COVHINDIA: DEEP LEARNING FRAMEWORK FOR SENTIMENT POLARITY DETECTION OF COVID-19 TWEETS IN HINDI

Purva Singh

School of Computer Science and Engineering, VIT University,
Vellore,India

ABSTRACT

On 11th March 2020, the World Health Organization (WHO) declared Corona Virus Disease of 2019 (COVID-19) as a pandemic. Over time, the exponential growth of this disease has highlighted a mixture of sentiments expressed by the general population from various parts of the world speaking varied languages. It is, therefore, essential to analyze the public sentiment during this wave of the pandemic. While much work prevails to determine the sentiment polarity for tweets related to COVID-19, expressed in the English language, we still need to work on public sentiments expressed in languages other than English. This paper proposes a framework, Covhindia, a deep-learning framework that performs sentiment polarity detection of tweets related to COVID-19 posted in the Hindi language on the Twitter platform. The proposed framework leverages machine translation on Hindi tweets and passes the translated data as input to a deep learning model which is trained on an English corpus of COVID-19 tweets posted from India [18]. The paper compares nine deep learning models' performances in classifying the sentiment polarity on an English dataset. Performance comparison of these architectures reveals that the BERT model had the best polarity detection accuracy on the English corpus. As part of testing the Covhindia's accuracy in performing sentiment classification on Hindi tweets, the paper employs a separate dataset developed using a python library called Tweepy to extract Hindi tweets related to COVID-19. Experimental results reveal that Covhindia achieved state-of-the-art accuracy in classifying COVID-19 tweets posted in the Hindi language. The use of open-source machine translation tools paved the way for leveraging Covhindia for performing multilingual sentiment classification on COVID-19 tweets. For the benefit of the research community, the code and Jupyter Notebooks related to this paper are available on Github[1].

KEYWORDS

Natural language Processing, multilingual sentiment polarity detection, LSTM, BERT, COVID-19, pre-trained word embeddings.

1. INTRODUCTION

In December 2019, several patients emerged with pneumonia from an unknown cause in Wuhan, China [1]. After tracing the contact line for these patients, the authorities linked these patients back to the seafood and wet animal wholesale market in Wuhan [1]. Chinese authorities' in-depth investigation further confirmed that the novel coronavirus was the carrier of the disease [1], which is now known as the Corona Virus Disease (COVID-19). The coronavirus is a highly infectious disease, leading to a high incubation period [2]. On average, an individual can spread the infection to 2 to 4 other individuals [3]. The world recorded a total of 38,619,674 cases of confirmed COVID-19 and 1,093,522 deaths as of October 16th, 2020 [4]. Coronavirus has not just become a source of a highly infectious contagious disease, but it has indirectly become a cause for depression and anxiety in the general population due to the misleading information posted on social media platforms.

23

The fake news and misleading information posted on these platforms are having a direct impact on mental health. With strict norms being imposed on social distancing and given the current lockdown situation, many of the population are dependent upon the internet. Social media has experienced the highest spike in its usage this year [5]. Through social media platforms such as Twitter, Instagram, and Facebook, we express our ideas, emotions and post about the things that we do in our daily lives. Given the strict imposition of lockdown, the only way people get to know about the outside world is through social media. Therefore, these social media platforms must be channels through which authentic information about coronavirus reaches the general population.

But in contrast, upon analyzing the posts on these platforms, it is observed that false figures and wrongful data have misled social media users. On social media platforms, the general population is getting attacked by unlawful information, which is very dangerous. Thus, the admins of social media platforms should perform heavy checks on the information posted online. There has been growing research to determine the correlation between the posts on Twitter and public health. A systematic review paper identified six main uses of Twitter for public health: analysis of shared content, surveillance of public health topics or diseases, public engagement, recruitment of research participants, Twitter-based public health interventions, and network analysis of Twitter users [6]. Other studies analyzed Twitter data for sentiment classification [7] and the use of Twitter to propagate credible vaccine-related web pages.

This paper aims to provide a deep learning model for sentiment polarity detection of COVID-19 tweets posted on the Twitter platform in Hindi. Though a lot of work already exists related to sentiment classification for COVID-19 tweets using various machine learning models, we still need to work on constructing model architectures that determine the sentiment polarity of posts related to COVID-19 in a language other than English. As a first step towards this problem, this paper proposes a framework to determine tweets' sentiment polarity associated with the novel coronavirus, posted in the Hindi language on the Twitter platform. On a high-level, the immediate solution to this problem statement is to use machine translation tools that are open-sourced [8]. Researchers in sentiment classification has been reluctant to leverage existing machine translation tools due to their low performance in translation tasks [8]. However, the performance of machine translation models has significantly improved in the past years.

The proposed framework accepts a Hindi tweet related to COVID-19 as an input. The tweet in Hindi is translated into the corresponding English tweet using the open-source machine translation tool provided by Google called Google Translate. It passes the translated tweet to a preprocessed pipeline. The cleaned tweet serves as an input to a deep learning sentiment classification model trained on a corpus containing English tweets related to COVID-19. The paper compares the performance of nine different deep-learning models to determine the best model configuration for classifying the sentiment polarity of English tweets.

This set of deep-learning models are a combination of long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) based models combined with pre-trained word embeddings such as GloVe [9], FastText [10], and Crisis [12] embeddings. The paper also employs the latest advancements in the field of Natural Language Processing (NLP), such as the BERT model, to compare its performance accuracy in classifying sentiment of English tweets related to COIVID-19. Upon extensive analysis, the paper observed that the BERT model had the best classification accuracy of 99.7% on training data and 93.8% validation accuracy. Leveraging Google Translate, and the pre-trained BERT model, the proposed framework for Covhindia performs sentiment classification on Hindi tweets. The paper employs a separate testing dataset containing Hindi tweets related to COVID-19, curated from the Twitter platform using a python library called

Tweepy. The article employs this testing dataset to calculate the classification accuracy of the Covhindia framework on Hindi tweets. Experiments reveal that the Covhindia framework achieves state-of-the-art accuracy of 88.9% on the Hindi dataset.

Research Objective and Questions

This paper's primary goal is to leverage open-source machine translation tools to perform sentiment polarity detection of Hindi tweets posted on the Twitter platform related to the novel coronavirus. In this investigation, the paper aims to answer the following research questions (RQ):

1. RQ1: To what extent can we use NLP with deep learning for performing sentiment classification?
2. RQ2: How do variations of deep learning models (such as LSTM, Bi-LSTM, with or without pre-trained word embeddings, BERT) perform on sentiment classification tasks?
3. RQ3: Can we leverage open-source machine translation tools for performing multi-lingual sentiment polarity detection?

Contributions

The primary contributions of this paper are as follows:

1. The deep learning model for performing sentiment classification on tweets related to coronavirus.
2. Leveraging open-source machine translation tools to perform sentiment classification of tweets posted in the Hindi language.
3. Achieve state-of-the-art accuracy in classifying the sentiment polarity for Hindi tweets.

The structure of the paper is as follows. Section II defines the literature survey, which discusses the extant published works that perform sentiment classification on COVID-19 tweets and existing models that leveraged machine translation for conducting multi-lingual sentence-level sentiment classification. Section III covers the detailed explanation of the Kaggle dataset being used to train the deep learning model, followed by Section IV, which conducts the performance comparison of nine different deep learning models and also delineates the system architecture of the proposed framework, Covhindia. Section V elaborates on the discussion and analyzes the results achieved after various experiments, followed by Section VI, which provides concluding remarks on the paper and potential future research directions.

2. LITERATURE SURVEY

A lot of work exists in performing sentiment polarity classification for tweets posted on the Twitter platform. Hence, this section discusses the extant methods and their unique proposals, along with their limitations. Shiayng Liao et al. [9] have adopted a unique way for classifying sentiments of Twitter data. They have leveraged Convolutional Neural Networks (CNNs) to determine the sentiment polarity of tweets. They have achieved better accuracy performance compared to traditional classification techniques such as SVM and Bayes method. In [10], the authors have developed two Naive Bayes unigram models, one Naive Bayes bigram model and a Maximum Entropy model for sentiment polarity detection. They reported that their Naive Bayes classification model performed better than the Maximum Entropy model. Ji X et al. [11] have proposed the Measure of Concern (MOC) to measure public health concerns. In their paper, they have developed a sentiment polarity classification in a two-step approach. As the first step, they

classify the tweets related to health into two separate categories: Personal tweets v/s News tweets. The second step has leveraged an emotion-oriented clause-based method to extract the training dataset and generate another classifier to predict whether a Personal tweet is Negative or Non-Negative [11]. The authors in [12] have employed Machine Learning (ML) algorithms to classify tweets' sentiment into positive or negative. They have analyzed the progress of emotions, especially the fear-sentiment amongst the U.S. population, and have compared two ML algorithms for classifying Coronavirus tweets of different lengths. In their experiment, the authors observed a classification accuracy of 91% using the Naive Bayes algorithm and 74% accuracy using logistic regression classification on short tweets. However, they achieved relatively low performance while conducting sentiment analysis on longer tweets. G. Barkur et al. [13] have performed sentiment analysis to observe the various emotions expressed by Indians after the imposition of strict lockdown by the Indian government. To extract the tweets that expressed sentiments related to the lockdown, they employed two primary hashtags: #IndiaLockdown and #IndiafightsCorona.

Earlier, the researchers were skeptical of using existing open-source machine translation tools to translate a tweet to the English language due to their poor performance. But recent developments in neural machine translations have optimized the performance of Simple Machine Translation (SMT) tools such as Google Translate. A. Joshi et al. [14] have proposed three different ways to perform sentiment analysis on tweets posted in the Hindi language. In the first approach, they have used Hindi documents as both training and testing data, using which they have trained their classifier. Their second approach has leveraged Machine Translation (MT) to perform sentiment analysis on Hindi tweets, assuming that the sentence's sentiment remains intact even after the translation. Finally, in the third approach, they have leveraged a Hindi annotated sentiment corpus called Hindi-SentiWordNet (H-SWN) to derive each word's sentiment score and then used the maximum score to assign the polarity to each word. The overall sentiment of the document is the majority of the polarity of its terms [14]. Their results show that in-language sentiment analysis performed the best, followed by the MT-based sentiment analysis. G. Shalunts et al. [15] have explored the impact of machine translation on sentiment analysis tasks. For this, they have employed a sentiment analysis tool called SentiSAIL and SDL Language Weaver as their MT tool. Their original corpora were in three different languages: Spanish, Russian, and German, translated into English. Their results showed that the translational quality of SDL Language Weaver allowed them to achieve comparable performance rates on both original and translated parallel corpora. At the same time, the SentiSAIL performed sentiment polarity classification. The authors in [16] have conducted a sentiment analysis on French Movie reviews. They have proposed a supervised classification of movie reviews where sentiment polarity detection depends on POS tagging, negation forms, and chunking. To improve the classification score, they have extracted word semantic orientation from a lexical resource called SentiWordNet [16]. To perform polarity classification on French Movie reviews, they have employed machine translation tools to translate the movie reviews in French to English and have shown significant performance improvement in terms of bag of words (BOW) baseline.

3. DATASET

Two different datasets are employed in training and testing the proposed framework. The framework uses the COVID-19 dataset available on Kaggle called, "Covid19 Indian Sentiments on Covid19 and Lockdown", published by Suraj Kumar containing a curated list of 3090 English tweets posted from India [18] as the first dataset (Dataset-I). This dataset includes clean tweets from India on topics like coronavirus, covid 19, lockdown, etc. The tweets have been collected between dates 23rd March 2020 and 15th July 2020 and are labeled into four sentiment categories

fear, sad, anger, and joy. The proposed framework uses this dataset for training the sentiment classification models on English tweets.

The second dataset (Dataset-II), employed by the Covhindia framework, is used as the testing dataset to determine the framework's accuracy to classify the sentiment polarity of a Hindi tweet. The tweets posted in the Hindi language are collected using a python library called Tweepy, which helps us to extract Twitter data with proper authorization based on various query parameters. The Hindi tweets are extracted based on the following hashtags: #coronavirus, #COVID19, #StayHomeChallenge, #IndiaFightsCorona and #AarogyaSetu. The framework extracted 363 Hindi tweets related to COVID19 based on the mentioned hashtags and were manually annotated with sentiment polarity tag (negative, positive). Out of these Hindi tweets, 178 tweets conveyed negative sentiment, and 183 tweets displayed a positive tone.

4. METHODOLOGY

Data Preprocessing

As the first step towards sentiment classification of Hindi tweets, the proposed framework trains a deep learning model for detecting tweets' sentiment polarity in English. This section analyzes and compares the training and validation accuracy of different deep-learning models to select the best performing model in terms of sentiment classification of English tweets. The paper combines long short-term memory (LSTM) based and bidirectional LSTM based models with existing pre-trained word embeddings such as GloVe, FastText, and Crisis embeddings. The framework leverages the PyTorch library as a back-end deep learning engine. As a part of training the models on English tweets, this section leverages Dataset-I [18] described in Section III. Before feeding English tweets to the model, the framework performs the below-mentioned steps as a part of preprocessing the tweets:

1. Convert tweets to lower case and remove stopwords, punctuations, user mentions, hyperlinks, and special characters.
2. Tokenization. Create a mapping dictionary that converts vocabulary to corresponding integral value.
3. Encoding tweets. In this step, the framework creates an array that contains an integer encoded version of words in tweets. The term appearing the most has the least integer value.
4. Outlier tweets removal. The framework gets rid of too lengthy or concise tweets.
5. Padding. To feed tweets of consistent length to the model, the framework pads/truncates the remaining data.
6. Split input data into three categories: train, test, and validation.
7. Create data loaders and perform batching of input data.

Analysis of Deep Learning Models for Performing Sentiment Classification on English Tweets

This section analyzes the classification accuracy of nine different deep-learning models on an English dataset. The framework employs deep learning models that leverage LSTM cells. LSTMs are a variant of Recurrent Neural Networks (RNNs) that can capture long-term dependence. They are useful in performing various NLP tasks such as sentiment classification, speech recognition, text classification, and many more. Hence, the paper experiments on deep learning models that leverage LSTM or bi-directional LSTM cells to classify the sentiment polarity of COVID-19 tweets posted in English. This paper employs various pre-trained word embeddings to increase

the classification accuracy of the vanilla LSTM/Bi-LTSM model. A word-embedding is a way of presenting words in such a way that allows words with similar meanings to have similar representations. There are many word-embeddings in NLP that have been pre-trained to perform transfer learning for NLP tasks. Some of the most popular pre-trained word embeddings are Google's Word2Vec, Stanford's GloVe, Facebook's FastText, to name a few. The proposed framework first experiments with eight different versions of deep learning models: vanilla LSTM + 3 pre-trained word embeddings and vanilla Bi-LSTM + 3 pre-trained word embeddings. Paper [22] describes more information on integrating deep learning models with pre-trained word embeddings. In this section, we leverage the Dataset- I, which contains 3090 total tweets posted from India. The framework uses 80% of these tweets as the training data for the models. The testing and validation dataset was created by dividing the remaining 20% of the tweets in a 50/50 ratio.

The final model which the paper uses to perform sentiment classification on COVID19 tweets posted in English is BERT (Bi-directional Encoder Representations for Transformers), using the HuggingFace Transformer library, PyTorch, and Python. In this paper, the framework uses the BERTbase uncased model, which comprises 12 transformer blocks, 768 hidden layers, and 12 attention heads. The framework employs the AdamW optimizer provided by the HuggingFace to mirror the training procedure mentioned in the original BERT paper [23]. This optimizer implements the Adam algorithm and corrects the weight decay [24]. The model employs the Softmax activation function to the model's output to get the predicted probabilities from our trained model and leverages Cross-Entropy Loss as the loss function. Finally, the model avoids the exploding gradients problem by employing gradient clipping. During the training procedure, we save the best model that achieved the highest validation accuracy. The framework accelerates the model's training process by using GPU (Graphics Processing Unit) and training it up to ten epochs to avoid overfitting. At the end of this procedure, we have obtained a BERT model trained on English tweets concerning COVID-19.

After conducting experiments on the nine mentioned models, this section concludes with the BERT model achieving the highest training accuracy of 99.7% on training dataset and 93.8% accuracy on validation dataset. The training accuracy and validation accuracy of the best model (BERT) against the number of epochs is represented by Figure 1 and Table 1 summarizes the training and validation accuracy obtained for all the nine models.
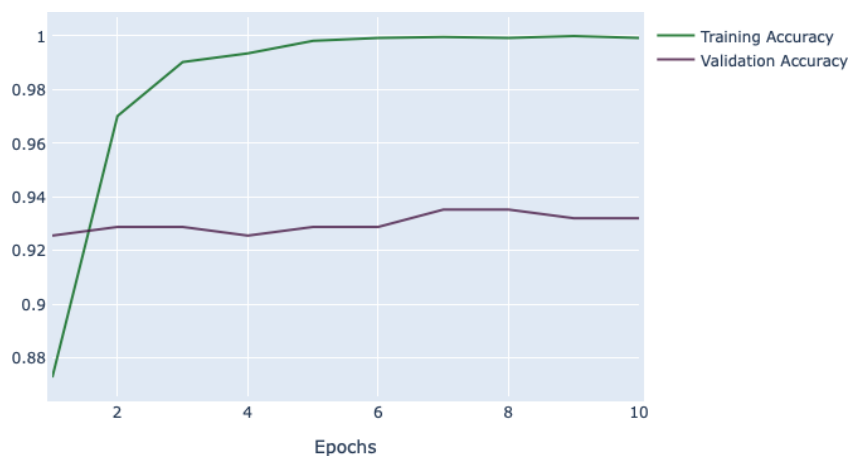

Figure 1: Training and validation accuracy of BERT model when trained on English tweets.

The probable reason for the results obtained in Table 1 is that traditional embedding layers such as Word2Vec, GloVe, and FastText provide a single, context-based representation for every token. But BERT takes a tweet as an input, and it calculates the token-level representations by leveraging the information from the entire tweet. The BERT model leverages Transformer, which is an attention mechanism to understand the relationship between words in a text. As opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), the Transformer encoder reads the entire sequence of words at once [23]. Hence, it is considered bidirectional. This characteristic allows the model to learn the context of a word based on its surroundings (left and right of the word). More details related to leveraging BERT for multiple NLP tasks such as question-answering model, text-classification process, etc. can be found in [23].

Table 1: Training and Testing Accuracy of nine deep learning models

S.No.	Model Name	Model's Configurations	Training Accuracy	Validation Accuracy
1.	Basic LSTM model	Embedding Layer with LSTM (256) and Dropout with p=0.3, followed by a fully Connected Layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	85.1%	86.7%
2.	Basic LSTM + GloVe Embeddings	Embedding Layer with glove.6B.100d.txt, LSTM(256) with Dropout = 0.2, fully-connected layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	88.9%	90.9%
3.	Basic LSTM + FastText Embeddings	Embedding Layer with wiki-news-300d- 1M.vec, LSTM(256) with Dropout = 0.2, fully-connected layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	92.5%	88.9%
4.	Basic LSTM + Crisis Embeddings	Embedding Layer with Crisis-embeddings.300d.txt, LSTM(256) with Dropout = 0.2, fully-connected layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	83.1%	84.7%
5.	Basic bi-directional LSTM	Embedding Layer with bidirectional LSTM (256) and Dropout with p=0.3, followed by a fully Connected Layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	87.3%	86.0%
6.	Basic bi-directional LSTM + GloVe Embeddings	Embedding Layer with glove.6B.100d.txt, bi-directional LSTM(256) with Dropout = 0.2, fully-connected layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	91.2%	90.6%
7.	Basic bi-directional LSTM + FastText Embeddings	Embedding Layer with wiki-news-300d- 1M.vec, bi-directional LSTM(256) with Dropout = 0.2, fully-connected layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	88.3%	88.6%
8.	Basic bi-directional LSTM + Crisis Embeddings	Embedding Layer with Crisis-embeddings.300d.txt, bi-directional LSTM(256) with Dropout = 0.2, fully-connected layer (1) with Sigmoid activation function. Training with Adam optimizer and Binary Cross Entropy (BCE) loss as the loss function.	86.0%	85.1%
9.	BERT	$BERT_{base}$, uncased model, 12 layers (transformer blocks), 12 attention heads, 100 million parameters. Fully connected linear layer with Softmax activation function.	**99.7%**	**93.8%**

In the next section, the paper employs the pre-trained BERT model (the best performing model) to identify the sentiment polarity of Hindi tweets concerning COVID19.

Sentiment Polarity Detection of Hindi Tweets Concerning COVID19 using BERT and Machine Translation

In this section, we leverage the pre-trained BERT model from section 4.2. to perform sentiment classification on COVID19 tweets posted in Hindi. For this section, we would be leveraging Tweepy, an open-sourced Python library for accessing the Twitter API, to extract tweets in Hindi to build our testing dataset. Figure 2 shows the system architecture of the pipeline of the proposed framework. In the consequent sections, this paper delineates every step of algorithm employed by Covhindia to perform sentiment classification on COVID-19 tweets posted in Hindi language.
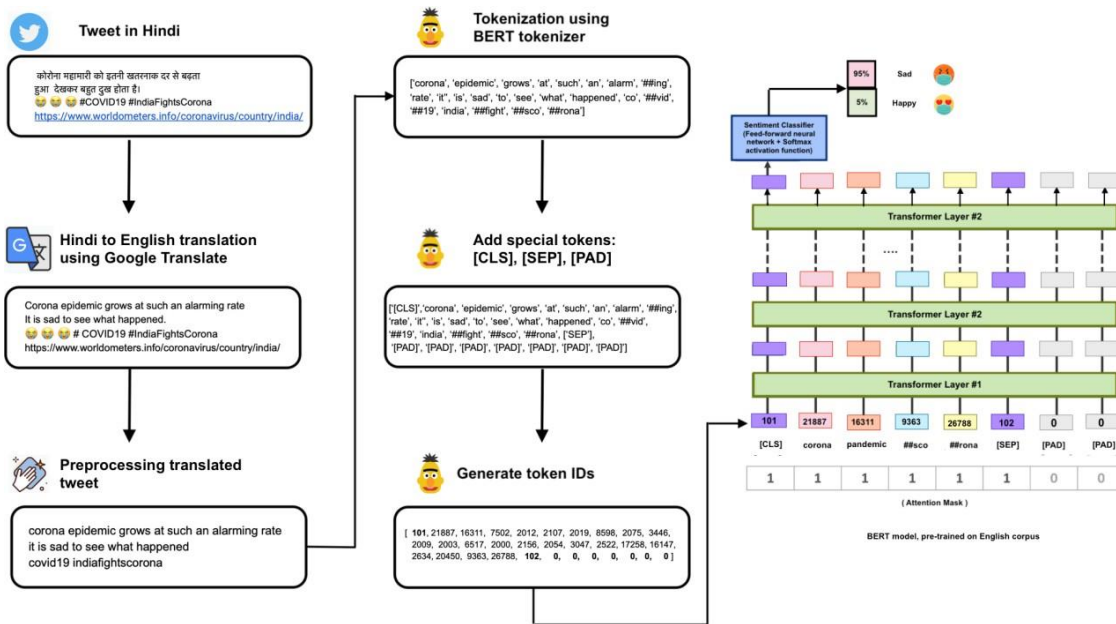

Figure 2: System architecture of the proposed model, Covhindia, for detecting sentiment polarity of Hindi tweets concerning COVID-19. (Icons made by https://www.flaticon.com)

Preprocessing Translated Tweets

First, we translate the input tweet in Hindi to English. Here, we are making an assumption that the sentiment of the translated tweet would remain intact even after the translation. According to the authors in [27], they have found a considerable overlap in the set of features generated from human-translated and machine translated texts and have concluded that that Google Translate is a useful tool for comparative researchers when using bag-of-words text models. The authors in [28] have leveraged three different open-source Machine Translation (MT) systems: Google, Bing and Moses and their extensive calculations have shown that MT systems can be used for multi-lingual sentiment classification.

We can't directly feed the translated raw English tweet to the BERT model. We need to convert the text into a format that is understandable by BERT. Hence, Covhindia performs the following preprocessing steps on the translated data –

1. Convert text into lower-case, remove text in square brackets, remove hyper-links, remove punctuations, remove words containing numbers, and remove emojis.

2. Create a BERT tokenizer from the pre-trained BERT model (BERTbase uncased) using the Transformers library provided by HuggingFace.
3. Using the tokenizer created in Step-2, tokenize the cleaned tweet.
4. Convert the tokens extracted from Step-3 into token IDs.
5. Add the following special tokens for additional information –

   a. ([CLS], 101) : The "CLS" token is added to the start of each text to inform BERT that we would be performing a text-classification task. Here "[CLS]" is the token, and 101 is its corresponding token ID.
   b. ([SEP], 102) : The "SEP" token is appended at the end of each sentence. This token marks the end of the sentence.
   c. ([PAD], 0) : To maintain a uniform length of tweets across the entire training dataset, the BERT tokenizer pads the tweets by appending the token "[PAD]" at the end of each sentence to maintain the consistency of the length of each input sentence.

Building Sentiment Classification Model

Now that our translated tweets are preprocessed, we will build our BERT-based Sentiment Classifier model. Our model consists of three parts:

1. BERT Model: Using the Transformers library, we create an instance of BERT model and pass the text token IDs and attention mask as the input. Attention mask is a binary dictionary (consisting of 0s and 1s) that tells BERT which tokens should be attended to and which should not [26]. This argument is helpful when we are trying to put together encoded version of text of different lengths. Output of the model is a vector of size corresponding to the size of the hidden layer in BERT base, i.e., 768. For classification purposes, we are only concerned with the output of the first position (where we passed the "[CLS]" token).
2. Dropout Layer: The pooled output is then forwarded through a dropout layer for some regularization.
3. Fully Connected Linear Layer: Output from the dropout layer is passed as input to the fully-connected linear layer. The result from fully-connected linear layer is then passed into the Softmax activation function to determine the final polarity of the COVID-19 Hindi tweet.

Model Evaluation and Results

To evaluate the performance of our BERT model to correctly classify tweets related to COVID-19 posted in the Hindi language, we leverage Dataset-II. This dataset contains 363 tweets posted in Hindi, which are related to COVID-19. Out of these tweets, 178 Hindi tweets conveyed negative sentiment, and the rest 185 tweets indicated a positive opinion. We manually annotated the testing dataset as positive or negative. We leveraged it to test our model's classification accuracy on tweets concerning COVID-19 posted in the Hindi language on Twitter. The results show that our proposed framework achieved 83.2% accuracy in classifying Hindi tweets that conveyed negative sentiment and 79.2% classification accuracy on Hindi tweets, indicating positive view. The overall accuracy of Covhindia on the testing dataset containing Hindi tweets was 88.9%.

5. DISCUSSION

The paper proposes a framework called Covhindia that leverages machine translation and deep-learning model to perform sentiment classification of COVID-19 tweets posted in the Hindi

language on the Twitter platform. The framework accepts a Hindi tweet as its input and leverages an open-source machine translation tool, Google Translate, to convert Hindi tweet to English. In this step, we are assuming that the sentiment of the translated tweet remains intact. This translated data is fed into a preprocessing pipeline to generate clean text. The article analyzes the classification accuracy of nine deep-learning models based on BERT, LSTMs, and Bi-LSTMs combined with three different pre-trained word embeddings, GloVe, FastText, and Crisis embeddings. Finally, the paper selects the BERT-based model, which achieved the highest classification accuracy on the English corpus. This pre-trained BERT model is further employed to determine the sentiment polarity of the translated tweet. In the end, the proposed framework achieved state-of-the-art classification accuracy on Hindi tweets related to COVID-19. These results encourage the use of open-source machine translation tools for performing multi-lingual sentiment classification tasks.

Future Work

Following are the directions for future work related to the proposed framework, Covhindia:

1. The testing dataset for Covhindia contains 363 Hindi tweets by performing web-scrapping using Tweepy. In the future, we can increase the number of Hindi tweets employed as testing data to validate the framework's accuracy further.
2. The experiment in Section 4.2. involves analyzing the classification accuracy of nine different deep-learning models on an English corpus. We can analyze the classification accuracy of a convolutional neural network (CNN-based) model for future purposes.
3. The proposed framework experimented with BERTbase uncased model. We can also experiment with BERTlarge (cased/uncased) and BERTbase cased model for multi-lingual sentiment classification.
4. In the future, we can further validate the machine translation framework by converting English tweets to Hindi and then back to English in order to manually check whether the sentiment of the translated tweet has remained intact.
5. As a part of future analysis, the performance of Covhindia can be compared with models that employ different techniques for performing multilingual sentiment classification.

REFERENCES

[1] World Health Organization. 2020 Jan 21. Novel coronavirus (2019-nCoV) situation report - 1 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4

[2] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. and Cheng, Z., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet, 395(10223), pp.497-506.

[3] Velavan, T.P. and Meyer, C.G., 2020. The COVID-19 epidemic. Tropical medicine & international health, 25(3), p.278.

[4] Report by World Health Organization, https://covid19.who.int

[5] Internet Users Worldwide Statistic, Available at: https://www.broadbandsearch.net/blog/internet-statistics, Anonymous, retrieved 10th September, 2020.

[6] Sinnenberg, L., Buttenheim, A.M., Padrez, K., Mancheno, C., Ungar, L. and Merchant, R.M., 2017. Twitter as a tool for health research: a systematic review. American journal of public health, 107(1), pp.e1-e8.

[7] Steffens, M.S., Dunn, A.G., Wiley, K.E. and Leask, J., 2019. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. BMC public health, 19(1), pp.1-12.

[8]   Balahur, A. and Turchi, M., 2012, July. Multilingual sentiment analysis using machine translation?. In Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis (pp. 52-60).

[9]   Liao, S., Wang, J., Yu, R., Sato, K. and Cheng, Z., 2017. CNN for situations understanding based on sentiment analysis of twitter data. Procedia computer science, 111, pp.376-381.

[10]   Parikh, R. and Movassate, M., 2009. Sentiment analysis of user-generated twitter updates using various classification techniques. CS224N Final Report, 118.

[11]   Ji, X., Chun, S.A., Wei, Z. and Geller, J., 2015. Twitter sentiment classification for measuring public health concerns. Social Network Analysis and Mining, 5(1), p.13.

[12]   Samuel, J., Ali, G.G., Rahman, M., Esawi, E. and Samuel, Y., 2020. Covid-19 public sentiment insights and machine learning for tweets classification. Information, 11(6), p.314.

[13]   Barkur, G. and Vibha, G.B.K., 2020. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian journal of psychiatry.

[14]   Joshi, A., Balamurali, A.R. and Bhattacharyya, P., 2010. A fall-back strategy for sentiment analysis in hindi: a case study. Proceedings of the 8th ICON.

[15]   Shalunts, G., Backfried, G. and Commeignes, N., 2016. The impact of machine translation on sentiment analysis. Data Analytics, 63, pp.51-56.

[16]   Ghorbel, H. and Jacot, D., 2011. Sentiment analysis of French movie reviews. In Advances in Distributed Agent-Based Retrieval Tools (pp. 97-108). Springer, Berlin, Heidelberg.

[17]   https://www.kaggle.com/gpreda/covid19-tweets

[18]   https://www.kaggle.com/surajkum1198/twitterdata

[19]   Nguyen, D.T., Joty, S., Imran, M., Sajjad, H. and Mitra, P., 2016. Applications of online deep learning for crisis response using social media information. arXiv preprint arXiv:1610.01030.

[20]   Pennington, J., Socher, R. and Manning, C.D., 2014, October. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

[21]   Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H. and Mikolov, T., 2016. Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651.

[22]   Kastrati, Z., Imran, A.S. and Kurti, A., 2019. Integrating word embeddings and document topics with deep learning in a video classification framework. Pattern Recognition Letters, 128, pp.85-92.

[23]   Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[24]   Loshchilov, I. and Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

[25]   https://blog.api.rakuten.net/api-tutorial-google-translate-api/

[26]   Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Brew, J., 2019. HuggingFace's Transformers: State-of-the-art Natural Language Processing. ArXiv, pp.arXiv-1910.

[27]   De Vries, E., Schoonvelde, M. and Schumacher, G., 2018. No longer lost in translation: Evidence that Google Translate works for comparative bag-of-words text applications. Political Analysis, 26(4), pp.417-430.

[28]   Balahur, A. and Turchi, M., 2014. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech & Language, 28(1), pp.56-75.

AUTHOR

Purva Singh has completed her B.Tech in Computer Science and Engineering from Vellore Institute of Technology, Vellore. She has been exploring the field of Artificial Intelligence and Deep Learning for the past two years. She is passionate about Natural Language Processing, Deep Reinforcement Learning, and Big Data Analytics.