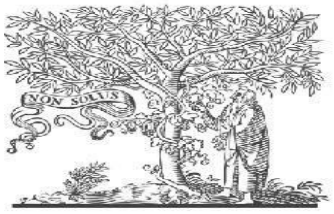
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 29th Aug 2021.
Link :http://www.ijiemr.com/downloads.php?vol=Volume-10&issue=ISSUE-08
DOI: 10.48047/IJIEMR/V10/I08/19

Title:- **A PRE-TRAINED MODEL BERT FOR MACHINE TRANSLATION FROM ENGLISH TO TELUGU**

Volume 10, Issue 08, Pages:

Paper Authors
\Mr. K.Venkatesh[1], Pakalapati Sushma [2], Karri Meena Sravani [3] , Ejna Praveena Saride [4]

Editor IJIEMR

www.ijiemr.com

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

A PRE-TRAINED MODEL BERT FOR MACHINE TRANSLATION FROM ENGLISH TO TELUGU

Mr. K.Venkatesh[1], Pakalapati Sushma [2], Karri Meena Sravani [3] , Ejna Praveena Saride [4]

[1]Assistant Professor, Dept. of CSE, [2]17ME1A0544, [3]17ME1A0528, [4] 17ME1A0514
Ramachandra College of Engineering, A.P., India

ABSTRACT:

At present, Neural Machine Translation (NMT) is an innovative and latest approach for machine translation, which is way better than statistical machine translation. It has drawn so much attention to it, that most of researchers are exploring new states of art approaches to get better translations. As the area of deep learning and transfer learning are also being implemented a lot, we are trying to fit a pre-trained model's unique way of tokenization into a NMT architecture so that the pre-trained weights gives better translation. In this work we study how BERT pre-trained models might be exploited for supervised NMT. We compare various ways to integrate pre-trained BERT model with NMT model and study the impact of the monolingual data that is used to train BERT which we are proposing to use in the translation of parallel corpus.

1. INTRODUCTION

Translation using machines is being done using various methods like Rule-Based MT (RBMT), Statistical MT (SMT), Neural MT (NMT) for a while now. Statistical MT uses the predictive algorithms in teaching a computer to translate the text. NMT is predicated upon the model of neural networks within the human brain, where information is shipped to the various “layers” for processing before giving output. Statistical MT doesn't work well for language pairs with significantly different ordering. Using Neural Networks, translation of longer sentences into required language is feasible. The Previous add neural MT is completed using different models like Attention mechanism in Encoder – Decoder. duringthisproposal we might wish to introduce a fine-tuned model Bidirectional Encoder Representations from Transformers (BERT). BERT uses a completely unique technique named Masked Language Model (MLM) which allows bidirectional training in models. Transformer may be a widely known and popular attention model, to language modelling. This contrasts with the previous efforts which checked out a text sequence either from left to right or combined left-to-right and right-to-left training. We implement BERT for translating one language to a different language.Machine Translation may be a field of common language preparing which uses machines in converting normal language. Information driven machine interpretation has become the overwhelming field of concentrate due to the supply of considerable parallel corpora. the primary goal of data driven machine interpretation is to convert considered Language, as long as the frameworks absorb interpretation learning from sentence adjusted.

Image recognition is one of the most common uses of machine learning. There are many situations where you can classify the object as a digital image. For example, in the case of a black and white image, the intensity of each pixel is served as one of the measurements. In colored images, each pixel provides 3 measurements of intensities in three different colors – red, green and blue (RGB).

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

Speech recognition is the translation of spoken words into the text. It is also known as computer speech recognition or automatic speech recognition. Here, a software application can recognize the words spoken in an audio clip or file, and then subsequently convert the audio into a text file. The measurement in this application can be a set of numbers that represent the speech signal. We can also segment the speech signal by intensities in different time-frequency bands.
Machine learning can be used in the techniques and tools that can help in the 3 diagnosis of diseases. It is used for the analysis of the clinical parameters and their combination for the prognosis example prediction of disease progression for the extraction of medical knowledge for the outcome research, for therapy planning and patient monitoring. These are the successful implementations of the machine learning methods. It can help in the integration of computer-based systems in the healthcare sector.

2.REATED WORK

The problem statement considered here is Machine Translation using neural networks and pre-trained models like BERT. Sequence modeling in MT has been largely focused on supervised learning which generates a target sentence word by word from left to right. Specifically, the encoder consists of layers. Encoder is that the layer function which is used to encode the sentences of the source language into vectors which are then send to decoders to predict the next word for the target sentence for the target language. The targetsentence is predicted word by word by the decoder. It is proved in many cases that using transfer knowledge for machine learning leads to a better knowledge, so we try to transfer BERT model weights to our NMT model and perform translation.

The purpose of this project is to try using the Pre-trained model BERT in Neural Machine Translation which is supported by many papers that it has better understanding of the language.The fusing BERT with NMT means that we have to replace of add BERT encoder in the existing NMT architecture.The scope of this project is to replace the NMT encoder with BERT or using BERT encoder embedding's or Tokens in NMT architecture so that it provides better target language predictions. The compatibility of a normal encoder embedding's and BERT encoder embedding's are different.The objective of the system is to use BERT word piece tokenized input in the attention based NMT model and try to get the better results by experimenting with the values of embedding size in the encoder and decoder.

Existing system:

The existing state of the art NMT models were mainly focused on the context of the translating sentences, for that people are using attention based models and they also using transformers to achieve the goal of context based translation.

Proposed system:

In the proposed system, we try to fit the BERT tokenized input into the keras25 embedding layer because the BERT tokenizer is based on the concept of word piece model which divides a word into subwords .And we try to keep experimenting to get the better results compared to the actual system.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

3. METHODOLOGY

System design is the procedure or art of describing the architecture, components, modules, interfaces, and data for a system in order to satisfy the particular requirements. There is some overlap and synergy with the disciplines of system analysis, system architecture and system engineering.A system architecture is a conceptual model which describes the structure, behavior and more views of a system. The caption generation system architecture is organized in the following way that outputs efficient captions about the images.

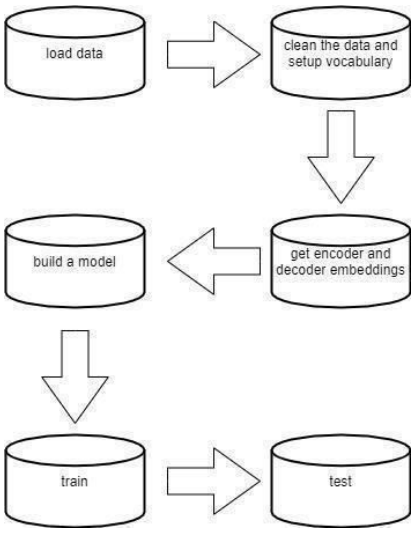


Figure 1:System Architecture

Dataset Preparation

Implementation of this systems is The Dataset considered for training of the machine learning model contains few thousands of lines which are prepared from students and few thousands of lines from internet sources. The dataset is prepared in such a way that each individual prepares two files one with English lines and the other file with its corresponding telugu translation. A new line separates each line in both the files. Now each student's data sets are appended to make a final dataset, which contains all the collected telugu lines in one file and all the corresponding English lines in another file. While implementing the experiment we read both the data sets line by line and make a tuple with each sentence as an element for English and Telugu by using the function **create_dataset()**method.

Tokenization using BERT:

BERT was made based on **Word Piece** model, which creates a fixed-size vocabulary of individual characters, words and sub-words that fits our language data The Tokenizer first checks if the consideredwordisinthevocabulary. If not, it partitions the considered word into largest possible sub-words contained in the vocabulary. And even then, if it could not find the word it breaks into individual characters. The output of the BERT tokenizer is shown in Result Analysispart.

Encoder

The encoder receives the source language inputs and encoder hidden state, the encoder hidden state is first initialized and sent as an argument to the encoder, the encoder returns a new hidden state which is used in the as decoders hidden state in the decoder, The other input to the encoder which are the source sentences in the form of list of indexes. This list of indexes is converted into fixed length of vectors which are called embedding's and returned as encoder 49 outputs.

Decoder:

The encoder hidden state obtained from encoder is taken as decoder hidden state and the initial token [CLS] is tokenized using BERT tokenizer added with another dimension by **tf.expand_dims()** which is inserted at index axis. Now for each iteration in the length of target sentence the decoder is called with three inputs which are decoder hiddenstate, decoder

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

input and encoder output. The decoder then returns the prediction, decoder next hidden state and attention weights. The predictions are used to calculate the loss function. using the attention weights we calculate the target word of the targetlanguage.

Attention:

The attention is used in the decoder with hidden state and encoder outputs as input. The attention calculates two things, Context vector and attention weights. The attention module first generate.

score based on batch size and hidden size for **query_with_time_axis**and based on batchsize, hidden size and maxlength for **values**.It generates score by tanhfunction. And further it generates attention weights using activation function **softmax**. And the context vector is generated by the Cartesian product of attention weights and values. The resulted context vector is reduced by the sum along the dimensions given in the axis.

Training The Model

To train the model we set the epoch to the convenient number, for each epoch we consider a part of the data set to train the model and also for each epoch we initialize the encoder hidden state. using data set considered and initial encoder hidden state we called **train_step(inp,targ,enc_hidden)** function. The following chart will explain the flow of steps while training the model

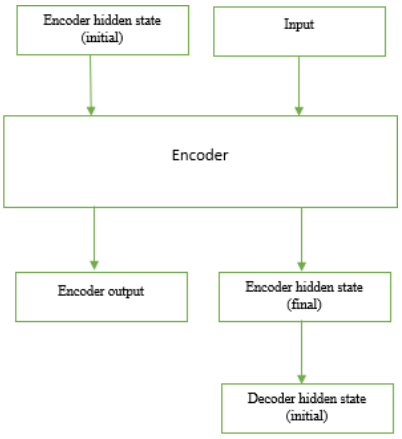


Figure 2: Encoder Function

After the above step we iterate over the target sentence word by word until the last in the following sequence as shown in the second diagram. After the controls returns back from decoder the predictions are used to calculate the batch loss and the loss per batch is returned to calculate the loss for each epoch. After all the epochs, that is the model is trained properly we use translate() function to test the model with a sentence in English to translate it into telugu. The translate function also follows similar mechanism of training it pre-process the input sentence and sends it to encoder and gets embeddings. Which are later sent into decoder with first token as decoder input and the decoder iterates each word in the input sentence and predict the next word in the target language.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

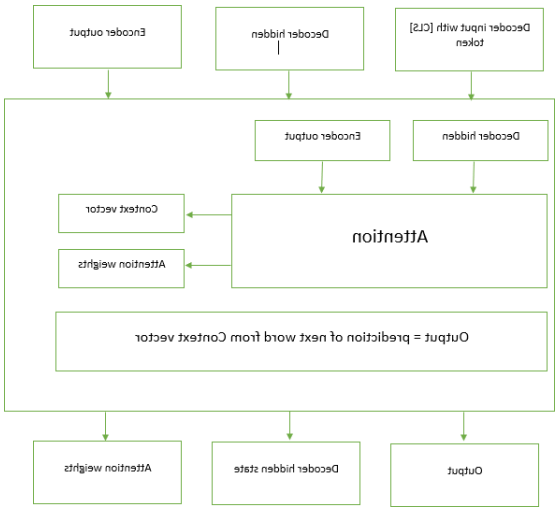


Figure 3: Attention Function

4. STUDY OF RESULTS:

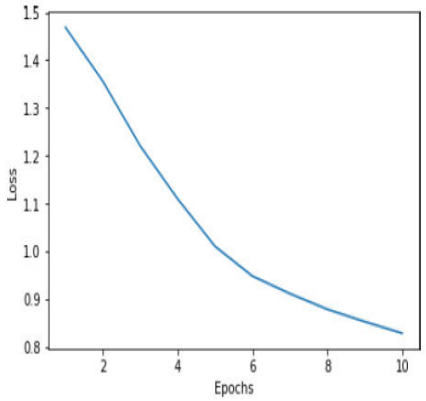


Figure 4:Graph of loss Function

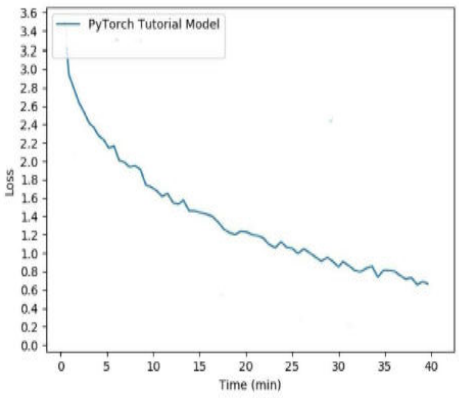


Figure 5: Graph of loss Function of Pytorch model

5.CONCLUSION :

We fit the BERT tokenized input into the keras embedding layer because the BERT tokenizer is based on the concept of word piece model which divides a word into subwords.The training is done by using attention based Neural Machine Translation model of Tensorflow, and we have used distilled BERT decoder for translating a sentence. We keep working on fitting the BERT embedding's of 768 parameter size.

The project has a very vast scope in future. The project can be implemented on intranet in future. Project can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion. With the proposed software of database Space Manager ready and fully functional the client is now able to manage and hence run the entire work in a much better, accurate and error free manner. The following are the future scope for the project.This can also be collaborate with the Feemanagement system. We can add Bar code Reader. We can take the print out of no due form without using search engine by adding the one module. We can add the some modules to recovery of the password, when the user forget the password.

6.REFERENCES :

[1] Ajay anandVerma and Pushpak Bhattacharyya, "Literature Survey: Neural Machine Translation", Indian Institute of Technology Bombay,June.2018.

[2] BrianTubayandMartaR.Costa-jussa,"NeuralMachineTranslationwiththeTransformerand Multi-Source Romance Languages for the Biomedical WMT 2018 task", Proceedings of the Third Conference on Machine Translation (WMT), Volume 2: Shared Task Papers, pages.667– 670 Belgium, Brussels, October 31 - November 1, 2018. c 2018 Association for Computational Linguistics,2018.

[3] DzmitryBahdanau, Kyunghyun Cho and YoshuaBengio, "Neural Machine Translation by Jointly Learning to Align and Translate", Published as a conference paperatICLR,2014.

[4] Fabian Hirschmann, Jinseok Nam and Johannes Furnkranz, "What Makes Word-level Neural Machine Translation Hard: A Case Study on English-German Translation", Proceedings of COLING 2016, the 26th International Conference on Computational,2016.

[5] HimanshuChoudhary, Aditya Kumar Pathak, Rajiv Ratn Shah and PonnurangamKumaraguru, "Neural Machine Translation for English-Tamil", Proceedings of the Third Conference on Machine Translation (WMT), Volume 2: Shared Task Papers, pages 770–775 Belgium, Brussels, October 31 - Novermber 1, 2018. c 2018 Association for Computational Linguistics, Jan2018.

[6] J. Hou, S. Zhang, L. Dai and H. Jiang, "Feedforward sequential memory networks based encoder-decoder model for machine translation", 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 71Kuala Lumpur, 2017, pp. 622-625, March2017.