Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023)
IEEE Xplore Part Number: CFP23BC3-ART; ISBN: 978-1-6654-5630-2

Qualitative Sentiment Analysis of YouTube Contents based on User Reviews

Mohammad Rakin Uddin
Department of EEE
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh.
rakin_96@yahoo.com

Nafisa Sadaf Prova
Department of EEE
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh.
n.prova23@gmail.com

Atik Jawad
Department of EEE
University of Liberal Arts Bangladesh
Dhaka- 1205, Bangladesh.
atik.jawad@ulab.edu.bd

Md. Rajib Rayhan
Department of CSE
Pabna University of Science and Technology
Pabna, Bangladesh.
rajib2jsr@gmail.com

S.M. Rhydh Arnab
Department of EEE
Islamic University of Technology
Gazipur, Dhaka, Bangladesh.
smrhydharnab@iut-dhaka.edu

S. M. Hasan Sazzad Iqbal
Department of CSE
Pabna University of Science and Technology
Pabna, Bangladesh.
sazzadice@gmail.com

2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 978-1-6654-5630-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAAIC56838.2023.10141517

***Abstract*— Sentiment analysis (a branch of Natural Language Processing) is one of the most crucial components of assessing online contents in current digital age. One of the leading competitors in the digital content market is YouTube, where thousands of new videos are uploaded every day. To assess the opinions of the users and customize advertisements and future contents accordingly, an appropriate sentiment analysis model is very crucial. Therefore, this study proposes a novel combination of deep learning technique called Encoder-decoder based Attention model to conduct sentiment analysis on YouTube reviews. In the proposed approach, the squeeze-and-excitation attention layer is utilized. Nearly 7,00,000 user comments from 8000 YouTube channels are utilized to train and test the proposed model in this study. Three different sentiments—positive, negative, and neutral—are assigned to the comments in the dataset for training purposes. The results show that the maximum accuracy of the model is 92.8% whereas maximum F1-score is 91.9%, surpassing several ML based state-of-the-art approaches. The evaluated metrics verify the balanced performance of the proposed approach. Moreover, the impact of the attention mechanism on the proposed approach is also assessed. Model accuracy for sentiment analysis is significantly improved by combining attention mechanism with encoder-decoder. As a generalized framework, this research can serve as a valuable guideline for future sentiment analysis on different languages.**

Keywords*— *LSTM, attention, sentiment analysis, encoder, decoder, NLP

I. INTRODUCTION

Social media is the platform where individuals connect to exchange information and ideas while creating contents and sharing. In this regard, YouTube is a widely recognized website for sharing videos on the internet, where individuals have the ability to view, appreciate, distribute, give feedback, and publish their own videos. Approximately 2.1 billion people globally utilize YouTube on a monthly basis [1]. The estimated number of users is expected to continue to rise every year, indicating that there is no indication of a decrease in this figure. The different responses expressed in the comments section by users can greatly influence the image of the video content and channel. YouTube video comments are very useful to determine what the users are thinking [2]. It helps the content creator to make desirable content for the viewers. Hence, a technique can be developed to determine YouTube users' perceptions of video content utilizing data gathered from textual content.

According to [3], utilizing a sentiment analysis approach is an effective way to comprehend the meaning of each comment. The classification of comments as positive or negative could assist a YouTube user in determining the significance of the published content based on the opinions of other users. Moreover, customizing advertisements according to the sentiment can boost channel exposure too. Sentiment analysis has become a crucial field of interest in natural language processing, with over 7,000 articles written on the subject [4]. The authors suggested using the Gini Index feature selection technique along with the SVM classifier to classify the data. [5]. M. R. Huq et al. [6] used SVM and K-NN for sentiment analysis. A technique for classifying the sentiment of a given sentence using a semi-supervised learning approach was suggested by the authors in [7]. CNN-based models stand out from previous techniques as they do not require the tedious task of feature engineering. To that purpose, a CNN based approach with K-means to improve the performance [8]. Then, Kang Liu et al. [9] used a bidirectional selection method, to identify sentiment words in proximity to product features and vice versa. Furthermore, Huiling Chen et al. [10] used an unsupervised LSTM network for fine-tuned sentiment analysis without considering implicit features of the reviews. However, these models suffer from different inadequacies such as extracting opinion targets only using opinion relations, unsatisfactory accuracy for a large versatile dataset. Therefore, there remains a critical research gap in developing model which has better accuracy, balanced performance, and applicability in practical solutions.

In response to these challenges, a novel sentiment analysis technique called the 'Encoder decoder-based Attention Model' has been introduced in this work. This technique is designed specifically for YouTube reviews and consists of embedding layers, an Encoder-decoder LSTM model, a Squeeze-and-excitation attention layer, and fully connected dense layers for classification. The proposed technique is trained on a large dataset of around 7,00,000 comments [11] and evaluated using various performance measures in comparison to other cutting-edge methodologies.

The remaining parts of this paper are organized as follows. The proposed methodology is discussed thoroughly with concise description of the used dataset in Section II. In section III, proposed model's performance is analyzed in detail. Finally, the conclusions of this research are briefly stated in the closing section.

II. METHODOLOGY

A. Dataset

The dataset [11] contains 691400 English comments made by YouTube users which are compiled from 200 popular YouTube videos. 313611 of the 691400 comments are positive, 141806 are negative, and the remaining comments are neutral. We divided the data into train, test, and validation datasets for this work. 80% of the comments in the train dataset. 10% of the dataset is kept for testing and remaining 10% are reserved for validation.

B. Preprocessing

The text data has undergone some processing before being used for sentiment categorization. Punctuation and other forms of symbols are initially taken out of the dataset. The most frequent words in a language, such as articles, pronouns, prepositions, etc., are referred to as stop words since they don't provide much context and that's why they were dropped from text dataset. Since the entire statement cannot be immediately placed into the model. This is why phrases are broken up into tokens using the Keras library's "Tokenizer" class.

C. Proposed Model

To enhance the accuracy of sentiment analysis for YouTube comment reviews, we integrated various features such as embedding layers, Encoder-decoder LSTM model, attention layer, and fully connected dense layers. In Fig. 1, the complete architecture of proposed model is shown.

Firstly, Tokenization is used to convert the preprocessed input data into vectors, which are then given to an embedding layer. We can turn each word into a fixed-length vector with a specified length due to the embedding layer. In this study, text embedding was accomplished using the "Word2Vec" [12] approach. The result is a group of word-vectors, where word-vectors close to one another in vector space have similar meanings, and word-vectors far distant from one another have distinct meanings, depending on the context. The "Word2Vec" model offers the Continuous Bag of Words (CBOW) and Skip Gram techniques for embedding text data. The CBOW version of the Word2Vec model, which predicts the target word by combining the dispersed representations of the surrounding words, is applied in this study. The weight matrix obtained using this technique is the embedding matrix after the Word2Vec model has been trained. Following that, the output of the embedding layer is normalized using batch normalization, which uses a transformation to keep the output mean and standard deviation near to 0 and 1, respectively.

Secondly, the encoder decoder LSTM model is coupled to the embedded layer's output following normalization. Two successive LSTM layers make up the encoder portion. There are 100 units in the first LSTM layer and 20 units in the next one. Before being inserted into the decoder section, the encoder layer's output is repeated 464 times. The decoder section features two LSTM layers, the first of which has 20 cells and the second of which has 100 cells.

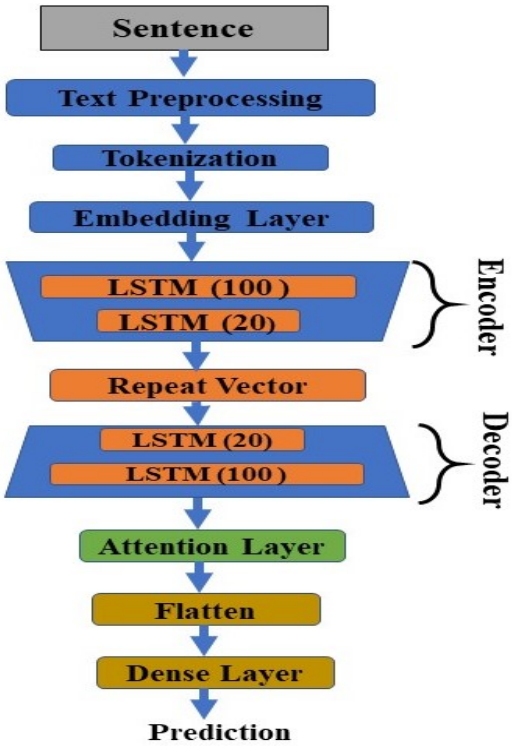

Fig. 1. Schematic diagram of proposed model.

Note that, the input and output data shapes are both unaltered in this architecture. Encoder-decoder models are typically used to remove unimportant information and to identify the terms that are most significant in determining the sentiment of the sentence. The encoder component of the model compresses the input data, and after that the decoder part of the model increases the compressed data's dimension until it is the same dimension as the input data. The LSTM encoder decoder model reconstructs the input data in this manner, but the output revealed from the encoder decoder part places greater focus on the most important features and less emphasis on other features.

Finally, the output of the encoder decoder is fed into the attention network after receiving the data. Although the encoder decoder LSTM model works well on sequential data and is capable of capturing semantic context, its key drawback is its inability to recognize strong contextual relationships from lengthy phrases, which ultimately reduces the model's overall accuracy. This study introduces an attention technique to capture the context that determines final classification in order to address this issue. In this study, squeeze-and-excitation attention [13] is used to identify context that the encoder decoder model was unable to discern. The structure of the squeeze-and-excitation attention block is graphically depicted in Fig.2. This module can add parameters to the model that will enable adaptive weighing of each unique feature map. This attention layer's output is flattened and fed into two thick layers that are completely interconnected. The classification outcome is produced by the last layer, which employs the softmax activation function.

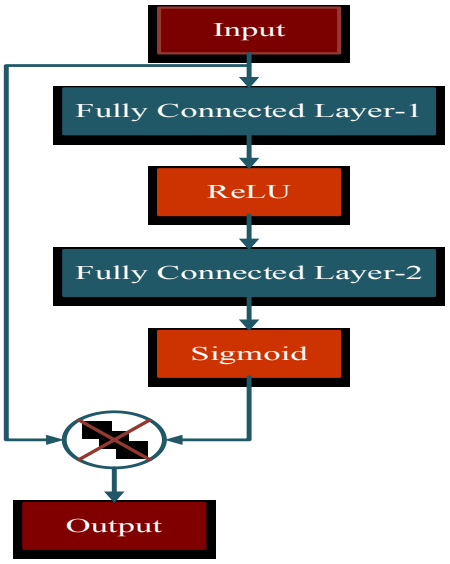

Fig. 2. Squeeze-Excitation Attention Block Diagram

III. Results and Analysis

There are two subsections within this section. The first segment provides a description of the performance metrics used in this work. The performance of the proposed approach is thoroughly discussed in the second part, along with a comparison to the performance of earlier studies.

A. Benchmarking

The proposed model performance is evaluated by looking at the recall, precision and F1 score [14].

The measure of accuracy indicates the model's overall performance across all classes and is particularly valuable when all classes are considered equally significant. The number of accurate predictions is divided by the total number of predictions to determine accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Recall is the ratio of a class's prediction accuracy to the total amount of facts categorized in the class. The recall formula can be written as follows:

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Precision is a metric to measure model's performance by calculating positive prediction made by model.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Furthermore, the F1 score is identified as the harmonic mean of precision and recall.

$$F1-score = 2\times\frac{Precision\times Recall}{Precision+ Recall} \quad (4)$$

Also, in this study Cohen's Kappa score is also calculated which is an insightful metric to assess model performance for multi-class classification problem.

$$Kappa = \frac{P0-Pe}{1-Pe} \quad (5)$$

B. Model Evaluation

The entire dataset is split up into training, training, testing, and validation sets for this study. 80% of the complete dataset is reserved for training set, the test set contains 10% and 10% is in the validation set. A total of 100 training epochs were used to train the model.

According to Fig.3 the accuracy increases for validation set as epoch number increases up to 25 epochs. After that the steady Validation accuracy reaches its peak at 92.89 %. But loss in validation set increases, which can be observed from fig.4. Therefore, a slight problem of over fitting prevails in our approach. The calculated overall accuracy and Kappa score for the test dataset are 92.8% and 88.65%, respectively. Table I provides a summary of the overall performance along with class specific accuracy.

Furthermore, Fig. 5 displays the estimated confusion matrix. The matrix indicates that the proposed approach performs evenly in terms of overall accuracy, F1-score, and Recall. The impact of the attention mechanism on the proposed approach is then assessed. After removing Attention network from proposed model, the performance of

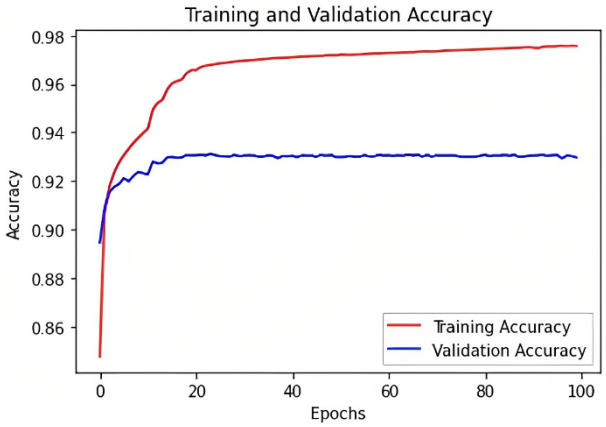

Fig. 3. Accuracy vs epoch curve for the proposed model

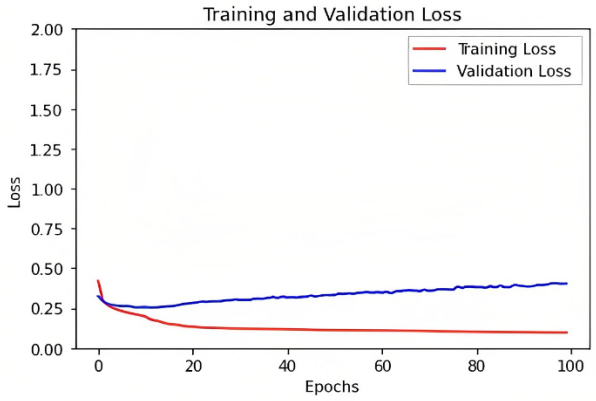

Fig. 4. Loss vs epoch curve for the proposed model

TABLE I. A Summary of Model Performance

Model's Overall Performance on Test Set			
Overall Accuracy	0.92802		
Cohen's Kappa	0.8865		
Model Performance by Class			
Metric Name	***Class Name***		
	Positive	***Negative***	***Neutral***
Accuracy	0.94	0.95	0.96
Precision	0.93	0.87	0.95
F1-Score	0.94	0.87	0.94
Recall	0.94	0.87	0.94

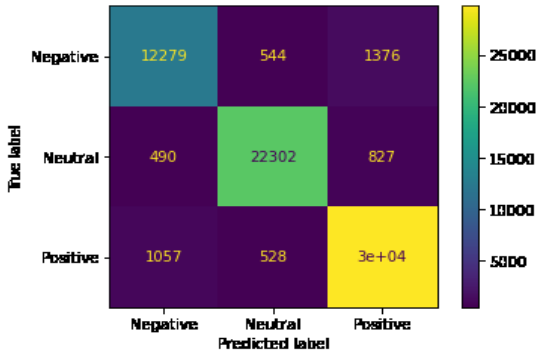

Fig. 5. Confusion matrix of the proposed model

the remaining of the proposed model is assessed and summarized. Later, Table II shows the model's performance with and without an attention network. Table II demonstrates that, in the absence of an attention mechanism, the model produced more False Positives (type-1 error) and False Negatives (type-2 error) while determining the sentiment classification. The quantity of false positives (FP) and false negatives (FN) has been considerably decreased by the addition of the attention network with encoder decoder model. In this approach, the attention mechanism expanded the model's parameter set and assisted in reducing the FP and FN rates.

Moreover, the proposed model is compared against state-of-the-art methodologies to further evaluate the model which includes Naïve Bayes [15], SVM [5], CNN [7], CNN with Attention network [16] and LSTM based approaches [17-18]. The findings in Table III demonstrate that the proposed model outperforms the corresponding state-of-the-art approaches significantly.

TABLE II. PERFORMANCE EVALUATION OF ATTENTION LAYER

	False Negative		False Positive	
Sentiment Type	***Without Attention Network***	***Complete Model***	***Without Attention Network***	***Complete Model***
Positive	1639	1585	2297	2203
Negative	1988	1920	1590	1547
Neutral	1385	1317	1125	1072

Table. III. Model Performance Comparison

Model	Accuracy
NaiveBayes[15]	57.9%
SVM[5]	67.7%
CNN[7]	90.9%
CNN+Attention[16]	91.4%
LSTM [17]	78%
CNN+Bi-LSTM [18]	90%
Proposed Model	92.8%

IV. CONCLUSIONS

In this research, a model for qualitative analysis of YouTube comments that can analyze user sentiments about the video based on user reviews is proposed. This model employs a preprocessing heuristic based on lexical analysis and then uses LSTM encoder–decoder model and squeeze-excitation based attention layers to predict user feelings from comments of any YouTube video. The findings indicate that the proposed approach performs in a balanced manner in terms of performance metrics. Moreover, the proposed model outperforms several current methods while comparing the accuracy and F1-score (considering the absence of labeled data). However, due to great numbers of parameters, it takes a bit longer for the training and testing period. Moreover, there is still room for the improvement of accuracy in this study. In future, we aim to implement our model in Bangla user review analysis by constructing useful resources such as building datasets and generating lexicons.

REFERENCES

[1] R. Gothankar, F. D. Troia, and M. Stamp, "Clickbait Detection for YouTube Videos," Advances in Information Security, pp. 261–284, 2022, doi: https://doi.org/10.1007/978-3-030-97087-1_11.

[2] Serbanoiu, Andrei, and Traian Rebedea. "Relevance-Based Ranking of Video Comments on YouTube." 2013 19th International Conference on Control Systems and Computer Science, IEEE, May 2013. Crossref, https://doi.org/10.1109/cscs.2013.87

[3] Kavitha, K. M., et al. "Analysis and Classification of User Comments on YouTube Videos." Procedia Computer Science, vol. 177, Elsevier BV, 2020, pp. 593–98. Crossref, https://doi.org/10.1016/j.procs.2020.10.084.

[4] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, pp. 82–89, 2013.

[5] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. Venugopal, ''Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier,'' World Wide Web, vol. 20, no. 2, pp. 135–154, Mar. 2017..

[6] M. R. Huq, A. Ali, and A. Rahman, ''Sentiment analysis on Twitter data using KNN and SVM,'' Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 6, pp. 19–25, 2017..

[7] Gamon M., Aue A., Corston-Oliver S., et al., "Pulse: mining customer opinions from free text". In: Proceedings of the 6th International Conference on Advances in Intelligent Data Analysis, Madrid, Spain: Springer-Verlag, 2005, pp.121-132..

[8] B. S. Lakshmi, P. S. Raj, and R. R. Vikram, ''Sentiment analysis using deep learning technique CNN with KMeans,'' Int. J. Pure Appl. Math., vol. 114, no. 11, 2, pp. 47–57, 2017.

[9] Liu K., Xu L., Zhao J., "Opinion target extraction using word-based translation model". In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Stroudsburg, USA: Association for Computational Linguistics, 2012, pp.1346-1356.

[10] Chen, Huiling, et al. "Fine-grained Sentiment Analysis of Chinese Reviews Using LSTM Network." Journal of Engineering Science and Technology Review, vol. 11, no. 1, International Hellenic University, Feb. 2018, pp. 174–79. Crossref, https://doi.org/10.25103/jestr.111.21.

[11] MITCHELL J. (2017, October). Trending YouTube Video Statistics and Comments, Version 1. Retrieved October 26, 2017 from https://www.kaggle.com/datasets/datasnaek/youtube.

[12] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013)

[13] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[14] Gulati, K., Saravana Kumar, S., Sarath Kumar Boddu, R., Sarvakar, K., Kumar Sharma, D., & Nomani, M. Z. M. (2022). Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to covid-19 pandemic. *Materials Today: Proceedings*, *51*, 38–41. https://doi.org/10.1016/j.matpr.2021.04.364

[15] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, ''Sentiment analysis of review datasets using Naive Bayes and K-NN classifier,'' Oct. 2016, arXiv:1610.09982. [Online]. Available: https://arxiv.org/abs/1610.09982

[16] B. Shin, T. Lee, and J. D. Choi, ''Lexicon integrated CNN models with attention for sentiment analysis,'' Oct. 2016, arXiv:1610.06272. [Online]. Available: https://arxiv.org/abs/1610.06272.

[17] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: a Hybrid Network for Targeted Aspect-Based Sentiment Analysis," Cognitive Computation, vol. 10, no. 4, pp. 639–650, Mar. 2018, doi: https://doi.org/10.1007/s12559-018-9549-x.

[18] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models," arXiv:1904.04206 [cs, stat], Apr. 2019, Available: https://arxiv.org/abs/1904.04206.