SSN_ARMM@ LT-EDI -ACL2022: Hope Speech Detection for Equality, Diversity, and Inclusion Using ALBERT model

PraveenKumar V ,Prathyush S , Aravind P, Angel Deborah S, Rajalakshmi S, Milton R S, Mirnalinee T T,
Department of Computer Science and Engineering
Sri Sivasubramaniya Nadar College of Engineering
Chennai , India
{vpraveenkumar0211,aravind2814,prathyushsunil2510}@gmail.com,
{angeldeborahs,rajalakshmis,miltonrs,mirnalineett}@ssn.edu.in

Abstract

In recent years social media has become one of the major forum for expressing human views and emotions. With the help of smart phones and high speed internet, anyone can express their views in Social media. However this can also lead to spread of hatred and violence in the society. Therefore it is necessary to build a method to find and support helpful social media content. In this paper we studied Natural Language Processing approach for detecting Hope speech in a given sentence. We developed a BERT model to help us classify whether the given sentence belongs to hope or non-hope class. Our model achieved 1st rank in Kannada language with a weighted average F1 score of 0.750, 2nd rank in Malayalam language a weighted average F1 score of 0.740 , 3rd rank in Tamil language a with weighted average F1 score of 0.390 and 6th rank in English language with a weighted average F1 score of 0.880.

1 Introduction

Social media has become an essential part of our lives. People tend to reflect their inner self through their online conversations. There is a huge increase in the number of individuals looking for support through the internet. In recent times there has been a surge in these online support sources. (Gowen et al., 2012) Online support Groups help people going through similar disabilities, health problems, etc and overcome their difficulties together. Especially after the pandemic, more people are looking for comfort to get through these tough times in the internet. Hope is defined as a feeling of trust or a expectation of something to happen.Hope Speech takes the positive feeling a step forward and helps us achieve a more inspirative environment on social media. Recently researchers (Ganda and Madison, 2014) have found out that Social media network and Online support Groups have great impact on people's self understanding. So it is necessary to support positive content in the internet.

2 Related Works

Hope speech detection has been one of the important areas of research in recent years. Most of the data obtained from Social media do not have a proper format and tend to be written with grammatical errors and native language of the country. In recent years many researchers have developed automatic methods for hope speech detection in social media. These methods rely on popular technologies like Machine Learning and Natural Language Processing. (Zhang et al., 2018) Did hate speech analysis for short text such as tweets. They proposed DNN method which helps in identifying features useful for classification. They evaluated their model with Twitter dataset and obtained good results. (Ribeiro et al., 2018) characterized hate speech in Online Social Network with the help of n DeGroot's learning model. They found how hateful users are different from normal users using centrality measure and user activity pattern. (Ghanghor et al., 2021) Carried out hope speech detection task with various models and found out that mBERT-cased model gave the best results. They employed zero short cross lingual model transfer which is used to fine tune the model evaluation. They found out that degradation of the model performance was due to freezing of base layers of transformer model . (Muralidhar et al., 2018) focused on YouTube sentiment analysis. The researchers did an analysis on these data to find their trends and it was found that real life events are influenced by user sentiments.Hope speech can also be termed as opposite of hate speech.Hate speech includes offensive and bad comments on a particular work or on a particular person. (Chakravarthi et al., 2020b) These offensive comments create bad impact on this society. Work done by (Puranik et al., 2021) include analyzing corpus of data collected from Youtube comments.

In recent days, NLP has gained many architec-

tural advancements and gained better results than state of art methods. The task focuses on classification of Hope speech in multiple languages with each language having different class imbalance. Hope speech detection can uplift the amount of positive content in social media and helps to build peaceful world.

3 Dataset Description

In this work, we made use of the datasets provided by the Association for Computational Linguistics for Hope Speech Detection for Equality, Diversity, and Inclusion competition. These are multilingual datasets constructed by (Chakravarthi et al., 2020a).It consists of comments made by users from the social media platform YouTube with 28,424, 17715 ,9918 and 6176comments in English, Tamil, Malayalam and Kannada respectively, manually labeled. In these datasets,the comments are classified into two different categories as Hope-speech and Non-hope-speech. The distribution of the each language dataset is shown in the table below.

Language	Train	Test	Dev
Tamil	14199	1761	1755
English	22740	2841	2841
Malayalam	7873	1071	974
Kannada	4940	618	618

Table 1: Summary of Dataset

4 Methodology

Text classification is the foundation of Hope speech detection where we classify sentences into different categories. Text classification problem have been improvised with better approaches and good results have been obtained. The NLP domain has provided many techniques to provide solution to text classification problems (Devlin et al., 2019). In our proposed method we used IndicBERT model which is a multilingual model trained on large-scale corpora covering 12 Indian languages (M K and A P, 2021). IndicBERT takes less number of parameters and still manages to give state of art performance. We first used label encoders to convert text labels into numerical labels since it is easy to find the probability of the models. One of the important parameter for training NLP model is Batch size. It controls the speed and stability of learning process. We used equal batch size to separate data, irrelevant of training data size with similar class weight ratio.

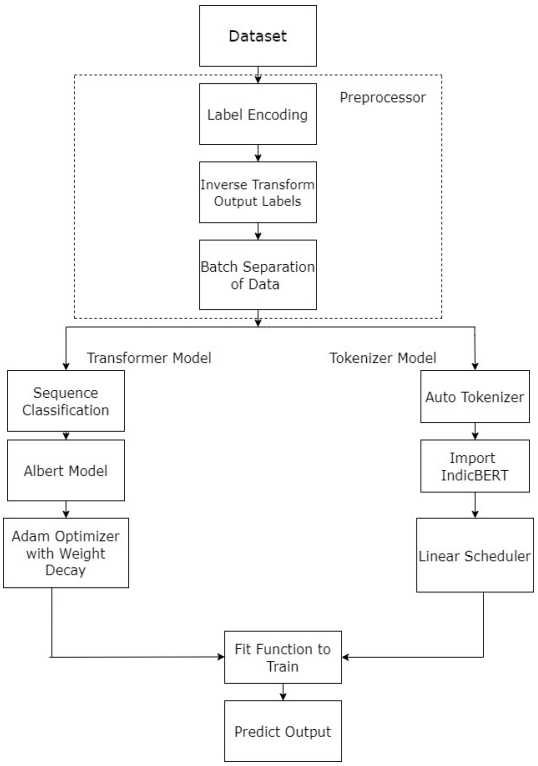

Figure 1: Overview of the process

We used pre-trained IndicBERT model from Huggingface transformers library. Tokenizer converts sentences into tokens which helps to understand and develop NLP model. We used ALBERT model from transformers which is a known as "Lite BERT for Self-supervised Learning of Language Representations" for sequence classification. ALBERT model has similar architecture as BERT model, but ALBERT model takes 18x less parameter compared to BERT model. This is achieved by splitting the embedding matrix into two levels where the input-level embedding takes low dimensions, where as the hidden-layer embedding takes higher dimensions. Transformer based neural network gives us another advantage through a technique called parameter-sharing where they use same parameters for different independent layers. By this way , the same layer is applied on top of each hidden layers.This approach reduces the size of parameters with a trade-off in accuracy. We used Pytorch library to run and test specific parts of code.

Neural Networks are difficult to train because there is large number of hyper parameters to specify and optimize. Choosing the right parameter can result in significant increase in the performance of the model. We used AdamW optimizer to up-

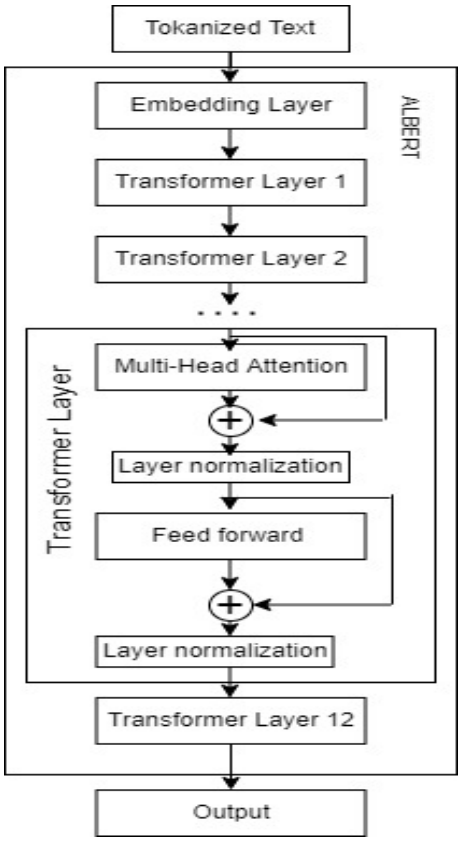

Figure 2: Overview of the process

Hyperparameters	Value
epoch	4
batch size	16
learning rate	$2e^{-5}$
max_length	400
activation	tanh
optimizer	AdamW

Table 2: Hyper-parameters of the model

date the network weights iteratively. AdamW optimizer yeilds better training loss than Adam optmizer. AdamW optimizer generalizes well compared to other models so that they can compete with stochastic gradient descent with momentum. Learning rate determines the extent to which the weights in the network are to be modified during backpropagation. In this research, we used linear scheduler with warm-up to increase the learning rate constantly. Backpropagation is a mathematical algorithm that is used to calculate gradient loss of a function with respect to other weights in the network. It tells us how exactly the network processes certain inputs. Backpropagation is used to calculate loss function and distribute it back in the network in backward direction. Our hyper parameters are present in table 2.

5 Experimental Result

In this section we described about the results we got in particular language datasets.

5.1 Evaluation for English

The result of the test dataset for English Language is given in the table 3. The dataset contains 22740 sentences for training, 2841 sentences for development and 2843 sentences for testing. We found that there is a class imbalance between Hope (N=1962) and Non-Hope (N=20778). Our model attained precision of 0.880, recall of 0.890 and F1-score of 0.880 achieving 6th rank among other 20 submissions. While our approach got good results there is still a room for improvement.

Label	Value
Non_hope_speech	2641
Hope_speech	202

Table 3: Test result for English Language

5.2 Evaluation for Tamil

The result of the test dataset for English Language is given in the table 4. The dataset contains 1755 sentences for development, 14199 sentence for training and 1761 for testing. Our model attained precision of 0.370, recall of 0.420, F1-score of 0.390 achieving 3rd rank among other submissions.

Label	Value
Non_hope_speech	1256
Hope_speech	505

Table 4: Test result for Tamil Language

5.3 Evaluation for Malayalam

The result of the test dataset for English Language is given in the table 5. The dataset contains 974 sentences for development, 7873 sentence for training and 1071 for testing. Our model attained precision of 0.700, recall of 0.780, F1-score of 0.740 achieving 2rd rank among other submissions.

5.4 Evaluation for Kannada

The result of the test dataset for English Language is given in the table 6. The dataset contains 618 sentences for development, 4940 sentence for training

Label	Value
Non_hope_speech	883
Hope_speech	188

Table 5: Test result for Malayalam Language

and 618 for testing. Our model attained precision of 0.740, recall of 0.760, F1-score of 0.750 achieving 1st rank among other submissions.

Label	Value
Non_hope_speech	444
Hope_speech	174

Table 6: Test result for Kannada Language

6 Conclusion

Due to pandemic there has been a sudden increase in active social media users which has lead to abundant online content. There is a need to promote and motivate positive content to spread peace and knowledge in this society.In this is paper we proposed a transformer based approach for Hope speech detection in 4 different languages (English, Tamil, Malayalam, Kannada). We used ALBERT model with Adam optimizer for classification. Our model got F1 score of 0.880, 0.390, 0.740, 0.750 in English, Tamil, Malayalam and Kannada. We can achieve good results by adjusting the hyperparameters for model training and also by increasing the training data set size. In future work we will be able to handle class imbalance with improvised dataset.

References

Bharathi Raja Chakravarthi, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P. McCrae. 2020a. A sentiment analysis dataset for code-mixed malayalam-english. In *SLTU*.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020b. Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805.

Ganda and Madison. 2014. Social media and self: Influences on the formation of identity and understanding of self through social networking sites. page 55. University Honors Theses.

Nikhil Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITK@LT-EDI-EACL2021: Hope speech detection for equality, diversity, and inclusion in Tamil , Malayalam and English. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 197–203, Kyiv. Association for Computational Linguistics.

Kris Gowen, Matthew Deschaine, Darcy Gruttadara, and Dana Markey. 2012. Young adults with mental health conditions and social networking websites: Seeking tools to build community. *Psychiatric rehabilitation journal*, 35:245–50.

Junaida M K and Ajees A P. 2021. KU_NLP@LT-EDI-EACL2021: A multilingual hope speech detection for equality, diversity, and inclusion using context aware embeddings. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 79–85, Kyiv. Association for Computational Linguistics.

Skanda Muralidhar, Laurent Nguyen, and Daniel Gatica-Perez. 2018. Words worth: Verbal content and hirability impressions in YouTube video resumes. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 322–327, Brussels, Belgium. Association for Computational Linguistics.

Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@LT-EDI-EACL2021-hope speech detection: There is always hope in transformers. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 98–106, Kyiv. Association for Computational Linguistics.

Manoel Horta Ribeiro, Pedro H. Calais, Yuri A. Santos, Virgílio A. F. Almeida, and Wagner Meira Jr. 2018. "like sheep among wolves": Characterizing hateful users on twitter. *CoRR*, abs/1801.00317.

Ziqi Zhang, David Robinson, and Jonathan Tepper. 2018. Detecting hate speech on twitter using a convolution-gru based deep neural network. In *The Semantic Web*, pages 745–760, Cham. Springer International Publishing.