scheme as an effective method of performing sentimental analysis over the big data is evaluated using accuracy and loss curve under increasing iterations of training. Figure 2 exemplars the accuracy of the proposed DBRNN-SA Scheme over CNN-DLSA, AC-CNN-SA and TSA-DNNM under a systematic increase in the number of iterations used for training. The accuracy of the proposed DBRNN-SA Scheme is confirmed to be sustained over the benchmarked sentiment analysis schemes due to the incorporation of deep recurrent neural networks used for the purpose of classification. Thus, the accuracy of the proposed DBRNN-SA Scheme is enhanced by 12%, 10% and 7% superior to the compared CNN-DLSA, AC-CNN-SA and TSA-DNNM schemes used for analysis.

Likewise, Figure 3 depicts the loss curve of the proposed DBRNN-SA Scheme over CNN-DLSA, AC-CNN-SA and TSA-DNNM under a systematic increase in the number of iterations used for training. The loss error of the proposed DBRNN-SA scheme is confirmed to be reduced compared on par with the benchmarked sentiment analysis schemes due to inheriting behavior of stemming process in the stage of preprocessing. The loss error of the proposed DBRNN-SA Scheme is minimized by 14%, 11% and 8% superior to the compared CNN-DLSA, AC-CNN-SA and TSA-DNNM schemes used for analysis.

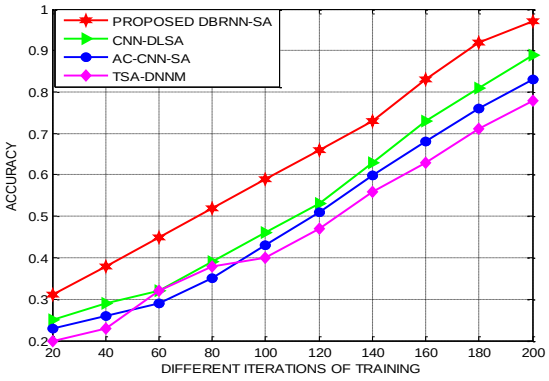

Figure 2: Accuracy of the proposed DBRNN-SA Scheme-varying iterations of training

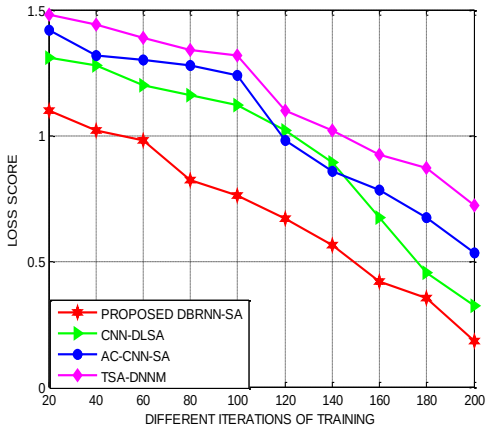

Figure 3: Loss Curve of the proposed DBRNN-SA Scheme-varying iterations of training

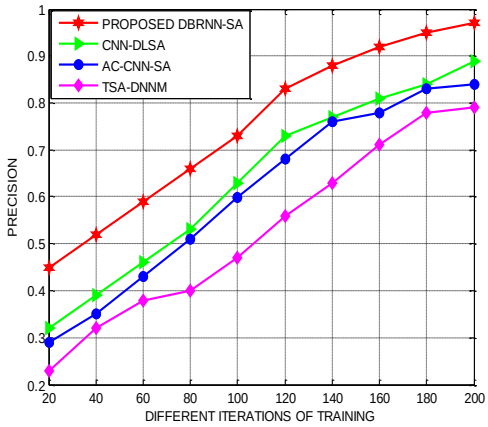

Figure 4: Precision of the proposed DBRNN-SA Scheme-varying iterations of training

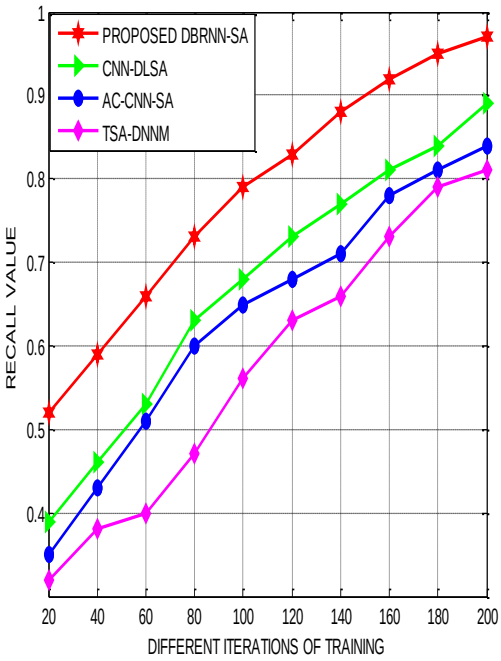

Figure 5: Recall value of the proposed DBRNN-SA Scheme-varying iterations of training

In the second part of the investigation, the predominance of the proposed DBRNN-SA scheme is quantified using precision and recall under increasing iterations of training. Figure 4 glorifies precision of the proposed DBRNN-SA Scheme over CNN-DLSA, AC-CNN-SA and TSA-DNNM under a systematic increase in the number of iterations used for training. The precision of the proposed DBRNN-SA Scheme is estimated to be balanced by a monotonic increase in the incorporated training iterations, since it inherits the significance of preprocessing for facilitating the option of

sentiment analysis over big data. The precision of the proposed DBRNN-SA Scheme is enhanced by 12%, 10% and 7% superior to the compared CNN-DLSA, AC-CNN-SA and TSA-DNNM schemes used for analysis. Likewise, Figure 5 depicts the recall of the proposed DBRNN-SA Scheme over CNN-DLSA, AC-CNN-SA and TSA-DNNM under a systematic increase in the number of iterations used for training. The improvement in the degree in recall of the proposed DBRNN-SA scheme is confirmed, since it explores the maximum number of dependent and independent factors that contribute towards sentiment analysis. The recall of the proposed DBRNN-SA Scheme is minimized by 13%, 11% and 9% superior to the compared CNN-DLSA, AC-CNN-SA and TSA-DNNM schemes used for analysis.

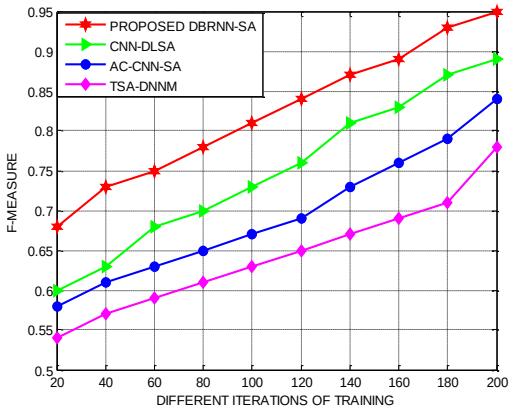

Figure 6: F-Measure of the proposed DBRNN-SA Scheme-varying iterations of training

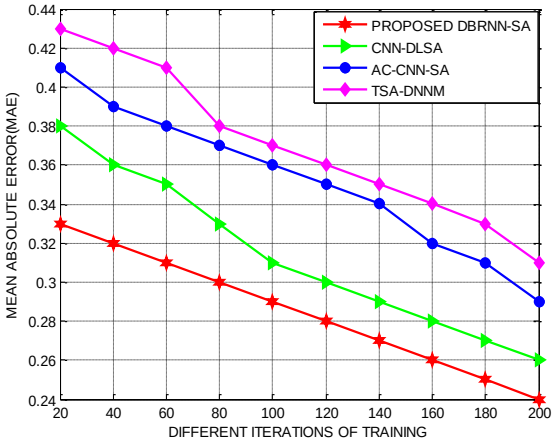

Figure 7: MAE of the proposed DBRNN-SA Scheme-varying iterations of training

In the third part of the investigation, the predominance of the proposed DBRNN-SA scheme is quantified using F-Measure and Mean Absolute Error (MAE) under increasing iterations of training. Figure 6 glorifies F-Measure of the proposed DBRNN-SA Scheme over CNN-DLSA, AC-CNN-SA and TSA-DNNM under a systematic increase in the number of iterations used for training. The F-Measure of the proposed DBRNN-SA Scheme is confirmed to be maintained independent to the number of incorporating training iterations, since it derives the merits of balancing the collection of words prior to the prediction algorithm execution. The F-Measure of the proposed DBRNN-SA Scheme is enhanced by 13%, 11% and 8% superior to the compared CNN-DLSA, AC-

CNN-SA and TSA-DNNM schemes used for analysis. Likewise, Figure 7 depicts the MAE of the proposed DBRNN-SA Scheme over CNN-DLSA, AC-CNN-SA and TSA-DNNM under a systematic increase in the number of iterations used for training. The reduction in the degree in the MAE of the proposed DBRNN-SA scheme is confirmed by investigating multiple linguistic features that are more common in big data documents. The MAE of the proposed DBRNN-SA Scheme is minimized by 11%, 8% and 6% superior to the compared CNN-DLSA, AC-CNN-SA and TSA-DNNM schemes used for analysis.

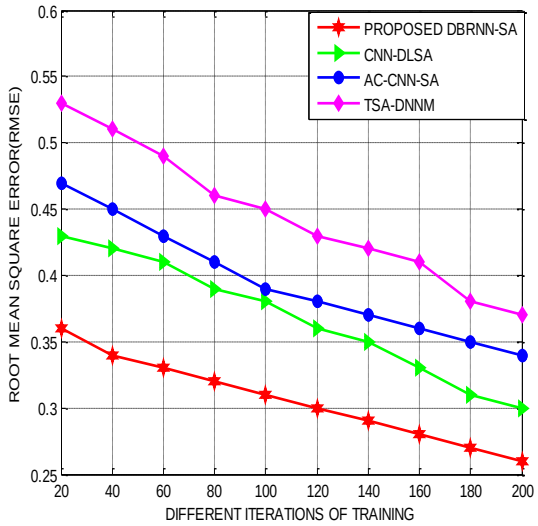

Figure 8: RMSE of the proposed DBRNN-SA Scheme under varying iterations of training

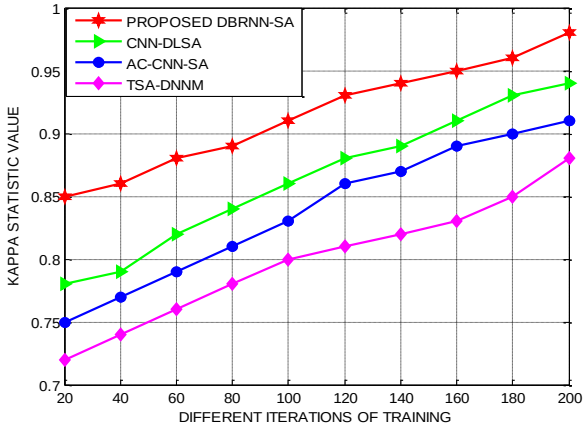

Figure 9: Kappa Statistic-the proposed DBRNN-SA Scheme-varying iterations of training

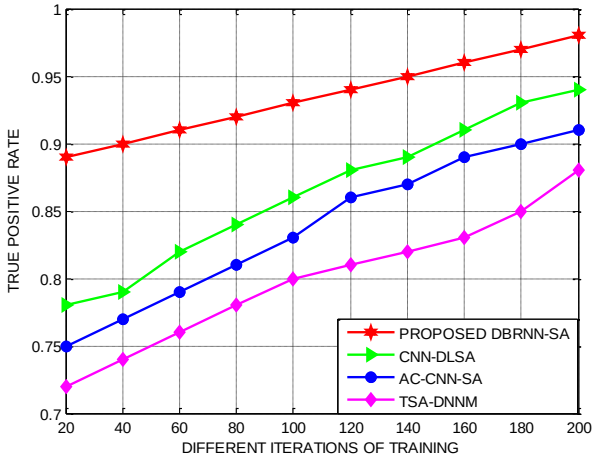

Figure 10: True positive rate-proposed DBRNN-SA Scheme-varying iterations of training

In addition, the significance of the proposed DBRNN-SA Scheme is investigated using Root mean Square Error, Kappa statistic and True positive rate under different iterations of training. Figure 8

glorifies the predominance of the proposed DBRNN-SA Scheme evaluated using RMSE with a corresponding increase in the number of iterations used for training. The RMSE value of the proposed DBRNN-SA Scheme remains reduced even under monotonic increase in the number of iterations used for training, since they inherit. The RMSE of the proposed DBRNN-SA Scheme is minimized by 15%, 13% and 10% compared to the baseline sentiment analysis scheme used for benchmarking. Figure 9 exemplars using Kappa statistic with a corresponding increase in the number of iterations used for training. The kappa value of the proposed DBRNN-SA Scheme remains higher even under monotonic increase in the number of iterations used for training, since they utilize the The kappa value of the proposed DBRNN-SA Scheme is improved by 14%, 11% and 10% compared to the baseline sentiment analysis scheme used for benchmarking. Figure 10 highlights the true positive rate with a corresponding increase in the number of iterations used for training. The True positive rate of the proposed DBRNN-SA Scheme remains higher even under monotonic increase in the number of iterations used for training, since they utilize the. The True positive rate of the proposed DBRNN-SA Scheme is improved by 16%, 13% and 11% compared to the baseline sentiment analysis scheme used for benchmarking.

6. CONCLUSION

The proposed DBRNN-SA Scheme is contributed as a vital deep recurrent neural network-based learning mechanism for potential prediction of peoples' behavior through the analysis of big social data. The proposed DBRNN-SA scheme first built dictionary of words using the polarity using a very small collection of hash tags that are positive, negative associated with the contextual subject. The proposed DBRNN-SA scheme incorporated DRNN for an effective classification process such that the extended words of the vocabulary can be balanced for effective classification of opinion tweets. The simulation results of the proposed DBRNN-SA Scheme proved an enhanced rate of 11% and 13% in kappa statistics and true positive rate under an increasing number of training iterations. As a part of future work, it is planned to formulate an integrated CNN and LSTM-based sentiment analysis method for big data with multiple linguistic parameters considered for facilitating the process of opinion mining.

8.REFERENCES

[1] Kalra, V., & Agrawal, R. (2019). Challenges of Text Analytics in Opinion Mining. Advances in Data Mining and Database Management, 1(2), 268-282.

[2] Ramteke, J., Shah, S., Godhia, D., & Shaikh, A. (2016). Election result prediction using Twitter sentiment analysis. 2016 International Conference on Inventive Computation Technologies (ICICT), 1(1), 45-56.

[3] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 67-78.

[4] Wang, H., & Castanon, J. A. (2015). Sentiment expression via emoticons on social media. 2015 IEEE International Conference on Big Data (Big Data), 1(1), 78-86.

[5] Ghosh, M., & Sanyal, G. (2018). An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. Journal of Big Data, 5(1), 12-23.

[6] Singh, P. (2018). Sentiment Analysis Using Tuned Ensemble Machine Learning Approach. Advances in Data and Information Sciences, 1(1), 287-297.

[7] Shrote, K. R., & Deorankar, A. (2016). Review based service recommendation for big data. 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 1(2), 34-43.

[8] Nhlabano, V., & Lutu, P. (2018). Impact of Text Pre-Processing on the Performance of Sentiment Analysis Models for Social Media Data. 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), 1(2), 65-74.

[9] Patil, M., & Darokar, M. S. (2018). A Supervised Joint Topic Modeling Process Using Sentiment Analysis. Journal of Advances and Scholarly Researches in Allied Education, 15(2), 720-725.

[10] Kaur, G. (2018). Text Mining Based Approach to Customer Sentiment Analysis Using Machine Learning. Journal of Advances and Scholarly Researches in Allied Education, 15(6), 58-65.

[11] Wehrmann, J., Becker, W., Cagnini, H. E., & Barros, R. C. (2017). A character-based convolutional neural network for language-agnostic Twitter sentiment analysis. 2017 International Joint

Conference on Neural Networks (IJCNN), 1(1), 56-65.

[12] Yazhi Gao, Rong, W., Shen, Y., & Xiong, Z. (2016). Convolutional Neural Network based sentiment analysis using Adaboost combination. 2016 International Joint Conference on Neural Networks (IJCNN), 1(1), 56-62.

[13] Tromp, E., Pechenizkiy, M., & Gaber, M. M. (2017). Expressive modeling for trusted big data analytics: techniques and applications in sentiment analysis. Big Data Analytics, 2(1), 45-56.

[14] Karpurapu, B. S., & Jololian, L. (2017). A Framework for Social Network Sentiment Analysis Using Big Data Analytics. Big Data and Visual Analytics, 1(2), 203-217.

[15] Salas-Zárate, M. D., Valencia-García, R., Ruiz-Martínez, A., & Colomo-Palacios, R. (2016). Feature-based opinion mining in financial news: An ontology-driven approach. Journal of Information Science, 43(4), 458-479.

[16] Mars, A., & Gouider, M. S. (2017). Big data analysis to Features Opinions Extraction of customer. Procedia Computer Science, 112, 906-916.

[17] Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. Journal of Big Data, 5(1), 78-89.

[18] El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., & Kobi, A. (2018). A novel adaptable approach for sentiment analysis on big social data. Journal of Big Data, 5(1), 78-89.

[19] Rani, S., & Kumar, P. (2018). Deep Learning Based Sentiment Analysis Using Convolution Neural Network. Arabian Journal for Science and Engineering, 1(1), 24-34.

[20] Chen, S., Peng, C., Cai, L., & Guo, L. (2018). A Deep Neural Network Model for Target-based Sentiment Analysis. 2018 International Joint Conference on Neural Networks (IJCNN), 1(1), 34-49.

[21] Kwabla, S., Kwame, N., & Katsriku, F. (2017). Sentiment Analysis of Twitter Feeds using Machine Learning, Effect of Feature Hash Bit Size. Communications on Applied Electronics, 6(9), 16-21.

[22] Wang, H., & Castanon, J. A. (2015). Sentiment expression via emoticons on social media. 2015 IEEE International Conference on Big Data (Big Data), 1(1), 12-22.

[23] Khatua, A., Khatua, A., Ghosh, K., & Chaki, N. (2015). Can #Twitter_Trends Predict Election Results? Evidence from 2014 Indian General Election. 2015 48th Hawaii International Conference on System Sciences, 1(1), 67-76.

[24] Nair, A., & Bhajani, M. (2016). Product Aspect Ranking Using Sentimental Analysis. International Journal of scientific research and management, 1(2), 15-27.

[25] Shyamasundar, L. B., & Rani, P. J. (2016). Twitter sentiment analysis with different feature extractors and dimensionality reduction using supervised learning algorithms. 2016 IEEE Annual India Conference (INDICON), 1(1), 34-45.