A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification

Preprint · September 2018

Md Saiful Islam
Shahjalal University of Science and Technology

A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification

Abdullah Aziz Sharfuddin
Department of CSE
Shahjalal University of Science and Technology
Sylhet, Bangladesh
abdullahshakkhor@gmail.com

Md. Nafis Tihami
Department of CSE
Shahjalal University of Science and Technology
Sylhet, Bangladesh
nafistiham@student.sust.edu

Md. Saiful Islam
Department of CSE
Shahjalal University of Science and Technology
Sylhet, Bangladesh
saiful-cse@sust.edu

***Abstract*—In the field of sentiment classification, opinions or sentiments of the people are analyzed. Sentiment analysis systems are being applied in social platforms and in almost every business because the opinions or sentiments are the reflection of the beliefs, choices and activities of the people. With these systems it is possible to make decisions for businesses to political agendas. In recent times a huge number of people share their opinions across the Internet using Bengali. In this paper a new way of sentiment classification of Bengali text using Recurrent Neural Network(RNN) is presented. Using deep recurrent neural network with BiLSTM, the accuracy 85.67% is achieved.**

***Index Terms*—Bengali text; Deep learning; Sentiment Classification; RNN; LSTM; BiLSTM; Facebook; NLP**

I. Introduction

With the elevation in the communication technology i.e. world wide web, a huge number of people from all lineages across the world take part in social networks and express their emotions or opinions on a wide range of topics. Now it is a dire need to summarize the data created by people over the social networks and see the insights from them. Besides, in the field of NLP, it has become a topic of enormous interest. Because, it is needed to make smart recommending systems, anticipating the results of political elections, understanding the feedback of people on public events and movements.

SA is a method of finding and classifying opinions expressed in a piece of text basing on computation technologies, especially in order to find out whether the writer's behavior towards a specific topic, product, etc. is positive, negative, or neutral. SA also refers to the administration of opinions, sentiments and subjective text [1]. It also gives the comprehensive data associated to public views, as it goes through all the different kinds of tweets, reviews and comments. It is a verified mechanism for the prediction of a numerous momentous circumstances, for instance movie ratings at box office and public or regional elections [2]. Public opinions are used to value a certain matter, i.e. person, product or place and might be found at different websites like Amazon and Yelp. The sentiments can be specified into negative, positive or neutral and even more classes. SA can automatically find out the expressive direction of user reviews or opinions whether the users have a good or positive impression or a negative impression. [3]

The usage of SA is broad and powerful. Its demand has grown due to the escalating need of extracting and inspecting hidden information from data coming through social medias. Different organizations around the world are using the ability to extract hidden data now-a-days. The change in sentiments can be connected to the change in stock market. The Obama administration used opinion mining in the 2012 presidential election to detect the public opinion before the announcement of policy. [4]

Deep learning has shown great performance in SA. In this article, a way of SA using deep recurrent neural network with BiLSTM is presented.

II. Related Works

Sentiment analysis is not new for English language. A significant number of research works have been done within this scope. Arvind et al. [5] applied Skip-Gram model where as Paramveer et al. [6] applied CCA(Canonical Correlation Analysis) for in-depth vectorization. Duyu Tang et al. [7] did their work on sentiment analysis of tweets. These contained not only the information but also the syntactic context.Bengali hasn't yet progressed in this particular research area. Approaches that have been already made are dependent on machine learning mostly. Dipankar Das [8] used Parts of speech tagger for to tag emotion. He achieved 70% accuracy. K.M Azharul Hasan et al. [9] used contextual valency analysis in Bangla text for SA. In their approach, using parts of speech tagger, they calculated the three classes (positivity, negativity and neutrality).The percentages was summed up to get result. They got an accuracy of 75%. K.M Azharul Hasan et. al [10] used unsupervised learning algorithm and created phrase pattern. Shaika Chowdhury [11] used SVM and Maximum Entropy on Bengali micro blog posts. They compared these classifiers using the accuracy metric. M Al-Amin et. al [12] used word2vec model for their base of sentiment analysis in Bangla. They used the positions of vector representation of the Bengali words to create a positive or negative score for each word and determine the sentiment of a specific text. They achieved a 72.5% accuracy. Sentiment analysis on the rohingya issue was done by Hemayet et al. [13] Similar works were also done by Al-Amin et al. [14].

978-1-5386-8207-4/18/$31.00

III. Methodology

This section describes the training and architecture of the models that are proposed for the classification of sentiments from Bengali texts.

A. Recurrent Neural Networks

Recurrent neural networks (RNN) are a type of network which form memory through recurrent connections. In feed forward networks, inputs are independent of each other. But in RNN, all inputs are connected to each other. This lets the network to exhibit vigorous temporal behavior for a time sequence which makes it favorable for sequential classification like sentiment analysis. As it can be seen in the figure, at first,

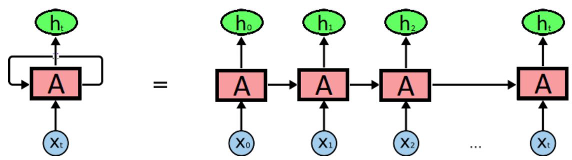

Fig. 1. RNN loop

it takes the x_0 from the sequence of input and then it outputs h_0 which together with x_1 is the input for the next step. So, the h_0 and x_1 is the input for the next step. Similarly, h_1 from the next is the input with x_2 for the next step and so on. This way, it keeps remembering the context while training.

$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b) \quad (1)$$
$$y_t = g(W_{yh}h_t + c) \quad (2)$$

B. Long Short Term Memory Units

In 1997, an alteration of RNN with Long Short-Term Memory units, or LSTM units [15], was proposed by the Sepp Hochreiter and Juergen Schmidhuber.Some errors back-propagate through time in general RNN. These LSTM units help to bypass these errors. While keeping a more consistent error, they let RNNs keep on learning over several time steps. LSTMs consist of information outside the basic flow of the rnn in a valved block [16].

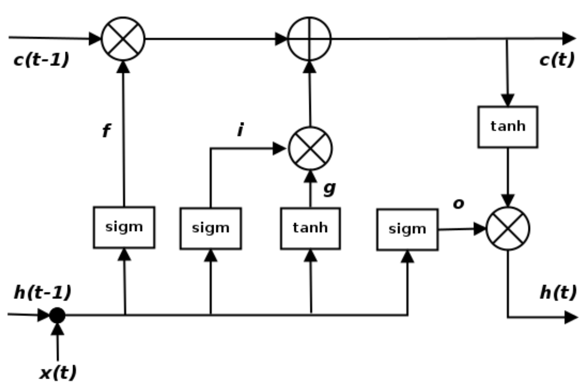

Fig. 2. LSTM Unit

Neural network's notes get triggered by the notes they get. Likewise, LSTM's gates pass on or block the data based on its weight. After that, these signals are grated with their own sets of weights. Subsequently, RNN's learning process modify these weights that control hidden states and input. Ultimately, these cells learn when to let information to get in, get out or be deleted through the consequent steps of taking guesses, back propagating error, and modulating weights via gradient descent. [17]

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$
$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (4)$$
$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$
$$c_t = f_t o c_{t-1} + i_t o \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (6)$$
$$h_t = o_t o \sigma_h(c_t) \quad (7)$$

C. Bidirectional LSTM

BiLSTMs are proven especially helpful in the occasions where the context of the input is needed. It is very useful for works like sentiment classification. In unidirectional LSTM information flows from the backward to forward. On the contrary in Bi-directional LSTM information not only flows backward to forward but also forward to backward using two hidden states. Hence Bi-LSTMs understand the context better. BiLSTMs were used to escalate the chunk of input information usable to the network. Structure of RNN with LSTM and RNN with BiLSTM. [18] Basically, BRNN follows such a process where the neurons of a normal RNN are broken into bidirectional ways. One is for backward states or negative time direction, and another for forward states or positive time direction. The inputs of the reverse direction states are not connected to these two states' results. The structure or BiLSTM is shown in the diagram below. By utilizing two time

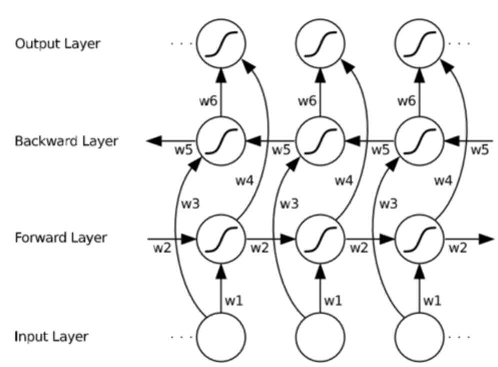

Fig. 3. Bidirectional LSTM

directions, input data from the past and future of the current time frame can be used. Whereas standard RNN requires the delays for including future data. [18]

IV. Experiment Setup and Results

A. Sentiment Analysis Dataset

It is very hard to find a benchmark dataset for Bengali Sentiment Classification. Because researchers do not publish their datasets along with their work. But we have managed a dataset from Md. Al-Amin et al. [12] Which contained about

15000 comments fetched from Facebook. But the dataset was not as good as we expected. so needed to reproduce the dataset by ourselves. We have fetched comments from Facebook and labeled them by hand before training our models. Total number of fetched comments was around 30,000. Among them about 30% were usable. There are in total of 10000 comments(5000 positive comments and 5000 negative comments). The dataset is evenly distributed. All the emojis, symbols, numbers, stickers were deleted. All English letters were removed. The dataset only contained plain Bengali text.

TABLE I
INFORMATION OF WORDS

Number of total words	152,976
Number of unique words	26,661
Highest occurrence of one word	2332
Length of the longest comment in words	447
Highest occurring words	না, করে, এই, আর, জন্য
Top negative words	খারাপ, বাজে, কিন্তু
Top positive words	ভাল, খুব, ঠিক

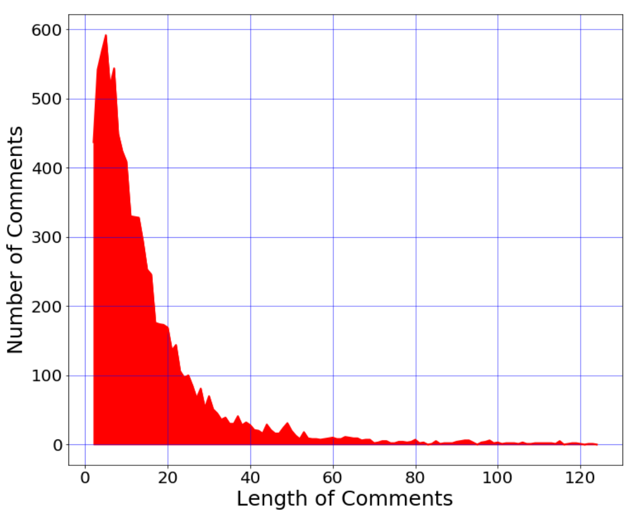

Fig. 4. Length vs number of comments

In the Fig:4, the distribution in the aspects of length of the comments can be seen. Retrospectively, we have seen that, the comments on various news article posts are rather short. In the graph there is a high peak in the area of approximately 6. and the peak steeply goes down to 20. And then a gradual fall.

Nearly all of the words are shorter than length of 50 words. The rest are negligible. It should also be taken into consideration that, some one word comments have been dropped which do not express a specific sentiment. Some other informations can be found in table I. If we have a look at the number of unique words and the number of most occurrences, we can anticipate the distribution of words against their occurrences. On average every word has occurred 5.73 times.

B. Results and Analysis

Our models have been trained on both datasets. Despite of being small, it performed better on the reproduced dataset. It also outperformed the previous work [12] with the proposed deep learning model.

TABLE II
ACCURACY COMPARISON OF MODELS ON DATASET [12]

Model	Accuracy(%)	Train(%)	Test(%)
Model [12]	75.50	90	10
Proposed Approach	82.53	67	33

TABLE III
ACCURACY COMPARISON WITH TRADITIONAL CLASSIFYING MODELS

Model	Accuracy(%)
Support Vector Machine	68.77%
Decision Tree Classifier	67.50%
Logistic Linear Regression	60.94%
Proposed Approach	85.67%

And for the dataset with 10000 comments using the same metrics the model achieved 85.67% accuracy. The dataset was also trained with the traditional models like support vector machine, decision tree classifier and logistic linear regression and accuracy obtained were respectively 68.77%, 67.50% and 60.94%.

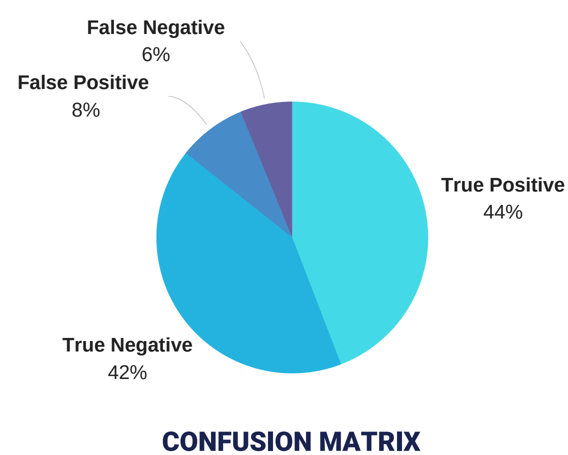

Fig. 5. Confusion Matrix

TABLE IV
RESULT OF TESTING ON 3300 DATA

	Predicted Positive	**Predicted Negative**
Actually Positive	44.15	8.13
Actually Negative	41.52	6.20

TABLE V
TEST AGAINST CUSTOM SAMPLES

Sentence	Model[17] with dataset[17]	Proposed model with our dataset	Actual
বাংলাদেশ আজ খেলায় হেরেছে।	**Positive**	Negative	Negative
আমার জীবন খানিকটা অতিষ্ঠ হয়ে উঠেছে	**Positive**	Negative	Negative
রাতে একা একা হলের রাস্তায় হাটার মজা বুঝি	**Positive**	Negative	Negative
পুর্নিমা রাতে সাইকেল চালাতে খুব মজা	Positive	Positive	Positive
আজকের দিনটা তেমন খারাপ নয়।	**Negative**	Positive	Positive
অনেক দিন পর একটু ছুটি পেলাম।	Positive	Positive	Positive

Then both models were tested (trained with dataset [12] and trained with small dataset) against some custom context based samples that were not in any of the datasets. The predictions are shown in table 4. In the table the mis-predictions are marked in block fonts. Out of six samples the model trained with 15000 comments got 4 wrong. And on the other hand model trained with 10000 comments got all correct.

Moreover, dataset [12] was not polished. It contained emojis, numbers, English letters punctuations and symbols. In spite of being smaller the reproduced dataset with 10000 comments is more fine-grained compared to the one with 15000 comments. The model was able to understand the sentiment from the context of the sentences. Hence the model performed better.

V. CONCLUSION

SA has become very important for the business owners. Because, with sentiment analyzers, it is now possible to understand the user activities and choices. Many works have been done for English. In contrast of that, work done in Bengali is very little. This research is a little step forward to fill the void.

Despite being one of the most used languages in the world, Bengali lacks in both benchmark datasets and a well furnished model for sentiment analysis. Moreover, researchers usually do not publish their dataset. The dataset that was made for this research is clearly step ahead since it will be enriched and published for research purposes. Moreover, the dataset was not stemmed for our purpose. In future, we can stem the dataset and our result might improve.

It is clear that deep learning models are the future here. Using the recurrent deep learning models, it is now possible to achieve state of the art performance in Bengali sentiment classification. In future it will be interesting to see the business applications or sentiment analyzers for Bengali text using deep recurrent models. Also hybrid deep learning models can be applied to the task of sentiment classification.

VI. ACKNOWLEDGEMENT

For our work, we are greatly thankful to the SUST NLP group and also thankful to the previous researchers who have put their efforts in Bengali sentiment analysis. Also, we are grateful to the researchers who have progressed the field of NLP and Neural Networks. We acknowledge M. Al-Amin et al. for providing us with their dataset without which our work would not be complete.

REFERENCES

[1] A. V. Yeole, P. Chavan, and M. Nikose, "Opinion mining for emotions determination," in *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on*. IEEE, 2015, pp. 1–5.

[2] B. Heredia, T. M. Khoshgoftaar, J. Prusa, and M. Crawford, "Cross-domain sentiment analysis: An empirical investigation," in *Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on*. IEEE, 2016, pp. 160–165.

[3] F. Luo, C. Li, and Z. Cao, "Affective-feature-based sentiment analysis using svm classifier," in *Computer Supported Cooperative Work in Design (CSCWD), 2016 IEEE 20th International Conference on*. IEEE, 2016, pp. 276–281.

[4] K. Bannister. (2018) Sentiment analysis uses @ONLINE. [Online]. Available: https://www.brandwatch.com/blog/understanding-sentiment-analysis/

[5] A. Neelakantan, J. Shankar, A. Passos, and A. McCallum, "Efficient non-parametric estimation of multiple embeddings per word in vector space," *arXiv preprint arXiv:1504.06654*, 2015.

[6] P. Dhillon, D. P. Foster, and L. H. Ungar, "Multi-view learning of word embeddings via cca," in *Advances in neural information processing systems*, 2011, pp. 199–207.

[7] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2014, pp. 1555–1565.

[8] D. Das, "Analysis and tracking of emotions in english and bengali texts: a computational approach," in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 343–348.

[9] K. A. Hasan, M. Rahman *et al.*, "Sentiment detection from bangla text using contextual valency analysis," in *Computer and Information Technology (ICCIT), 2014 17th International Conference on*. IEEE, 2014, pp. 292–295.

[10] K. A. Hasan, S. Islam, G. Mashrur-E-Elahi, and M. N. Izhar, "Sentiment recognition from bangla text," in *Technical Challenges and Design Issues in Bangla Language Processing*. IGI Global, 2013, pp. 315–327.

[11] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in bangla microblog posts," in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*. IEEE, 2014, pp. 1–6.

[12] M. Al-Amin, M. S. Islam, and S. D. Uzzal, "Sentiment analysis of bengali comments with word2vec and sentiment information of words," in *Electrical, Computer and Communication Engineering (ECCE), International Conference on*. IEEE, 2017, pp. 186–190.

[13] H. A. Chowdhury, T. A. Nibir, M. Islam *et al.*, "Sentiment analysis of comments on rohingya movement with support vector machine," *arXiv preprint arXiv:1803.08790*, 2018.

[14] M. S. Islam, M. Al-Amin, and S. D. Uzzal, "Word embedding with hellinger pca to detect the sentiment of bengali text," in *Computer and Information Technology (ICCIT), 2016 19th International Conference on*. IEEE, 2016, pp. 363–366.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional cnn-lstm model," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 225–230.

[17] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[18] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.