Journal of Theoretical and Applied Information Technology
31st May 2023. Vol.101. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FINETUNED ROBERTA ARCHITECTURE FOR MOOCS EVALUATION USING ADVERSARIAL TRAINING

MUNIGADAPA PURNACHARY[1], T ADILAKSHMI[2]
[1] Research Scholar, Department of Computer Science and Engineering, University College of Engineering, Osmania University, Telangana, India
[2] Professor and Head, Department of Computer Science and Engineering, Vasavi College of Engineering, Osmania University, Telangana, India

E-mail: [1]purnachary@gmail.com

ABSTRACT

Massive Open Online Course is an online learning platform that can provide the opportunity to the learners to gain knowledge in a particular domain or subject. During recent days, MOOCs grabbed the attention of learners and provide quality education for free. As many MOOC providers are offering same courses, Choosing the best one out of all available MOOCs become very challenging task. MOOCs Evaluation plays an important role in finding the sentiment of the learners through the reviews and helps the MOOC Providers to improve the curricular quality. This paper proposed **MOOC-RoBERTa**, a sentiment analysis architecture that can evaluate MOOCs using Student reviews. Firstly, we prepared a balanced MOOC Reviews Dataset containing 13200 reviews. Secondly finetuned a RoBERTa Model for sentiment analysis on MOOCs. Next, trained the proposed model by applying Adversarial attacks approach to gain domain specific knowledge and test the model to find the sentiment of the learners. Finally, we compared the performance of the proposed model with different variants of other transfer learning models like BERT, Albert, XLNet. The experimental results demonstrate that the proposed model outshines the state-of-the-art methods by achieving the accuracy of 96.6% on MOOC Reviews Dataset.

Keywords-*MOOCs, Sentiment, Transfer learning, BERT, XLNet, Adversarial Attack, Accuracy*

1. INTRODUCTION

Massive Open Online Course (MOOCs) is a popular online learning platform that enables the students to learn the skills and knowledge from very experienced professors from reputed universities all over the world. MOOCs put forward to break the huddles faced in traditional classroom learning and motivated the students to adopt blended learning. Recent pandemic has made online teaching and learning process essential in the student's curriculum [1]. Fast growing technologies, high speed and flexible internet, massive usage of smart phones in connection with increase in demand for high quality and efficient education platforms, are motivating the growth of the MOOCs Course market. Besides that, Different countries across the world are requesting education institutions like universities, schools to adopt and foster the MOOCs to give maximum benefits to the students [2]. As the demand for MOOCs got increased, All Standard MOOC providers offering the same courses based on huge demand. Finding the efficient MOOC course out of all available courses become a difficult task. To address this problem, extensive research has been carried out and sources said that sentiment analysis is the best approach for finding the sentiment of the learners on a particular MOOC Course. Sentiment analysis can be done on MOOCs using two different sources of data, one is Discussion forum posts and other one is MOOC Course reviews. Discussion forum posts are the discussions between the students and the faculty whereas Course reviews are the opinions of the learners expressed on a particular course. From the Literature it has been observed that MOOCs Evaluation using course reviews gave good results when it compared with forum posts because a smaller number of contextual words in forum posts [3,7].

In recent days, Transfer Learning algorithms are performing better in finding the sentiment of the learners on MOOCs when it compared with traditional Machine Learning models and Deep Learning Models [4,5].

Adversarial training is an approach to train a classification model to classify both the original text examples and adversarial example. It improves

the robustness and classification performance of model [11].

The process of training a model correctly to classify both the original examples and adversarial examples is called Adversarial Training. Transfer learning models like BERT, RoBERTa combined with Adversarial training give good results in text classification task [6].

Robustness of the Pretrained Models in sentiment analysis task can improve the performance. Our proposed research work found the new way to increase the robustness the pretrained model and classify the opinion of the learners on a specific MOOC course in a better way.

The Major contributions of our research work are as follows.

- Identified and collected different MOOC reviews from various MOOC platforms. Preprocessed and annotated the collected reviews and prepared a balanced dataset of size 13200 reviews.
- Fine-tuned the pretrained model RoBERTa on the Created MOOC Dataset yielding MOOC-RoBERTa to classify the MOOC reviews.
- Adversarial Training has been adopted to train our MOOC-RoBERTa model that can correctly classify both original and adversarial reviews.
- Performed the testing on the MOOC-RoBERTa model using our MOOC reviews to find the performance.
- proposed model experimental results are compared with different variants of other transfer learning algorithms like BERT, XLNet, Albert using our MOOCs reviews dataset.

2. RELATED WORK

Wei, Lin et al proposed Convolutional -LSTM, the sentiment analysis system to predict the sentiment of the learners on a particular MOOC using cross domain MOOCs forum posts dataset. the performance of Convolutional -LSTM is better compared to other conventional Machine Learning Models [1].

Even though MOOCs got more popularity online in recent days, The serious problem in the MOOCs environment is Dropouts. As the Student's dropout rate is gradually increasing, many researchers perusing research to find the factors that are affecting the Attendance of the learners by evaluating the MOOCs. [1,6]

Yi Gao proposed a PNN model for sentiment analysis on the MOOCs discussion forum posts. The Author introduced a self-attention mechanism to that can find the key features automatically in a MOOC post. In PNN, a self-attention mechanism along with CNN and LSTM ware used parallelly to find the sentiment and finally obtain the accuracy of 91.29% [7].

Cheng proposed an ensemble transfer learning a model named ALBERT-BILSTM for analysing the sentiment of learners using MOOC Course Reviews. Firstly, ALBERT was used for word embedding and generated the word vectors and then BiLSTM is used to find the contextual feature vectors and applied the attention mechanism is applied to calculate the weight of each word in a sentence to identify the best context words that are suitable for sentence classification. The results shown that the ALBERT-BILSTM model has given the accuracy of 91.6% [8]. Yinhan Liu and other researchers from Facebook introduced an advanced version of BERT called Robustly optimised BERT (RoBERTa) a pre trained modal that was trained on huge data with large batches and longer sequences. They also added the approach of dynamically changing the Masking pattern. Because of these changes RoBERTa can perform well compared to BERT [9].

Xiang Li et al proposed a shallow BERT-CNN model as a classifier that contains 6 layers of BERT, a CNN layer and a self-attention layer is used in architecture. MOOC reviews dataset is used to train the model and calculated the performance. BERT-CNN gave 81.3% of test accuracy and 92.8% of F1 score [10].

Di Jin, Zhijing Jin developed TextFooler, a very efficient Adversarial text generator that can be used rapidly used for text classification task. Adversarial Attacks plays vital role in improving the text classification performance of the model [11,12,15].

Jin Yong Yoo, Yanjun Qi introduced improved version of vanilla adversarial training approach called Attacking to training(A2T). It is mainly focused on word substitution attack method to increase the robustness of training. The A2T model was used to train BERT, RoBERTa models on IMDB,Yelp datasets. Experimental results proven that A2T is best suitable for cheaper adversarial training of robust NLP models and can improve NLP models' accuracy [13,14].

Hassan Ali et al compared the robustness of four deep learning algorithms namely, MLP, CNN, RNN, CNN-RNN by applying Adversarial

Training approach. The experimental results shown that RNN is the best and robust model for detecting the fake news. It is also observed that robustness is purely depends on the input sequence length [16].
Wei Zhang et al aimed to generate readable adversarial text and proposed a novel approach that can generate it with some perturbations that can also confuse human observers successfully [17].

3. Background

3.1. Sentiment Analysis

Sentiment Analysis (SA) is the process of finding the opinion of the customers/users on a particular product or service using text review or posts. In a business platform, it is the process of mining the context of the text review/post that can find and capture subjective information and helps a business to know the opinion of their service/ product/brand.
SA helps in identifying the opinion of the listeners on a particular MOOC to find the best one out of all available MOOCs and also helps the MOOC Providers to improve the quality of the MOOC. MOOCs Evaluation using Sentiment analysis can be done using two different text sources, one is using MOOCs Discussion Forum Posts and the other one is using Course Reviews. In our proposed work we used second one i.e., Course reviews to evaluate the MOOCs.

3.2. Transfer Learning

It is a sub field of machine learning that focused mainly on storing knowledge that is gained while doing a task and uses that knowledge in another relevant task. The models that are trained using Transfer Learning are called as Pretrained Models. It can be implemented to many NLP downstream tasks like Sentiment Analysis, Question Answering etc.

In our proposed work, For MOOCs Evaluation we use transfer learning approach and build the MOOC- RoBERTa model and also compared our proposed MOOC-RoBERTa with some of the other transfer learning algorithms like BERT, Albert, XLNet, RoBERTa. The detailed information about the pretrained models are as follows.

3.2.1. BERT

BERT means Bidirectional Encoder Representation from Transformers. It is a pretrained model from transformers. It uses Next Sentence Prediction and Masked Language Model to perform various NLP tasks like Sentiment analysis. It has two variants BERT $_{base}$, BERT $_{Large}$.
BERT $_{base}$ uses, 12 Encoders, 768 attention dimensions, 12 attention heads, 110M parameter neural network architecture. BERT $_{Large}$ uses 24 encoders, 1024 attention dimensions, 16 attention heads and 340M parameter neural network architecture.

3.2.2. Albert

ALBERT is also called as a light BERT, other variant of BERT. It is an encoder-decoder model at which self-attention is implemented at the encoder end, attention on encoder outputs is applied on decoder end. Albert uses four different techniques to improve the performance. Factorization of Embedding matrix, parameter sharing are some of them. Albert has four variants namely ALBERT $_{base}$ contains 12 encoders, ALBERT $_{Large}$

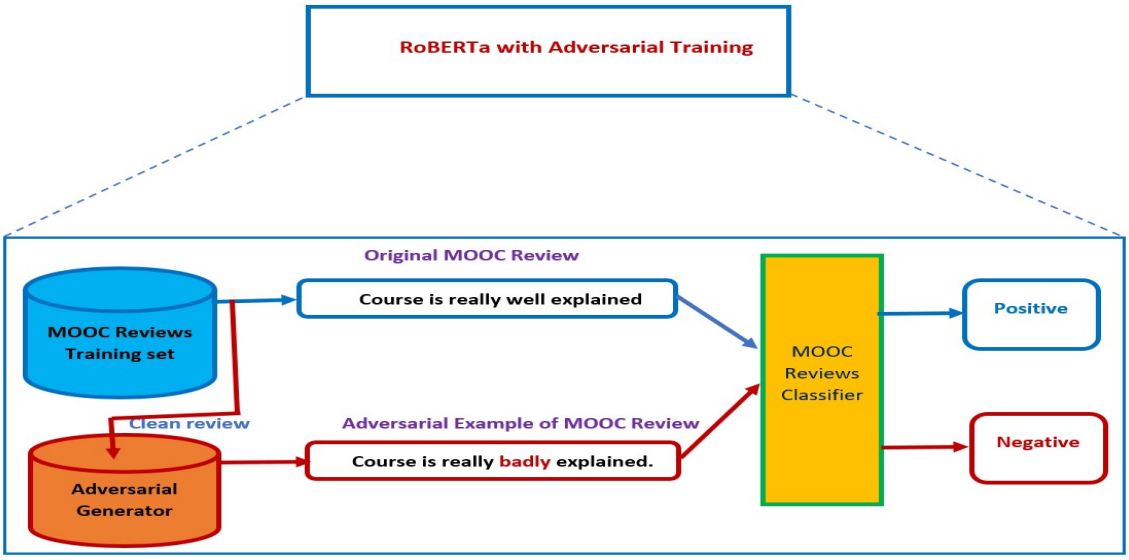


Fig. 1. Adversarial Training Mechanism in MOOC-RoBERTa that generates Adversarial examples using Adversarial attack along with the original input for comparison. Perturbed words in adversarial example are highlighted in red.

Journal of Theoretical and Applied Information Technology
31st May 2023. Vol.101. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2.3. XLNet

XLNet is also a pretrained model that was generated by integrating auto regression from Transformer-XL and auto encoders from BERT. XLNet readopts Permutation Language Model (PLM) that can learn contextual information of a given text in both the sides. Next sentence prediction mechanism was removed in XLnet. This model trains well by using PLM and gives state of the results in some of the NLP downstream tasks like sentient analysis.

3.2.4. RoBERTa

RoBERTa is a polished successor of BERT. It uses the same architecture of BERT. This model optimizes some of the hyper parameters available in BERT and obtained the state-of-the-art results. The model optimizes the following hyper parameters [9]

- Training data is increased 10 times i.e. (from 16GB to 160 GB)
- Used Dynamic masking instead of Masked Language Modelling
- As the Training data got increased, Training time also increased
- Batch size increased from 256 to 8k.
- Removed the Next Sentence Prediction Task that is available in BERT.
- vocabulary size increased from 30k to 50k.
- Longer sequences are used as input

to **a ′**, while a victim DNN would have high confidence on wrong prediction of **a ′. a′** can be formalized as:

$$\mathbf{a' = a + \mu, f(a) = b, a \in X}$$
$$\mathbf{f(a') \neq b}$$
$$\mathbf{or\ f(a') = b', b' \neq b,}$$

where μ is the worst-case perturbation. The goal of the adversarial
attack can be deviating the label to incorrect one **(f (a ′) ≠ b)** or specified one **(f (a ′) = b ′).**

It is important to pick the effective adversarial examples to train the model so that model can able to generalize both the original and adversarial examples and become robust. Adversarial training mechanism is presented in **Fig.1** above

3.3.2. Adversarial Attack

A program that repeatedly generates Adversarial examples for a model using input example is known as Adversarial Attack. An Adversarial Attack finds the perturbation to create adversarial example that can mislead the neural network model to do misclassification [21-23].

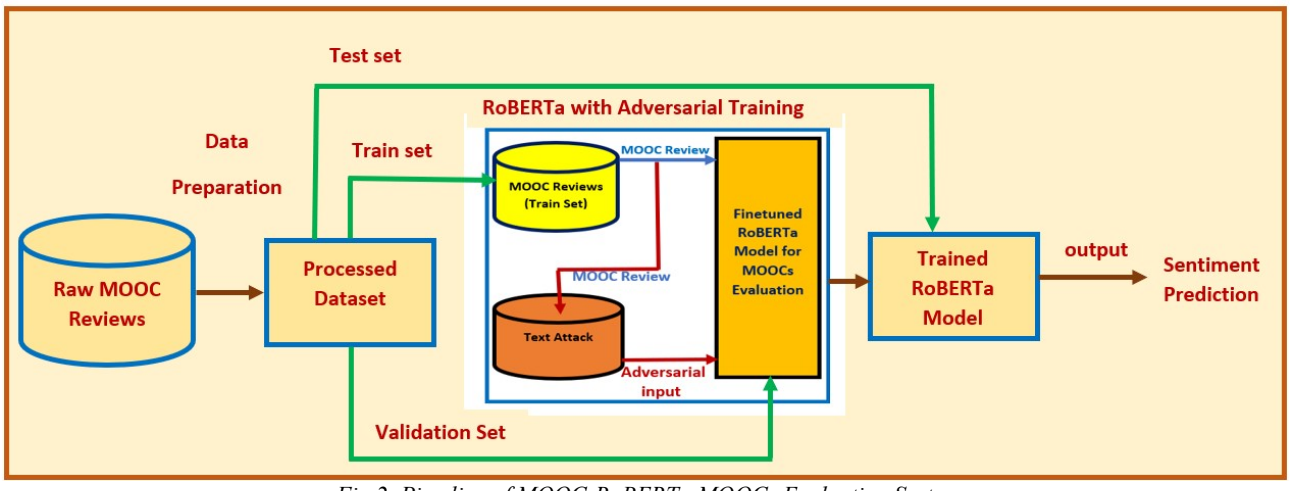


Fig.2. Pipe line of MOOC-RoBERTa MOOCs Evaluation System

4. PROPOSED SYSTEM

The MOOC-RoBERa is a MOOCs Evaluation system that is designed by combining basic RoBERTa architecture with Adversarial Training mechanism and apply training mechanism such that the model can classify original and adversarial examples correctly. An Adversarial Training mechanism used in our proposed system improves the robustness of the model. Our proposed model used

TextAttack framework to generate adversarial attacks that can improve the classification accuracy. The pipeline of MOOCs Evaluation system using MOOC-RoBERTa is presented below in Fig. 2.

4.1. Data Preparation

Data Preparation for our proposed MOOCs Evaluation system is done in different steps. The Data preparation flow is presented in **Fig.3.**

Collect Data: Identified and gathered around 20000 MOOC reviews from various online MOOC platforms like edX, Coursera, Udacity, Udemy.

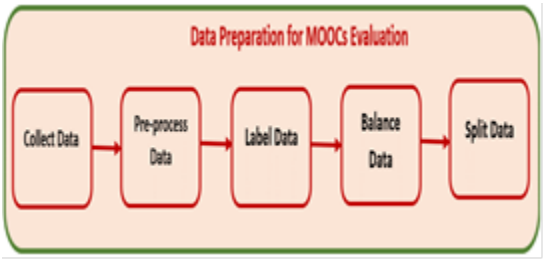


Fig.3. Data Preparation for MOOC-RoBERTa Model

Pre-Process Data: Collected reviews have been pre-processed to remove irrelevant information such as special characters, HTML Tags etc ...

Label Data: Assign labels (0 or 1) to each review by understanding the sentiment of that review. We gave label **0** for the review with negative sentiment and **1** for Positive sentiment.

Balance Data: To prevent the model from being biased towards a particular sentiment, we need to balance the data. To prepare balanced dataset, we have selected 13200 labelled reviews Out of which 6600 positive and 6600 negatives.

Split Data: Split the dataset into train, test, validation sets. The training set contains 70% of total reviews i.e., 9240 Reviews (4620 positive and 4620 negative) and the validation set contains 10% of total reviews i.e., 1320 Reviews (660 positive and 660 negative) and the test set contains 20% of total reviews i.e., 2640 Reviews (1320 positive and 1320 negative) Training set is used to train the model, validation set is used to finetune hyper parameters of a model and Test set is used to evaluate the performance of the model.

Now the Dataset in the form of train, validation and test set is ready for experimentation. Decomposition of MOOCs Dataset is presented in

Table.1. Mooc Reviews Dataset Description

Total Reviews	Trainset	ValidationSet	Testset
13200	9420(70%)	1320(10%)	2640(20%)

4.2. Adversarial Training in Proposed Model

Adversarial Training is implemented in the proposed model using TextAttack framework. It offers different attack generation methods that can perform very effectively and improves the robustness of the model. It uses four basic components to construct adversarial attacks. They are Transformation, Constraints, Goal Functions, Search Method. The functionality of each module in generating Adversarial Attack is presented below.

4.2.1. Transformation

it is a component in the TextAttack framework that generates a set of potential perturbations for a given input text sentence or review. To perform transformation, The textattack offers many techniques in terms of modules to build transformations. In our Proposed system to build transformations we used **word embedding word swap** module.

4.2.2. Constraints

Constraints identify the perturbation that is suitable for given input text review and filters out the bad perturbations that are generated using transformation. In proposed system we used **Maximum word embedding distance** as a measure to identify the bad perturbations.

4.2.3. Goal Function

It is used to find whether the attack generated is successful in terms of model output. It tells us when the model was
successfully fooled. Targeted classification is the best approach to do this.

4.2.4. Search Method

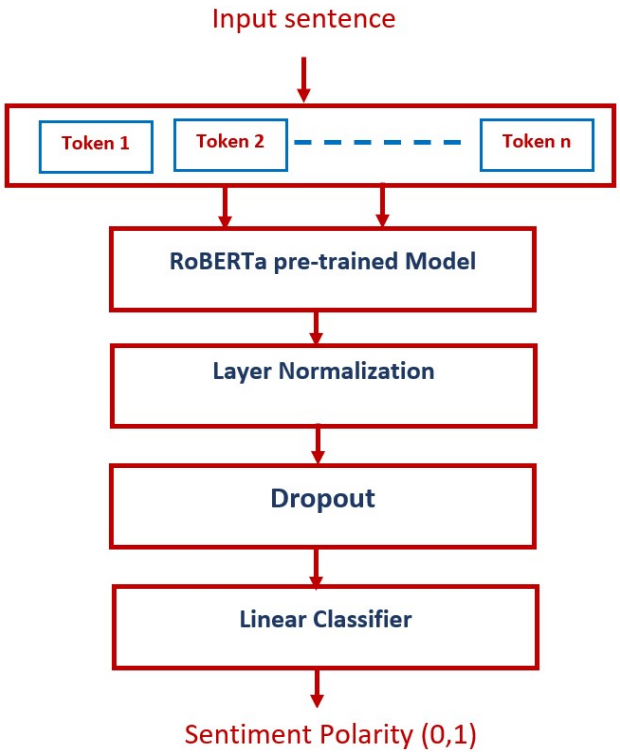
It asks the model and finds successful perturbations from a set of transformations. In the proposed system **Genetic Algorithm** is used to do this job.

4.3. MOOC-RoBERTa Architecture

In order to perform MOOCs evaluation, The MOOC RoBERTa architecture is designed based a pretrained RoBERTa Transfer learning model. To build our proposed model, we upgraded the RoBERTa base model by adding three other functionalities in three consecutive layers. First one is **Layer Normalization;** that allows the model to speed up the training process, improves generalization accuracy. Second is **Dropout Layer** that allows to drop the node in the neural network and avoids overfitting issues. Third one is **Linear Layer;** it is also called as fully connected layer that is used for the classification of MOOC reviews into positive and negative.

The transfer learning approach is adopted to our model to gain more domain knowledge and can perform classification task effectively. Different variants of other pretrained language models like BERTbase, BERT Large, XLNet Base, XLNet Large, Albert were used to pre-process the input text and finetune pretrained models to perform text classification. The detailed architecture of MOOC-RoBERTa is presented below in **Fig.4.** Proposed architecture contains 12 hidden layers, 12 attention heads, 1 Normalization layer, 1 dropout layer, 1 linear layer and the hidden size is of 768.

Fig.4. MOOC-RoBERTa Architecture



5. Implementation

5.1. Environment Setup

All Experiments are carried out using the **Google Colab** runtime environment along with GPU support. **Hugging face Transformers (V4.24.0)** API is used to access pretrained models. The **PyTorch 2.0** framework is used for implementing the proposed model. **Python 3.11** is used as base. TextAttack 0.3.4 was used for implementing adversarial training in our proposed model.

Table.2: Hyper parameter setting for MOOC-RoBERTa

Parameters	Tuned Range	optimized
Max Sequence Length	**[50, 100, 150, 200]**	**150**
epochs	**[3, 5, 10]**	**3**
batch size	**[16, 32, 64]**	**32**
learning rate	**[2E-5, 2E-3, 2E-4]**	**2E-05**
Optimizer	**AdamW**	**AdamW**
Dropout	**[0.1, 0.2, 0.3]**	**0.1**
max_grad_norm	**10**	**10**
Activation Function	**Tanh, Softmax**	**Tanh,Softmax**

Table.3. Classification Results of different Pretrained models from Transformers.

Model	Accuracy	Precision	Recall	F1 score
BERT base	0.941	0.943	0.941	0.941
Albert base	0.931	0.933	0.931	0.931
XLNet_Base_cased	0.956	0.958	0.956	0.956
RoBERTa_Base	0.951	0.953	0.951	0.951
BERT_Large	0.951	0.951	0.951	0.951
Albert -Large	0.94	0.941	0.94	0.94
XLNet_Large_cased	0.958	0.958	0.958	0.958
BERT_Text Attack	0.953	0.954	0.953	0.953
Albert_TextAttack	0.946	0.947	0.946	0.946
XLNet_TextAttack	0.959	0.961	0.959	0.959
Proposed Model (MOOC-RoBERTa)	**0.966**	**0.966**	**0.966**	**0.966**

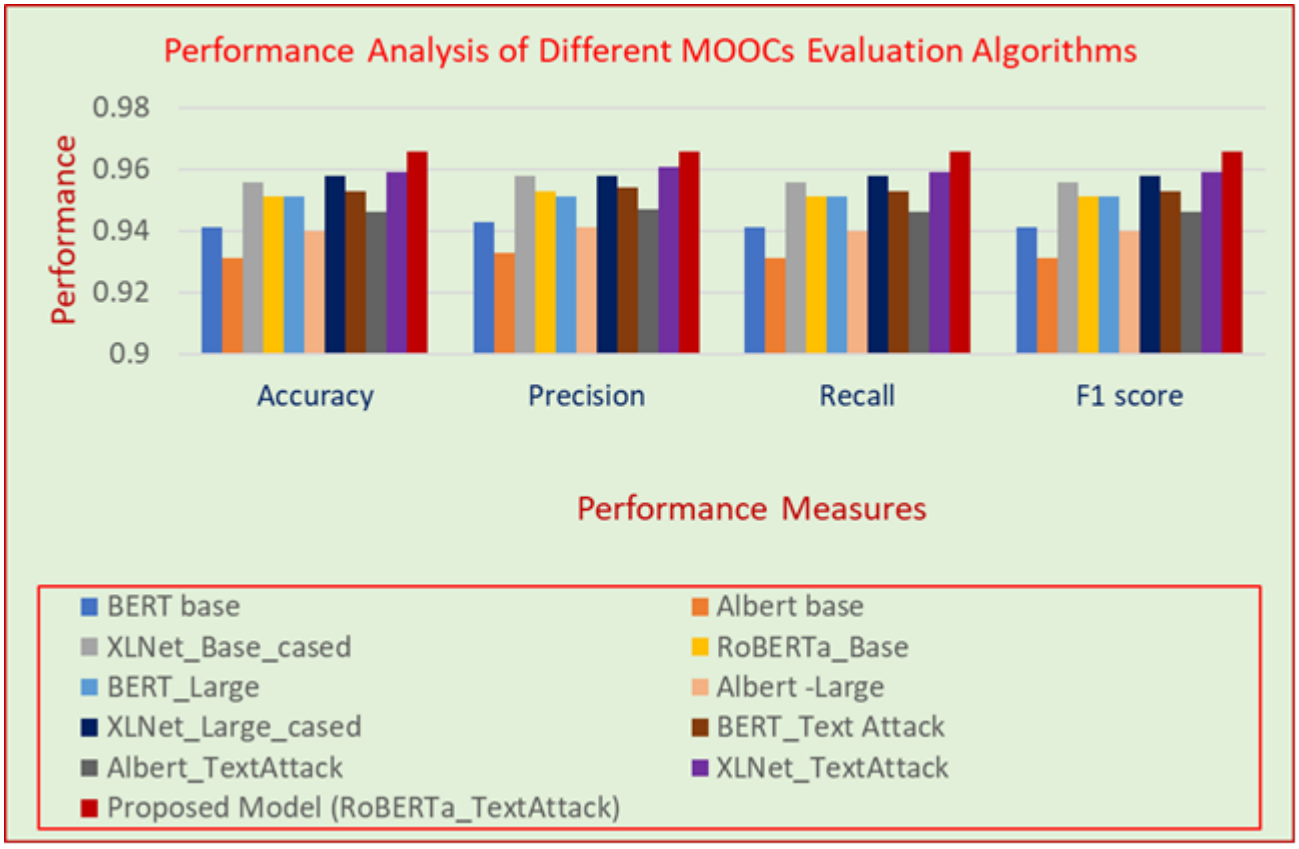


Fig.6. Performance Analysis of Different Pretrained Transfer Learning Models for MOOCs Evaluation.

3.1. Performance Measures

To evaluate the performance of our proposed MOOC-RoBERTa model and to compare the performance of the proposed model with other pretrained models, we use precision, recall, F1 Score and accuracy as measures and the detailed information about these measures are presented in Fig.5. .

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Fig.5. Performance Measures for Classification

3.2. Experimentation

Initially MOOC Reviews from MOOCs Dataset were pre-processed and divided in to tokens. Then some special tokens were added to separate the tokens based the model we use. Padding and truncation techniques were used to maintain fixed length of tokens for the entire input batch that is suitable for a particular model. Finally, each input text tokens are encoded into input ID that can represent the input word.

We finetuned our proposed model by changing the values of different hyper parameters and find the optimum results.

The details of the hypermeters used, range of parameters tried and the optimal hypermeters used to train the model are presented in **Table.2** above. We used these optimal hyper parameters to compare the performance of proposed model with other pretrained language model like BERT, XLNet, Albert.

3.3. Results and Analysis

Many Experiments are carried by changing the hyperparameters to train the proposed model. Finally proposed model gave the highest accuracy of 96.6% by using the following optimal hyperparameters. Those are maximum sequence length -150, The number of Epochs is 3, batch size 32, Learning rate 2E-05, optimizer is AdamW, dropout value of 0.1, Activation Function are Tanh, Softmax.

To compare the performance of MOOC-RoBERTa with other pertained models, same parameters have been used and the performance analysis of these models along with MOOC-RoBERTa are presented in **Table.3**.

From all the experimental results, it is observed that out of all pretrained base models, XLNet $_{base\_cased}$ model was performed well and given the accuracy of 95.6%. From all the large models, The $XLNet_{large\_cased}$ projected highest accuracy of 95.8%. After applying Adversarial Attack on different pretrained language models, accuracy was increased in all models. the accuracy of the models with TextAttack are as follows. 95.3% in $BERT_{base\_TextAttack}$, 94.6% in $Albert_{base\_TextAttack}$, 95.9% for XLNet $_{TextAttack}$ finally MOOC-RoBERTa $_{TextAttack\ projected}$ the highest accuracy of **96.6%.** The performance analysis of all the pretrained Transfer learning algorithms for MOOCs Evaluation is projected using a Bar graph in **Figure.6**. In the bar graph, Accuracy, Precision, Recall, F1 Score values shows that the pretrained RoBERTa with TextAttack is performed well compared to all other pretrained model

4. CONCLUSION AND FUTURE WORK

Transfer Learning algorithms are performing well for evaluating sentiments of learners on a particular MOOC. In this paper we proposed the MOOC-RoBERTa model by finetuning the pretrained RoBERTa base model and added Adversarial training mechanism to increase the robustness. This integration increased the robustness of the proposed model and could perform the classification task very well compared to other pretrained models. This is the best model for evaluating the MOOCs using various MOOC Reviews. The proposed model projected the state-of-the-art results in the field of MOOCs Evaluation.

Integration of Graph Neural Networks with pretrained embedding techniques can improve the performance of Language understanding models for Classification task.

REFERENCES

[1] X. Wei, H. Lin, L. Yang, and Y. Yu, ‘‘A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification,’’ Infor mation, vol. 8, no. 3, p. 92, Jul. 2017.

[2] Massive Open Online Course Market Analysis - Industry Report - Trends, Size & Share -mordorintelligence.com.

[3] Manuel J. Gomez” Large scale analysis of open MOOC reviews to support learners’ course selection” in Expert systems with Applications -Elsevier Journal 30 December 2022.

[4] Nusrat Jahan “Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning” in Artificial Intelligence in Sensors **2022**, *22*(11), 4157,

MDPI, 30 May 2022.
[5] Devlin, Chang, et al "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv* **2018**, arXiv:1810.0480, ACL, 2018.
[6] Takeru Miyato, Andrew M Dai "Adversarial Training Methods for semi-supervised Text Classification", ACL 2021.
[7] Y. Gao et al, "A parallel neural network structure for sentiment classification of MOOCs discussion forums", IEEE Access Journal of Intelligent & Fuzzy Systems. 4915–4927 (2020)
[8] Cheng Wang1, Sirui Huang2" Sentiment analysis of MOOC reviews via ALBERT-BiLSTM model" MATEC Web of Conferences 336, 05008 (2021) https://doi.org/10.1051/matecconf/202133605008 CSCNS2020
[9] Yinhan Liu, Myle Ott "RoBERTa: A Robustly Optimized BERT Pretraining Approach" ACL 2019 (Facebook).
[10] Xiang Li et al "A Shallow BERT-CNN Model for Sentiment Analysis on MOOCs Comment" University of New South Wales, IEEE November, 2020.
[11] WEI EMMA ZHANG, QUAN Z. SHENG, "Adversarial Attacks on Deep Learning Models in Natural Language Processing: A Survey" ACM, , Vol. 1, No. 1, Article , 2019, arXiv:1901.06796v3
[12] Di Jin, Zhijing Jin "Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment" AAAI, 2020, arXiv:1907.11932v6.
[13] Jin Yong Yoo, Yanjun Qi "Towards Improving Adversarial Training of NLP Models" AAAI, 2021.
[14] Takeru Miyato et al "Adversarial training methods for semi supervised text classification" Google Brain, ATR Cognitive Mechanisms Laboratories, Kyoto University
[15] Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2015. Explaining and Harnessing Adversarial Examples. In Proceedings of the 3rd International Conference on Learning Representation (ICLR). San Diego.
[16] Ali et al.: "Evaluating Adversarial Robustness of Fake-news Detectors Under Black-Box Settings" DOI 10.1109/ACCESS.2021.3085875, IEEE Access, VOLUME 4, 2016.
[17] W. Zhang et al.: "Deep Learning Based Robust Text Classification Method via Virtual Adversarial Training" Digital Object Identifier 10.1109/ACCESS.2020.2981616, IEEE Access, 2020.
[18] Tao Bai, Jinqi Luo et al "Recent Advances in Adversarial Training for Adversarial Robustness" International Joint Conference on Artificial Intelligence (IJCAI-21) arXiv:2102.01356.
[19] Silva, S.H.; Najafi ad, P. "Opportunities and Challenges in Deep Learning Adversarial Robustness: A Survey". arXiv 2020, arXiv:2007.00753.
[20] PanelBen Rodrawangpai, Witawat Daungjaiboon "Improving text classification with transformers and layer normalization" Machine Learning with Applications, Volume 10, 15 December 2022, 100403, https://doi.org/10.1016/j.mlwa.2022.100403.
[21] Abdullah Al-Dujaili, Alex Huang, Erik Hemberg, and Una-May O'Reilly. 2018. Adversarial Deep Learning for Robust Detection of Binary Encoded Malware. In Proc. of the 2018 IEEE Security and Privacy Workshops (SPW 2018). Francisco, CA, USA, 76–82.
[22] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, ''A CNN-BiLSTM model for document-level sentiment analysis,'' Mach. Learn. Knowl. Extraction, vol. 1, no. 3, pp. 832–847, 2019.
[23] Jin Yong Yoo and Yanjun Qi. 2021. Towards improving adversarial training of nlp models. arXiv preprint arXiv:2109.00544.
[24] Yisen Wang, Difan Zou, Jinfeng Yi, James Bailey, Xingjun Ma, and Quanquan Gu. 2019. Improving adversarial robustness requires revisiting misclassified examples. In International Conference on Learning Representations.
[25] Chen Zhu, Yu Cheng, Zhe Gan, Siqi Sun, Tom Gold stein, and Jingjing Liu. 2019. Freelb: Enhanced adversarial training for language unde.