IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, October 2022

Sentimental Analysis using Bert Algorithm over LSTM

Mohammed Athar Rangila[1], Soham Khandke[2], Yash Mohite[3], Komal Kamble[4].
Students, Department of Computer Engineering[1,2,3,4]
All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, India

Abstract: *Sentiment analysis also referred to as opinion mining, is an approach to natural language processing (NLP) to find out whether the meaning of the given data is positive or negative. There has been a lot of research done in this sector to increase the accuracy of sentiment analysis systems, ranging from basic linear models to complicated neural network models. Other models or algorithms were checked previously but these took longer time for operations and had low accuracy scores overall. So, in this paper, the BERT model, which has been pre-trained on a huge corpus, is offered to address concerns with sentiment analysis systems. Further fine-tuning improves results according to the use cases. The BERT model gives high accuracy, and efficient performance with good predictions as compared to other models in experimental testing. We are particularly interested in performing sentiment mining techniques over transcribed audio recordings to detect the speakers' sentiment using a Google specialised algorithm BERT (Bidirectional Encoder Representations from Transformers), which is a faster and more accurate algorithm for text fields because it can also consider the context of the text. Our study was initially driven by call centre use cases, but it can potentially have applications in other domains such as security and other areas. So in this research project work we will be giving input as an audio file from which the text will be generated, and out of which we will be finding the sentiment of that audio stream by using the BERT model.*

Keywords: LSTM, BERT, Sentimental Analysis, NLP, contextual learning

I. INTRODUCTION

Sentiment analysis (SA) is an important task in natural language processing and is broadly used in industry. People study subjective sentiment information in articles and reviews to examine product sales, service tactics, and literary trends. Amazon and Tmall, for example, provide automated fine-grained review analysis services. The BERT Algorithm, a bidirectional algorithm, is used to explain and predict public emotional responses. To be specific, we fine-tune BERT for sentiment classification upon statements with two potential categories of sentiment, positive, and negative, getting high accuracy. The percentage of each sentiment category is computed after sentiment categorization.

Bidirectional Encoder Representations from Transformers (BERT), a machine - learning approach for pre-training in natural language processing, has been widely used in sentiment analysis. The BERT model may be fine-tuned with appropriate input and output layers to produce cutting-edge results.

However, most of these learning models have traditionally concentrated solely on the substance of expression, ignoring contextual information. As a result, they failed to capture the context of emotion. Using BERT we propose a system where the context of a statement is being taken into consideration and then it's categorized as positive or negative.

II. METHODOLOGY

2.1 Bert Algorithm

BERT is an open source machine learning framework for natural language processing. BERT is intended to assist computers in understanding the meaning of ambiguous words in the text by establishing context from surrounding content. BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model in which each output element is connected to each input element and the weightings between them are calculated dynamically based on their connection. Historically, language models could only interpret text input sequentially, either left-to-right or right-to-left, but not both. BERT is unique in that it can read in both directions at the same time.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

2.2 How Bert Works

The purpose of any given NLP approach is to comprehend real human language. In the context of BERT, this usually implies anticipating a word in a blank. To do this, models are frequently trained using a massive pool of specialised, labelled training data. Manual data tagging by linguist teams is required.

BERT, on the other hand, was trained using simply an unlabeled plain text dataset (namely the entirety of the English Wikipedia, and the Brown Corpus). It continues to learn and improve unattended from unlabeled text. As previously stated, BERT is made feasible by Google's Transformer research. The transformer is the component of the model that allows BERT to recognize context and ambiguity in language. The transformer does this by analyzing every given word in connection to all other words in a phrase, rather than one at a time. The Transformer helps the BERT model to grasp the whole context of the word by looking at the surrounding words, allowing it to better understand the searcher's intent.

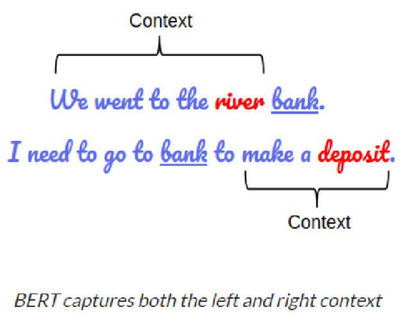

Fig.1. Context Classification in BERT.

2.3. LSTM Concept

LSTM is an abbreviation for long short-term memory networks, which are utilised in Deep Learning. It is a kind of recurrent neural networks (RNNs) that may learn long-term dependencies, particularly in sequence prediction tasks. Apart from single data points such as photos, LSTM contains feedback connections, which means it can process the complete sequence of data. This is used in speech recognition, machine translation, and other areas. LSTM is a kind of RNN that performs exceptionally well on a wide range of tasks.

2.4 The Logic Behind LSTM

A memory cell known as a 'cell state' that maintains its state throughout time plays the fundamental function in an LSTM model. The cell state is shown by the horizontal line that runs across the top of the picture below. It may be seen as a conveyor belt on which information just passes, unmodified.. Data can be added to or deleted. The in LSTM is derived from the cell state and is controlled by gates.These gates allow information to enter and exit the cell. To aid the mechanism, it features a pointwise multiplication operation as well as a sigmoid neural net layer.

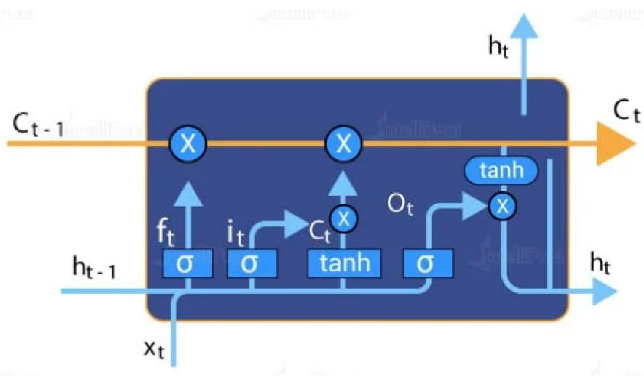

Fig.2. LSTM Architecture.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 1, October 2022

Impact Factor: 6.252

2.5 Disadvantages of LSTM.

It is seen that LSTMs cannot perform well in classification tasks for larger datasets. LSTSMs also take quite huge amounts of time for training purposes. It also requires more amount of memory than transformers. When compared to other recent models, LSTMs appear to be more prone to overfitting.

The epochs required by LSTMs are greater in number as compared to transformers. Even when LSTMs require more epochs they cannot give the accuracy up to the mark as compared to transformers and other modern models.LSTMs work on data in only one direction, from left to right, resulting in inadequate learning of the real meaning and context of the data provided. LSTMs are not pretrained like BERT, so for some specific use cases, this training task should be performed additionally. LSTMs follow the old conventional language model which uses previous tokens to predict the next tokens. LSTMs are not specialized in next sentence prediction and other tasks which may not be an appropriate choice for some applications or use cases.

III. IMPLEMENTATION

3.1 Data Pre-Processing

Tokenization is the first stage in preprocessing. In the tokenization process, each text data is separated into smaller components, either words or phrases. The NLTK library may be used to execute tokenization operations. The next step is to get rid of any unnecessary information.

The bulk of the data contains some noise, necessitating a preprocessing step to eliminate the unwanted elements. Stop words, punctuation, and special characters, for example, can be found in data.. As a result, any elements that do not contribute to the classification objective are removed prior to the feature extraction procedure. By deleting the undesired things from the text contents, a data set may be generated.

3.2 Classification

Deep learning is the process of developing an optimum solution from an algorithm to solve a typical machine learning problem by learning a deep representation of data. Deep learning is an advanced learning strategy that outperforms the capacity to recognise a word representation of data in each test.

It can automatically extract fresh features from the training data's shifting sets of characteristics without the need for human intervention. In other words, it extracts more features when the collection lacks labels.

IV. RESULT DISCUSSIONS

BERT vs LSTM

4.1 Efficiency of the Algorithm

We used a dataset of 50,000 reviews. Out of which 25,000 were positive and 25,000 were negative. To provide a fair assessment, we utilised the same dataset for both the BERT Model and the LSTM Model. Here are the findings we obtained.

```
Model: "tf_bert_for_sequence_classification"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 bert (TFBertMainLayer)      multiple                  109482240

 dropout_37 (Dropout)        multiple                  0

 classifier (Dense)          multiple                  1538

=================================================================
Total params: 109,483,778
Trainable params: 109,483,778
Non-trainable params: 0
_________________________________________________________________
```

Fig.3. Number of the trainable parameters in BERT.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
IJARSCT
Impact Factor: 6.252

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 200, 100)          300000

bidirectional (Bidirectiona  (None, 128)               84480
l)

dense (Dense)                (None, 24)                3096

dense_1 (Dense)              (None, 1)                 25

=================================================================
Total params: 387,601
Trainable params: 387,601
Non-trainable params: 0
```

Fig.4. Number of the trainable parameter in LSTM

There is a significant difference in the amount of parameters divided by the BERT Algorithm vs the LSTM Algorithm. This helps us understand how efficiently sentimental analysis occurs in the BERT algorithm, which is one of the main reasons we chose BERT as our Sentimental Analysis system.

4.2 Accuracy of the Algorithm

```
num_epochs = 5
history = model.fit(train_padded, train_labels,
                    epochs=num_epochs, verbose=1,
                    validation_split=0.1)

Epoch 1/5
1055/1055 [==============================] - 103s 90ms/step - loss: 0.6935 - accuracy: 0.5031 - val_loss: 0.6933 - val_accuracy: 0.4888
Epoch 2/5
1055/1055 [==============================] - 94s 89ms/step - loss: 0.6928 - accuracy: 0.5114 - val_loss: 0.6924 - val_accuracy: 0.5107
Epoch 3/5
1055/1055 [==============================] - 94s 89ms/step - loss: 0.6924 - accuracy: 0.5157 - val_loss: 0.6926 - val_accuracy: 0.5117
Epoch 4/5
1055/1055 [==============================] - 93s 88ms/step - loss: 0.6924 - accuracy: 0.5195 - val_loss: 0.6921 - val_accuracy: 0.5131
Epoch 5/5
1055/1055 [==============================] - 93s 88ms/step - loss: 0.6923 - accuracy: 0.5180 - val_loss: 0.6921 - val_accuracy: 0.5155
```

Fig.5. Epochs of LSTM

```
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=3e-5, epsilon=1e-08, clipnorm=1.0),
              loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
              metrics=[tf.keras.metrics.SparseCategoricalAccuracy('accuracy')])

model.fit(train_data, epochs=2, validation_data=validation_data)

Epoch 1/2
1250/1250 [==============================] - 2148s 2s/step - loss: 0.0325 - accuracy: 0.9895 - val_loss: 0.4581 - val_accuracy: 0.8874
Epoch 2/2
1250/1250 [==============================] - 2096s 2s/step - loss: 0.0223 - accuracy: 0.9930 - val_loss: 0.6509 - val_accuracy: 0.8728
<keras.callbacks.History at 0x7f788eb82a10>
```

Fig.6. Epochs of BERT

Only after 2 epochs, BERT can produce an accuracy of 98% but we can see that after 5 epochs, LSTM is not even close to the accuracy of BERT. It was able to produce a 51% accuracy.

4.3 Contextual Learning

```
pred_sentences = [' Book contents are best ','Product was very bad",'It would have been great if it had more feature']

tf_batch = tokenizer(pred_sentences, max_length=128, padding=True, truncation=True, return_tensors='tf')
tf_outputs = model(tf_batch)
#print('Model output : ',tf_outputs[0])
tf_predictions = tf.nn.softmax(tf_outputs[0], axis=-1)
print('Predictions : ',tf_predictions)
labels = ['Negative','Positive']
label = tf.argmax(tf_predictions, axis=1)
print('label : ', label)
label = label.numpy()
for i in range(len(pred_sentences)):
  print(pred_sentences[i], ": \n\n", labels[label[i]])

Predictions :  tf.Tensor(
[[4.6773627e-03 9.9532264e-01]
 [9.9952567e-01 4.7433696e-04]
 [7.2036022e-01 2.7963981e-01]], shape=(3, 2), dtype=float32)
label :  tf.Tensor([1 0 0], shape=(3,), dtype=int64)
 Book contents are best  :

 Positive
Product was very bad :

 Negative
It would have been great if it had more feature :

 Negative
```

Fig.7. Result of BERT.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, October 2022

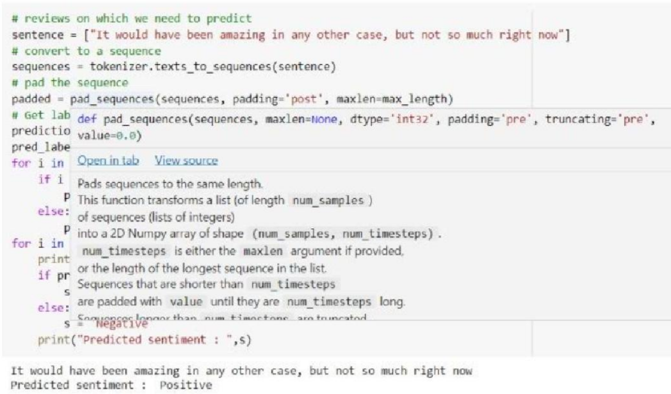

Fig. 8. Result of LSTM

When the same sentence was provided to the models one showed it as positive sentiment and one showed a negative sentiment.

The sentence: "It would have been amazing in any other case but not right now".

Since the word amazing is present in the sentence the LSTM algorithm is forced to respond to this as a positive statement as it doesn't consider the context. But BERT is a bidirectional algorithm and it does consider context so BERT responds to this with a negative statement which indeed it is.

V. CONCLUSION

The specified and expected activities and procedures were carried out successfully. The entire project is centred on developing a system that accurately determines the sentiment of an audio file. The implication if this would mean that we can know The hidden or true meaning of the audio and what it meant, this would help in easy knowledge gathering and will be time-saving for everyone.

Future Work: The proposed method can be extended to analyze multiple audio files under a parallel approach. Audios in different languages can be analyzed to extract sentiment. Various other features can be used and the method can be combined with other methods for improving the nature of sentiment analysis.

ACKNOWLEDGEMENT

The work carried out was successful and possible because of the guidance provided by our guide and also the facilities and support provided by our HOD, Principal, etc, thus we sincerely extend our gratitude and thank everyone for it.

REFERENCES

[1]. Z. Gao, A. Feng, X. Song, and X. Wu, "Target-Dependent Sentiment Classification With BERT," in IEEE Access, vol. 7, pp. 154290-154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[2]. T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," in IEEE Access, vol. 8, pp. 138162-138169, 2020, doi: 10.1109/ACCESS.2020.3012595.

[3]. X. Li et al., "Enhancing BERT Representation With Context-Aware Embedding for Aspect-Based Sentiment Analysis," in IEEE Access, vol. 8, pp. 46868-46876, 2020, doi: 10.1109/ACCESS.2020.2978511.

[4]. ˙I. Ayg¨un, B. Kaya and M. Kaya, "Aspect Based Twitter Sentiment Analysis on Vaccination and Vaccine Types in COVID-19 Pandemic with Deep Learning," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2021.3133103.

[5]. Y. Chen, L. Kong, Y. Wang and D. Kong, "Multi-Grained Attention Representation With ALBERT for Aspect-Level Sentiment Classification," in IEEE Access, vol. 9, pp. 106703-106713, 2021, doi: 10.1109/ACCESS.2021.3100299.

[6]. R. Abas, I. El-Henawy, H. Mohamed and A. Abdellatif, "Deep Learning Model for Fine-Grained Aspect-Based Opinion Mining," in IEEE Access, vol. 8, pp. 128845-128855, 2020, doi: 10.1109/ACCESS.2020.3008824.