OPINION MINING AND EVENT DETECTION ANALYSIS OF CORONAVIRUS TWITTER DATA USING ENSEMBLE DEEP LEARNING MODELS

Supported by tetfund TERTIARY EDUCATION TRUST FUND

Abdullah, Khadijha-Kuburat Adebisi*, Sodimu Segun Michael, Odule Tola John, Efuwape Biodun Tajudeen and Solanke Olakunle O.

Department of Mathematical Sciences, Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria

*Corresponding author: abdullah.adebisi@oouagoiwoye.edu.ng

Received: August 18, 2021 **Accepted:** September 27, 2021

Abstract: Internet technology has grown to the extent that people create, share ideal, opinion and content on Twitter. Useful information is obtained from Twitter on Coronavirus (COVID-19) for response and management of the pandemic. The classification of Twitter is affected by addressing negation for sentiment analysis and domain dependence for Twitter events or topic. In this paper, ensemble deep learning models are used to learn and train network on real world COVID-19 Twitter data in analysing and provide a broad classification to user's opinion and events discuss in multiclass and multilabel, respectively. A Textblob lexicon and Natural Language ToolKit (NLTK) are employed for polarity sentiment and event discuss for categorisation and event similarity respectively, this reduce the bias in the network and enhanced models' performance. A comparative analysis is performed on Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and C-LSTM with pretrained distributed GloVe embedding. The classification is performed in batches to reduce the complexity and memory usage, however, dropout and early stopping strategies are employed to prevent overfitting with Adaptive Momentum (ADAM). Evaluation of the models are performed using precision, recall, F1 score and accuracy with 90%, however, LSTM shows more accuracy compared to CNN and C-LSTM.

Keywords: Multiclass, multilabel, COVID-19, polarity, event discuss

Introduction

The outbreaks of Coronavirus (COVID-19) are detected and announced by World Health Organization (WHO) has a global emergency respiratory syndrome that originated in a sea food in Wuhan, China at the end of December 2019 (Sohrabi *et al*., (2020). Almost all countries in the world have been affected with Coronavirus, this pandemic according to WHO is an International concern as a Public Health Emergency control (WHO, 2020). Most research on COVID-19 are basically on Epidemiology, causes, clinical and X-Ray diagnosis, environmental issues, vaccine development, preventions, and control (Francesco *et al*., 2020) etc. but fewer research have been done on social media data. Palen and Hughes, (2018) presented that social media data is useful for emergency and crisis as communication tool and real-time information for situational awareness. However, the roles of Twitter during crisis outbreak has been enormous. People post on Twitter not for communication but to share information on updates regarding the pandemic such as requesting for help, report on infected people, precautions, preventions, and other useful information (Vieweg *et al*., 2014). According to Ahmad *et al*. (2019), WHO posted information of COVID-19 on Twitter, extracting useful and meaningful information from people' tweets have attracted the attention of a very large number of researchers (Kwak *et al*., 2010). Most research on COVID-19 using social media were based on 5G networks (Lewis & Full, 2020; Shanapinda, 2020). In this era of massive information, analysing hidden relationships in Twitter data is time consuming as well as difficult to process and classify due to short and unstructured text. Extracting meaningful data and patterns reflect the knowledge to make decisions (Brindha *et al*., 2016). But as number of Twitter data increases, the computational complexity also increased (Stas *et al*., 2014).

However, users' tweet on emotions and event discuss on the pandemic have not been fully identified on COVID-19. As the pandemic is increasing with the second wave, it requires planning and methods of processing information for opinion mining (sentiment analysis) as well as classification of the tweets' event discuss. Sentiment analysis identifies and analyses subjective information on opinion mining on Twitter data (Pham & Le 2016; Yang *et al*., 2017). While event domain classification captures local (syntactic) and global (semantic) dependencies by identifying different kinds of topic or events discuss and useful information posted on Twitter. This information cannot be achieved by only keywords 'Coronavirus or COVID-19' because of its limited utilization on people' tweets. Despite advances in natural language processing (NLP) techniques, automatic classification processing of short texts (40 characters) and informal Twitter messages are task challenging. Interpreting the semantics of short informal texts from Twitter automatically pose problem hence, difficult to understand without enough context. To address these problems, most existing methods for expressing the representation of words on social media use neural language models which is based on word embedding. Distributed representation of word embedding used Word2Vec (Mikolov *et al*., 2013) and/or GloVe (Pennington *et al*., 2014) in generating embeddings by converting words into meaningful vectors. This involves vast amounts of big data that correspond to good learning systems. Traditional methods of classifying Twitter data are based zon shallow learning such as Naıve Bayes (NB) (Diab & Hindi, 2017; Xu, 2018), Support Vector Machine (SVM) (Lilleberg *et al*., 2017), K-Nearest Neighbour (K-NN) (Palen & Hughes, 2018) with advantages in accuracy and stability. Due to data sparseness and high dimensionality, traditional classification methods depend on the word frequency or word co-occurrences, hence, requires more feature selections to solve contextual problem.

Recently, existing textual data classification research are based on Deep Neural Network (DNN) in solving big analytic data such as Twitter. This improves and provides efficient accuracy with low computational complexity. Though, deep learning models require a large-scale corpus to manage parameters, therefore, require large memory for processing embedding from scratch. Therefore, in this study, a pretrained Glove word embedding is adopted and fed into the classifiers to increase the accuracy and performance for analysing COVID-19 Twitter data. The ensemble DNN jointly learns feature extraction automatically without domain knowledge for text classification models (Kim, 2014; Shen *et al*., 2014). Furthermore, features map captures local context and global statistical features, train on the non-zero elements in a cooccurrence matrix. This serves as inputs to DNN models to learn words, sub-words, phrases, and their relationship. In this work, the classification models involves Convolutional Neural Networks (CNN) (Shen *et al.*, 2014), modified Recurrent Neural Networks (RNN); Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) as well as hybrid models of CNN and LSTM (C-LSTM) for effective classification. The classification is performed in batches to reduce the complexity and memory usage to learn set of non-linear representations. These models have parameters such as learning rate, batches of data used, dropout, early stopping, epoch etc. (Abdi *et al.*, 2019; Kowsari *et al.*, 2017). The CNN models learn, recognise, and detect pattern across space, thus, improved state-of-art on events or topic classification (Wallace *et al.*, 2014) and sentiment analysis (LeCun *et al.*, 1998; Shen *et al.*, 2014). While, RNN model captures word dependency and structures of text with gradient vanishing problem (Bengio *et al.*, 1994), therefore, LSTM (modified RNN) is used to alleviates and handled gradient vanishing problem by introducing a memory cell to remember and capture long term dependencies in tweets over time ((Hochreiter & Schmidhuber, 1997).

Intuitively, sentiment requires categorisation of emotions and intension for detection of subjective and objective of factual based text before validation (Mukherjee & Bhattacharyya, 2013). A Textblob lexicon is employed to detect part-of-speech tagging, sentiment analysis, noun phrase extraction, translation as well as classification (Loria, 2019). This is used to discover context-based sentiments into positive, negative, and neutral. Saha *et al.* (2017) built a sentiment analysis using Textblob and validated by SVM and Naive Bayes classifier for polarity confidence calculation with Twitter content. In events modelling, multi-label classification problems involve tweet manipulation to interpret the class label correctly by transforming the single tweet into a multiple class using Natural Language ToolKit (NLTK) similarity task. This adapts the algorithm to train selected batches of tweets as a set of size k using different classifiers models. In this study, multiclass and multilabel methods of classification are adopted for sentiment analysis and event tweets on Coronavirus using SoftMax and Sigmoid output, respectively.

The sentiments are categorised into multiclass such that the class labels are given as a set of $y = \{x_i\}$, $y \in$ {very positive, positive, neutral, very negative and negative}. Consequently, multilabel events involve binary class labels with zero or more labels for each input tweet (events domain). The class labels are given as a set of $y = \{x_i\}$with training sample (x_i) and label $y \in \{1, \cdots k\}$ such that k is number of estimated probabilities $y \in \{0,1\}^j$ with data point $x \in R^d$.

The models have capacity to learn complex decision functions for large features but overfit. To obtain optimal performances of the classifiers, dropout regularisation function (Srivastava *et al.*, 2014) and early stopping (Prechelt, 2012) strategies are used to enhance the network. All trainable parameters are fine-tune from hidden layers to output layer by minimising loss function with cross-entropy during learning process as well as optimising the momentum of the loss with ADAM optimizer (Kingma *et al.*, 2014). The contributions are:

(i) We adopted Textblob to discover and distribute context emotion of users' tweet and intensity of their sentiments and NLTK similarity algorithms to distribute users' events on COVID-19.

(ii) We proposed CNN, LSTM and C-LSTM classifiers to validate users' opinion mining and events discuss on COVID-19 tweets into multiclass and multi-label respectively to understand user's view across the globe.

(iii) The proposed models' parameters are fine-tune with regularisation using dropout to prevent over fitting as well as early stopping by optimizing the loss function with Adaptive Momentum Optimization (ADAM) to have better performance.

Materials and Methods

The development and network structures of deep learning models are introduced in this section using CNN, LSTM and C-LSTM to classify tweets on COVID-19 by extracting meaningful information to understand the opinion mining and events discuss of users in order to improve decision making on the pandemic as shown in Fig. 1.

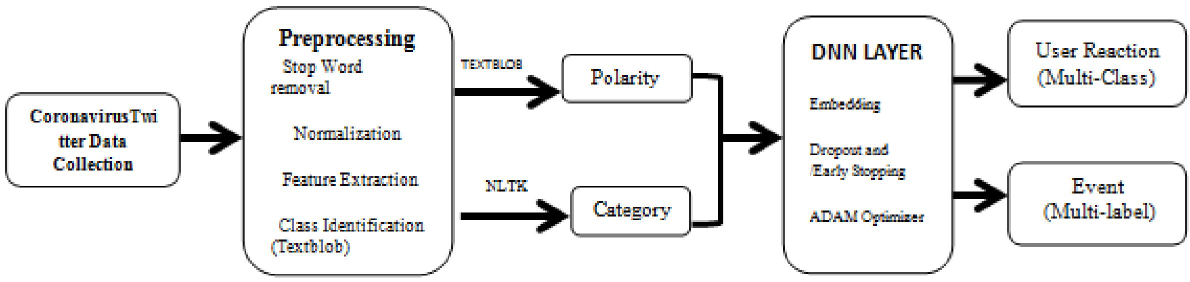

Fig. 1: The approach methodology of opinion mining and event domain

Classification models

Convolutional neural network (CNN)

Apparently, features generated required a fixed-length representation from embedding layer where the short length are padded with zero padding. In the feed forward network structure, the pattern extraction of CNN consists of three layers which include convolutional layer, max-pooling layer, and fully connected layer. The network combines the input generated from matrix vector sequence $A \in \Re^{m \times n}$ to generate feature maps, (where m and n the sizes of the matrix) and used max-pooling of the feature map to have features however, the layer composes of filer $c \in \Re^{m \times n}$ $(c = 32)$ corresponding kernel size s=24 one step at a time. The fully connected layer uses hidden layers of 128 neurons to improve the interaction between input layer and output layer to generate new weighted (W) matrices. Then, Rectified Linear Unit (ReLU) is added to the network to capture complex relationship. Therefore, for parameters, W x, b is given as:

$$f(W, x) = f(W^T x + b) \qquad (2)$$

W = weighted matrix (randomly initialised) for each layer; x = input vector; b = bias (constant)

Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) units are collections vector $x \in R^d$ with d the LSTM unit number each time step *t*. The LSTM adjusts the information in the memory cell state (c_t) with three (3) gates; input gate (i_t), forget gate (f_t) and output gate (o_t) which control the non-linear activation function (δ). Each LSTM cell consists of outputs hidden state (h) for each input tweet while the current state input (x) and previous hidden state (hs_{t-1}) is used by the 3 gates (i_t), (f_t) and (o_t). The input gate determines new information in the memory cell, the output gate controls amount of information in the internal memory cell and forget gate remember information at arbitrary interval. The LSTM unit can be represented for COVID-19 classification model is as follows;

$$i_t = \sigma(W_i x_t + U_i hs_{t-1} + b_i) \qquad (3)$$

$$f_t = \sigma(W_f x_t + U_f hs_{t-1} + b_f) \qquad (4)$$

$$o_t = \sigma(W_0 x_t + U_o hs_{t-1} + b_o) \qquad (5)$$

$$\hat{\tau}_t = tanh(W_{\hat{\tau}} x_t + U_{\hat{\tau}} hs_{t-1} + b_{\hat{\tau}}) \quad (6)$$

$$\tau_t = f_t^o \tau_{t-1} + i_t^o \hat{\tau}_t \qquad (7)$$

Convolutional long short-term memory (C-LSTM)

The C-LSTM model aligned CNN and LSTM to learn sequential features word lack in CNN with LSTM.at varying time step. This is achieved by passing the output of CNN into the LSTM layer to learn higher sequence representation as in sequence of CNN and LSTM equations above while the sequence order is preserved.

In the backpropagation, the parameters are updated to adjust the networks. The loss function (ℓ) adjusts the network weights to create better fitted network (minimising loss). The network is performed on training dataset and the decision boundaries of the output layers are determined for COVID-19 classification which indicates the possible probability labels. These probabilities are compared to the target labels and loss function is calculated using weighted cross entropy as equation (8) and (9) respectively for sentiment analysis and events discuss classification

Classification

The classification is in two forms: A SoftMax and Sigmoid functions are used for multiclass and multi-label classification, respectively.

(i) In Multiclass classification, a Softmax function (ρ), $\rho_n(x) = \frac{e^{x_n}}{\sum_{k=1}^{K} e^{x_k}}$ model for the generalization $(K > 2)$ classes for opinion mining (sentiment). This can also be minimise by categorical cross-entropy loss function (ℓ_{ce}) given by equation (8), *one-hot* encoding is used (very positive, positive, neutral, very negative, negative) for the target label y_n:

$$\ell_{ce} = -\frac{1}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} y_n^k \log(\rho_n(x_n, K)) \qquad (8)$$

(ii) A Sigmoid function (γ), $\gamma(x) = \frac{1}{1+e^{-x}}$ model probabilities on events classification extracted on COVID-19 is in binary form. This indicates $y = \{0,1\}^j$ and is minimize by binary cross-entropy loss function (ℓ_{be}) as in equation (9) with multi-label many (*k*) versus all max-entropy loss, suited for multi-label tasks.

$$\ell_{be} = -\frac{1}{N}\sum_{n=1}^{N}[\log(\gamma(x_n))y_n + \log(1-\gamma(x_n))(1-y_n)] \qquad (9)$$

However, Adaptive Momentum (ADAM) is being used as optimizer on equation 8 and 9 to improve accuracy, then, updates the parameter weights iteratively via back propagation as in equation (10):

$$\omega_{t+1} = \omega_t + \Delta\omega_t \qquad (10)$$

Instantaneous changes in parameter weight (as depicted in equation 11) with respect to time, affect the learning rate of the model (η):

$$\Delta\omega_t = -\eta \frac{\theta_t}{\sqrt{\bar{\theta}_t + \varepsilon}} \times \nabla_t \qquad (11)$$

$$\theta_t = \alpha_1 * \theta_{t-1} - (1-\alpha_1) * \nabla_t$$

$$\bar{\theta}_t = \alpha_2 * \bar{\theta}_{t-1} - (1-\alpha_2) * \nabla_t^2$$

Where η = learning rate; ∇_t = gradient with time t; θ_t= average gradient; $\bar{\theta}_t$= average of squares of gradient; α_1, α_2 are the hyper-parameters

Results and Discussion

In this experiment, we classify Coronavirus (COVID-19) ‘tweets into multiclass and multi-label representation using four hundred thousand (400,000) dataset sample size for each class due to size of the hardware. The tweets are pre-processed to removes top word, non-alphanumeric character and finally normalisation is performed. The distributions of the COVID-19 dataset for sentiment and event are performed with Textblob and NLTK similarity respectively for generating test data before training in Fig. 2(a and b).

The experiments are performed in batches of 50,000, 100,000, 200,000, 300,000 and 400,000 using ensemble models of CNN, LSTM and C-LSTM so that convergence can be much faster on the datasets using Python programming language and other packages. The code repository for the paper can be found in http://github.com/princesegzy01/ensemblesdeeplearning-multilabek-covid-19. The classifiers (CNN, LSTM and C-. LSTM) are trained and evaluated on cross validated set of 90:10 ratio. The hidden layer is computed with ReLU function of 128 hidden layers; hence, weight of each layer is regularised by dropout. A dropout of 0.2 is used to randomly switch off some neuron to resolves the problem of over fitting (co-adaptation). The Models are trained using cross entropy as a loss function, and hyper-parameter tuning of 20 epochs, learning rate of η = 0:001, optimisation is performed with ADAM. While early stopping stops the training after 3 consecutive epochs without improvement on the loss.

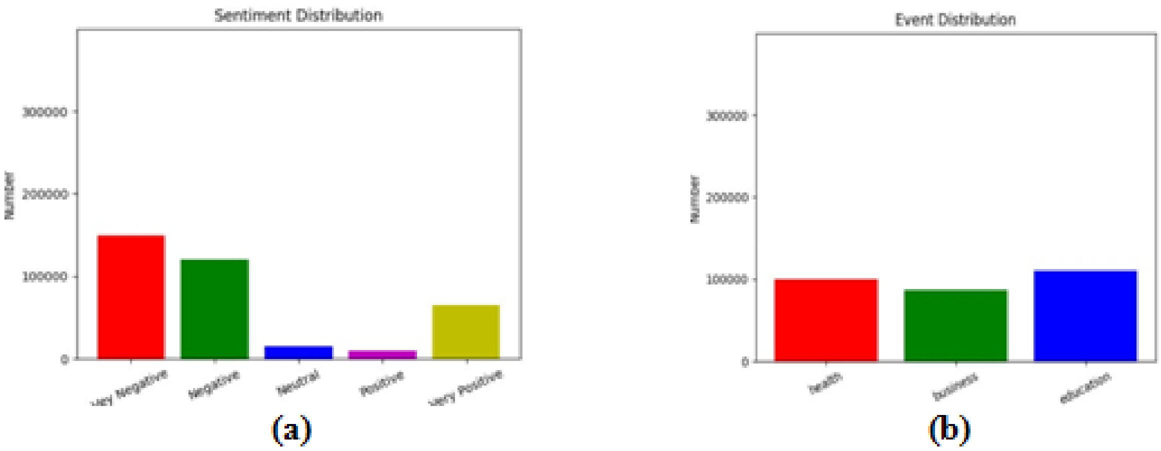

Fig. 2(a and b): COVID-19 dataset classification distribution for sentiment (polarity) and event tweet

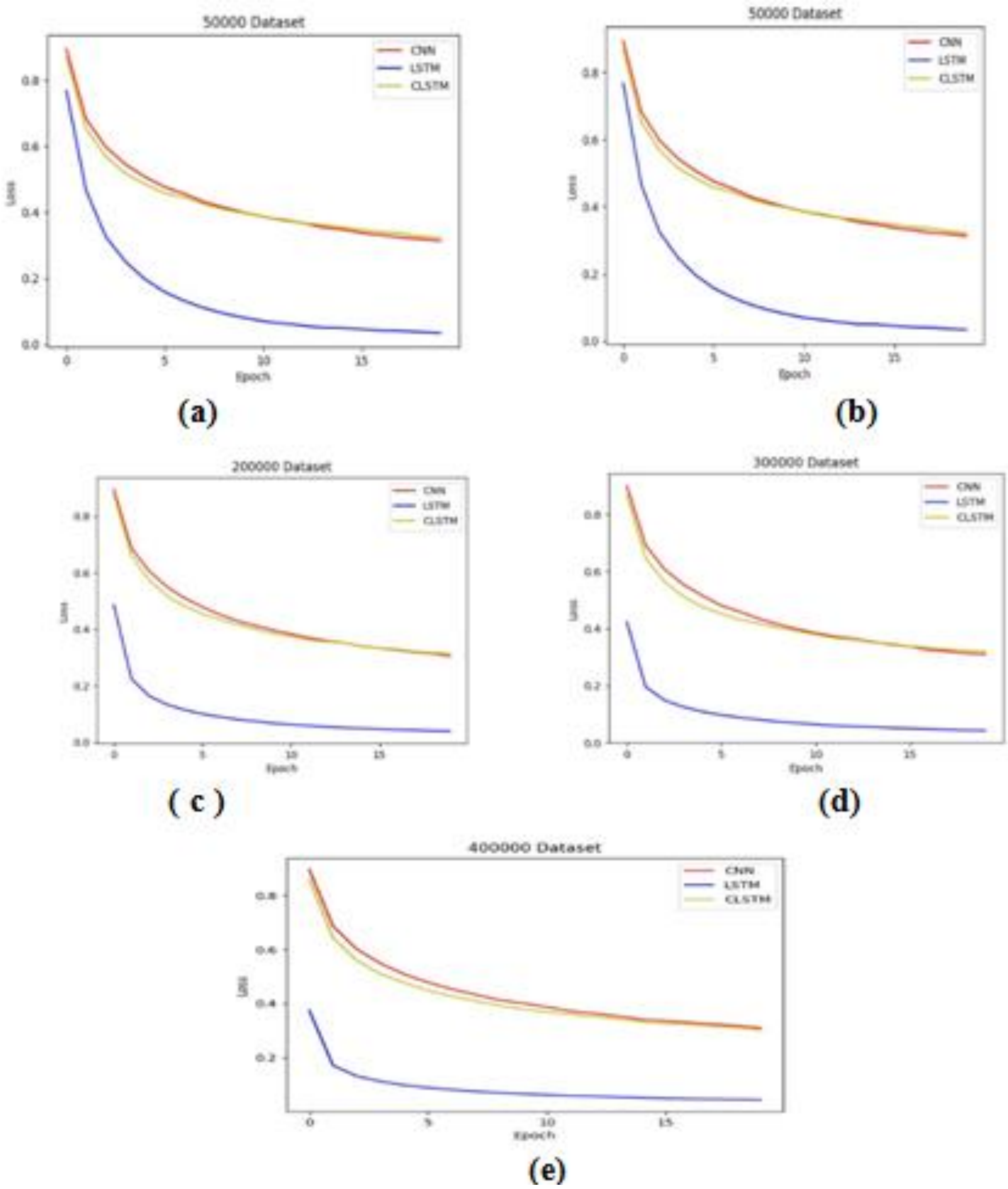

Fig. 3: Shows the loss for dataset in batches of 50000, 100000, 200000, 300000, 400000

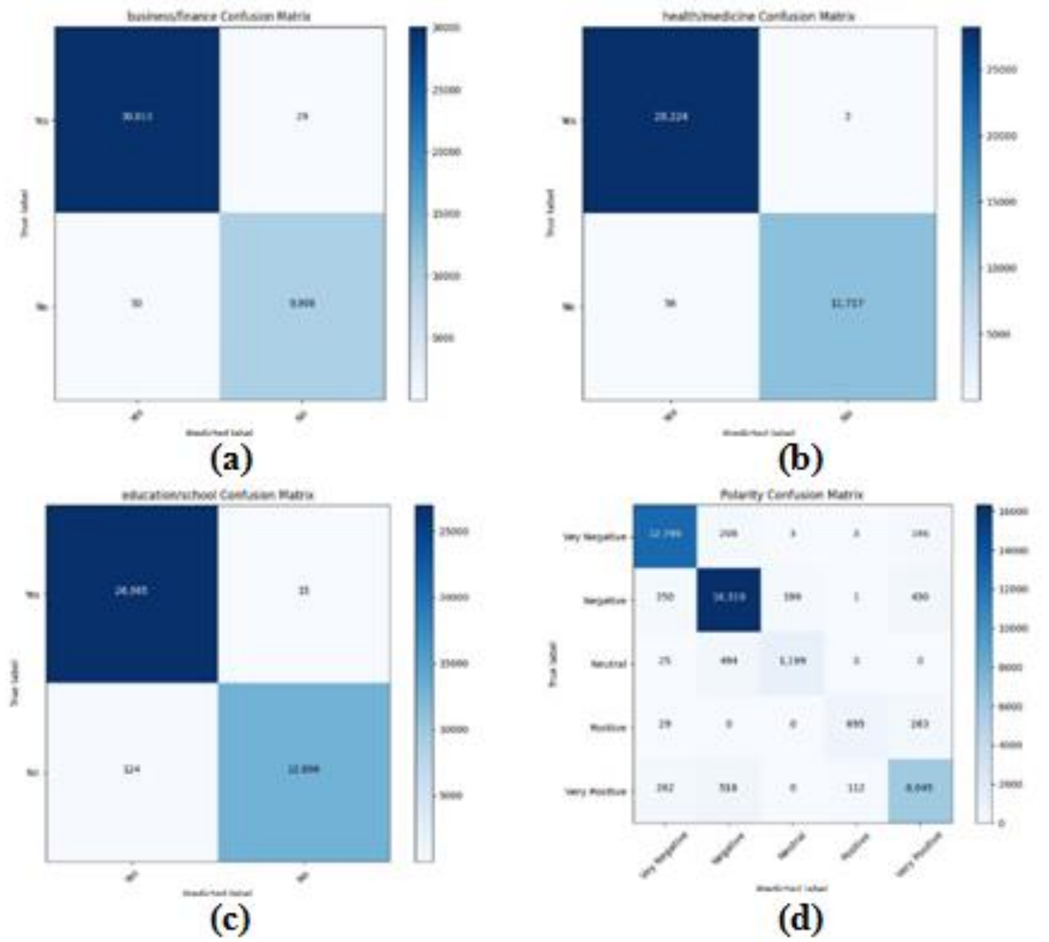

Fig. 4 (a-d): Confusion matrix for events and polarity

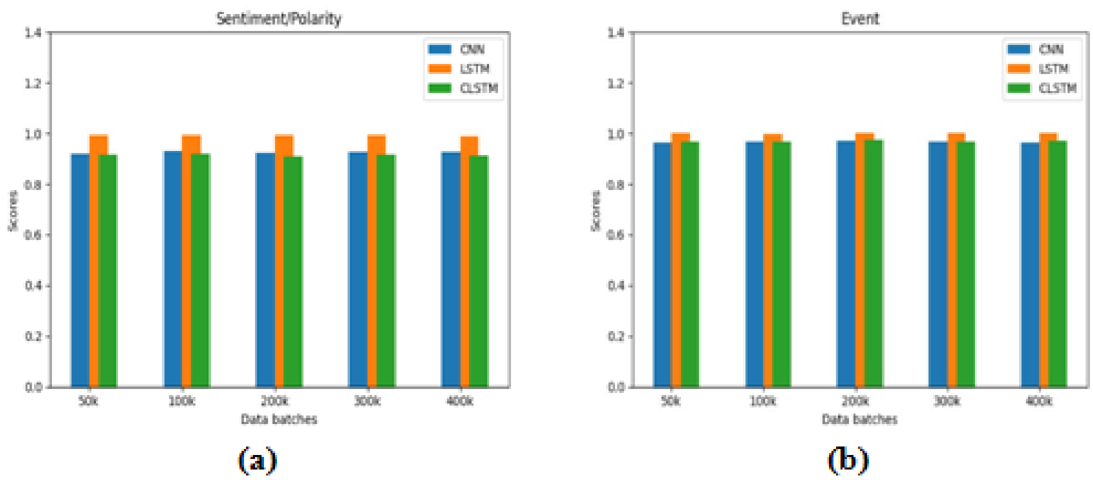

Fig. 5: Sentiments and Events on Precision, Recall, F1 Score and Accuracy for CNN, LSTM and C-LSTM Models respectively

Figure 3 shows the loss for dataset in batches with decrease in cross entropy loss as the number of epochs increases during the training. However, the loss for LSTM improve better than CNN and C-LSTM, thus, the loss for CNN and C-LSTM are very close as the number of epochs increases. In Fig. 4 confusion matrix for predicted set for both events and polarity are shown. It consists of True Positive (**TP**), **F**alse Positive (**FP**), False Negative (**FN**) and True Negative (**TN**), respectively.

In Tables 1 – 6, precision, recall, F1 score and accuracy are shown as performance measures for Sentiments and Event, respectively (http://github.com/princesegzy01/ensemblesdeeplearning-multilabek-covid-19). While Fig. 5 represents the bar chart for multiclass (Sentiment) and multi-label (Events), respectively for the predicted data.

Tables 1 – 3: Sentiments mean value for precision, recall, f1-score and accuracy of CNN, LSTM and C-LSTM Models, respectively

Table 1: CNN sentiments mean value

Dataset	Precision	Recall	F1-score	Accuracy
50000	0.924	0.852	0.882	0.921
100000	0.928	0.876	0.900	0.931
200000	0.932	0.852	0.886	0.925
300000	0.938	0.870	0.900	0.926
400000	0.904	0.892	0.896	0.929

Table 2: LSTM sentiments mean value

Dataset	Precision	Recall	F1-score	Accuracy
50000	0.994	0.980	0.980	0.993
100000	0.996	0.986	0.988	0.994
200000	0.990	0.990	0.990	0.993
300000	0.984	0.988	0.986	0.993
400000	0.978	0.986	0.980	0.990

Table 3: C-LSTM sentiments mean value

Dataset	Precision	Recall	F1-score	Accuracy
50000	0.892	0.856	0.874	0.916
100000	0.924	0.836	0.874	0.921
200000	0.904	0.858	0.874	0.911
300000	0.902	0.856	0.876	0.917
400000	0.884	0.862	0.62	0.912

Tables 4 – 6: Events mean value for precision, recall, f1-score and accuracy of CNN, LSTM and C-LSTM Models, respectively

Table 4: CNN model events mean value

Dataset	Precision	Recall	F1 score	Accuracy
50000	0.986	0.957	0.973	0.965
100000	0.990	0.960	0.973	0.968
200000	0.990	0.67	0.973	0.971
300000	0.990	0.956	0.973	0.968
400000	0.987	0.60	0.973	0.966

Table 5: LSTM events mean value

Dataset	Precision	Recall	F1-score	Accuracy
50000	1	1	1	0.999
100000	1	1	1	0.998
200000	1	1	1	0.999
300000	1	1	1	0.999
400000	1	1	1	0.999

Table 6: C-LSTM events mean value

Dataset	Precision	Recall	F1-score	Accuracy
50000	0.993	0.953	0.970	0.967
100000	0.990	0.963	0.976	0.970
200000	0.997	0.963	0.980	0.974
300000	0.993	0.957	0.977	0.968
400000	0.993	0.963	0.988	0.972

However, from the Table, it shows that as the number of dataset increases, accuracy increases for all dataset except 400000 datasets. This occurs for CNN, LSTM and C-LSTM on sentiment and events. Though, the performance measures for LSTM improve better than CNN and C-LSTM.

Conclusion and Recommendation

People's view on COVID-19 and analysis of Twitter provides opinion and events about users' emotion on the recent pandemic. This paper found the polarity and events of tweets using Textblob and NLTK similarity tool. A total number of 400,000 tweets are used for calculating CNN, LSTM and C-LSTM accuracy. The accuracy for LSTM is 99% which is better than CNN and C-LSTM. It shows that during pandemic, people' opinions were negative for COVID-19 and the effect is more on Education/school, Health/Medical and Business/Economics. It is recommended that a more sophisticated deep learning models such as Autoencoder, BERT can be employed for big datasets.

Acknowledgement

The information used in the experiment can be found in Twitter, Tensorflow, Keras, NLTK, Python, GloVe, Gensim

.

References

Abdi A, Shamsuddin SM, Hasan S & Piran J 2019. Deep learning based sentiment classification of evaluative text based on multi-feature fusion. *Infor. Processing and Mgt.*, 56: 1245-1259.

Ahmad S, Asghar MZ, Alotaibi FM & Awan I 2019. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Human-Centric Comp. and Infor. Sci.*, 9:24 https://doi.org/10.1186/s13673-019-0185-6

Bengio Y, Simard P & Frasconi P 1994. Learning long-term dependencies with gradient descent is difficult. *Institute of Electr. Electronic Engr. (IEEE) Transac. on Neural Networks*, 5(2): 157-166.

Brindha S, Sukumaran S & Prabha K 2016. A survey on classification techniques for text mining. Proceedings of the 3rd International Conference on Advanced Computing and Communication Systems Institute of Electrical Electronic Engineering (IEEE) Coimbatore, India, pp. 1: 1-5 https://doi.org/10.1109/ICACCS.2016.7586371

Diab DM & Hindi KME 2017. Using differential evolution for fine tuning naïve Bayesian classifiers and its application for text classification. *Applied Soft Computing*, 54(20): 183-199.

Francesco DG, Damiano P, Claudia M, Mario A, Vincenzo R Nicola V & Lee S 2020. Coronavirus diseases (COVID-19) current status and future perspectives: A narrative review. *Int. J. Envt. Res. and Public Health*, 17: 2690 http://doi:10.3390/ijerph17082690

Hochreiter S & Schmidhuber J 1997. Long short-term memory. *Neural Computation*, 9(8): 1735–1780.

Kim Y 2014. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, October 2014, Association for Computational Linguistics, pp. 1746-1751.

Kingma D & Ba J 2014. Adam: A method for stochastic optimization. International Conference of Learning Representation (ICLR), pp. 1-15.

Kowsari K, Brown DE, Heidarysafa M, Meimandi KJ, Gerber MS & Barnes LE 2017. Hdltex: Hierarchical deep learning for text classification. In 2017 16th *IEEE Int. Conf. on Machine Learning and Applic.* (ICMLA), pp. 364-371.

Kwak H, Lee C, Park H & Moon S 2010. What is twitter, a social network or a news media? In: *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10, 591–600

LeCun Y, Bottou L, Bengio Y & Haffner P 1998. Gradient-based learning applied to document recognition.

Proceedings of the Institute of Electrical and Electronics Engineers, 86(11): 2278–2324.

Lewis K & Full F 2020. 5G is not accelerating the spread of the new coronavirus URL: https://fullfact.org/health/5G-not-accelerating-coronavirus/ [accessed 2020-04-14] Mar 31

Lilleberg J, Zhu Y & Zhang Y 2015. Support vector machines and word2vec for text classification with semantic features, in 14th IEEE International Conference on Cognitive Informatics and Cognitive Computing, 136-140.

Loria S 2019. TextBlob: Simplified Text Processing. URL https://textblob.

Mikolov T, Sutskever I, Chen K, Corrado G & Dean J 2013. Distributed representations of words and phrases and their compositionality. In: *Proceedings of the 26th Int. Conf. on Neural Infor. Processing Sys*., 2: 3111-3119.

Mukherjee S, Bhattacharyya P 2013. Sentiment analysis: A literature survey. *CoRR*, abs/1304.4520, 1-51 http://arxiv.org/abs/1304.4520.

Nii M, Takahama K, Uchinuno A & Sakashita R 2014. Soft class decision for nursing-care text classification using a k-nearest neighbour based system. In: IEEE I*nternational Conference on Fuzzy Systems*, 1825-1830.

Palen L & Hughes AL 2018. Social Media in Disaster Communication. In: Handbook of Disaster Research, edited by H Rodríguez, W Donner & JE Trainor. Cham: Springer International Publishing, pp. 497–518.

Pennington J, Socher & R, Manning C 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543.

Pham DH & Le AC 2016. A neural network based model for determining overall aspect weights in opinion mining and sentiment analysis. *Indian J. Sci. and Techn.*, 9: 1–6.

Prechelt L 2012. Early Stopping-But When? in Neural Networks: Tricks of the Trade. 2nd edn; G Montavon, GB Orr, and K.-R. M¨uller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 53–67. readthedocs.io/en/dev/.

Saha S, Yadav J & Ranjan P 2017. Proposed Approach for Sarcasm detection in Twitter. *Indian J. Sci. Techn.*, 10(25) doi:10.17485/ijst/2017/v10i25/114443

Shanapinda S 2020. The Conversation. 5G Radiation doesn't Cause or Spread the Coronavirus. Saying it does is Destructive. URL: https://tinyurl.com/uppng7w [accessed 2020-04-14]

Shen Y, He X, Gao J, Deng L & Mesnil G 2014. Learning Semantic Representations Using Convolutional Neural Networks for Web Search. In Proceedings of the 23rd International Conference on World Wide Web. 2014. ACM, 1-2.

Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A & Al-Jabir A 2020. A review of the 2019 novel coronavirus (COVID-19). *Int. J. Surgery*, April 2020. doi: 10.1016/j.ijsu.2020.02.034

Srivastava N, Hinton G, Krizhevsky A, Sutskever I & Salakhutdinov R 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1): 1929–1958.

Stas J, Juhar J & Hladek D 2014. Classification of heterogeneous text data for robust domain-specific language modeling. *EURASIP J. Audio, Speech, and Music Processing*, 14: 1-12.

Vieweg S, Castillo C & Imran M 2014. Integrating social media communications into the rapid assessment of sudden onset disasters. *Social Informatics Springer*, pp. 444–461.

Wallace BC, Choe DK, Kertz L & Charniak E 2014. Humans require context to infer ironic intent (so computers probably do, too). In: *Proceedings of the 52nd Annual Meeting of the Assoc. Computa. Ling.*, 2: 512–516.

WHO 2020. Statement on the second meeting of the International Health Regulations (2005) 2020 Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). 2020. https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-healthregulations-(2005)-emergency-committee-regarding-the-outbreak-of-novelcoronavirus-(2019-ncov). Accessed 1 Feb 2020

Xu S 2018. Bayesian naive bayes classifiers to text classification. *Journal of Information Science*, 44(1): 48-59.

Yang K, Cai Y, Huang D, Li J, Zhou Z & Lei X 2017. An effective hybrid model for opinion mining and sentiment analysis. IEEE *International Conference on Big Data and Smart Computing* (BigComp), pp. 465–466.