TEAM HUB@LT-EDI-EACL2021: Hope Speech Detection Based On Pre-trained Language Model

Bo Huang
School of Information Science
and Engineering Yunnan University,
Yunnan, P.R. China
hublucashb@gmail.com

Yang Bai
School of Information Science
and Engineering Yunnan University,
Yunnan, P.R. China
baiyang.top@gmail.com

‘

Abstract

This article introduces the system description of TEAM _HUB team participating in LT-EDI 2021: Hope Speech Detection. This shared task is the first task related to the desired voice detection. The data set in the shared task consists of three different languages (English, Tamil, and Malayalam). The task type is text classification. Based on the analysis and understanding of the task description and data set, we designed a system based on a pre-trained language model to complete this shared task. In this system, we use methods and models that combine the XLM-RoBERTa pre-trained language model and the Tf-Idf algorithm. In the final result ranking announced by the task organizer, our system obtained F1 scores of 0.93, 0.84, 0.59 on the English dataset, Malayalam dataset, and Tamil dataset. Our submission results are ranked 1, 2, and 3 respectively.

1 Introduction and Background

According to relevant statistics, as of 2020, the number of people active in social media worldwide has reached 3.6 billion. Based on the analysis of the current growth rate, in the next 5 years, the total number of social media users will be 1.3 times the current number (Clement, 2020). The COVID-19 virus began to spread globally in 2019, and so far, more than 200 countries and regions have been affected by the epidemic to varying degrees. The number of fake news and negative comments circulating on online social media is 9 times the usual (Bhardwaj et al., 2020). In the past few years, eliminating abusive and hate speech in online social media comments has been widely used on various social media platforms. Most of these elimination systems are developed for languages with richer corpus resources, such as Chinese, Spanish, English, etc. (Malmasi and Zampieri, 2017a; Fortuna and Nunes, 2018; Chetty and Alathur, 2018; Pereira-Kohatsu et al., 2019). There is not much work on Hope Speech detection For example, Chakravarthi et al. developed a data set of hope speech to complete the recognition of Hope Speech in online social media (Chakravarthi, 2020).

The task goal proposed by the task organizer team of “Shared Task on Hope Speech Detection for Equality, Diversity and Inclusion” is to identify the content of posts related to *Hope speech, Not hope speech*, and *Not in intended language* from the dataset of hope speech obtained on YouTube (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). So this sharing task is a comment/post level classification task. Also, this is the first task of identifying hopeful speeches for equality, diversity, and inclusiveness in a multilingual environment. The annotation data set we get contains three different languages (English, Malayalam, and Tamil).

In recent studies in the field of natural language processing(NLP), the pre-trained language model can achieve state-of-the-art results in many NLP tasks (Wang et al., 2019). Also, combined with the task description and the multilingual characteristics of the data set, in terms of model and method selection, we chose to use fine-tuned XLM-RoBERTa (Conneau et al., 2019) to complete our task. We have not done much work on data preprocessing. Our work focuses mainly on how to reduce the impact of multi-language and mixed code on the results. This is also a difficult problem that needs to be solved in this shared task. Therefore, in the method, we consider combining the semantic information of Term Frequency-Inverse Document Frequency (Tf-Idf) and XLM-RoBERTa. Then use Inception blocks (Szegedy et al., 2015) as the shared layer to learn the output information of Tf-Idf and XLM-RoBERTa. Then use this combined model to train on the training set of this task to get the

Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, pages 122–127
April 19, 2021. ©2020 Association for Computational Linguistics

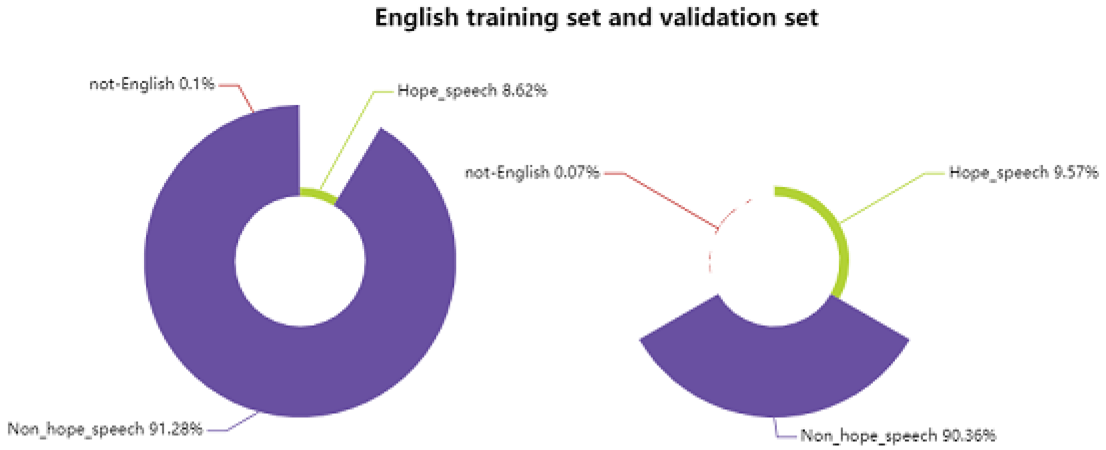

Figure 1: Labels distribution of English training set and validation set. In the training set, Non hope speech: 91.28%, Hope speech: 8.62%, Not English: 0.1%. In the validation set, Non hope speech: 90.36%, Hope speech: 9.57%, Not English: 0.07%.

trained weights. Finally, the trained weights are used to predict the result of the test set.

2 Related Work

So far, many organizations and teams have invested in the identification of negative content on social media platforms and have achieved some good results (Zampieri et al., 2020; Rangel et al., 2020; Chakravarthi et al., 2020; Malmasi and Zampieri, 2017b).

However, combined with our previous analysis, it is not difficult to know that just identifying negative content in social media is not complete. Because not only do we need to avoid negative information in our lives, we also need more positive and hopeful information (Chang, 2017). Some recent studies and reports have also revealed to us the importance of hope and the impact of hopelessness (Eslami et al., 2017; Hernandez and Overholser, 2020; Ranzadeh and Arsh Akmal, 2020). There is not much research and work on Hope Speech detection, but the recent research and results of Hope Speech detection by Palakodety et al. have revealed to us that Hope Speech detection has begun to attract the attention of the academic community (Palakodety et al., 2019).

Chakravarthi et al. used Support vector machine (SVM), Naive Bayesian (NB), k-nearest neighbors (KNN), Decision Tree (DT), Logistic Regression (LR), and other methods on the hope speech data set to get good results (Chakravarthi, 2020). Combine their work on the hope speech data set and the code-mixing in the data set in this task. We chose a pre-trained language model(XLM-RoBERTa) that is currently widely used in the field of natural language processing and fine-tuned it to complete the task.

3 Data And Methods

3.1 Data Description and Analysis

The task description shows us that the data used in the task comes from some comments and posts on YouTube. The data set contains three languages (English, Malayalam, and Tamil) training set and a validation set.

- **Hope_speech**: The comments express the praise of love, courage, interpersonal skills, beauty, perseverance, forgiveness, tolerance, future consciousness, and talent and wisdom. Comments that promote harmony, beauty, tolerance, and bring positive effects to people.

- **Non_Hope_speech**: The comments contain prejudices, attacks on individuals, antihumanity, racial discrimination, and other content that cannot make readers feel hopeful.

- **Not_English, Not_Malayalam, Not_Tamil**: There are languages other than English, Malayalam, or Tamil in the data set of languages English, Malayalam, and Tamil. For example, there are comments posted in the language Tamil in the English data set.

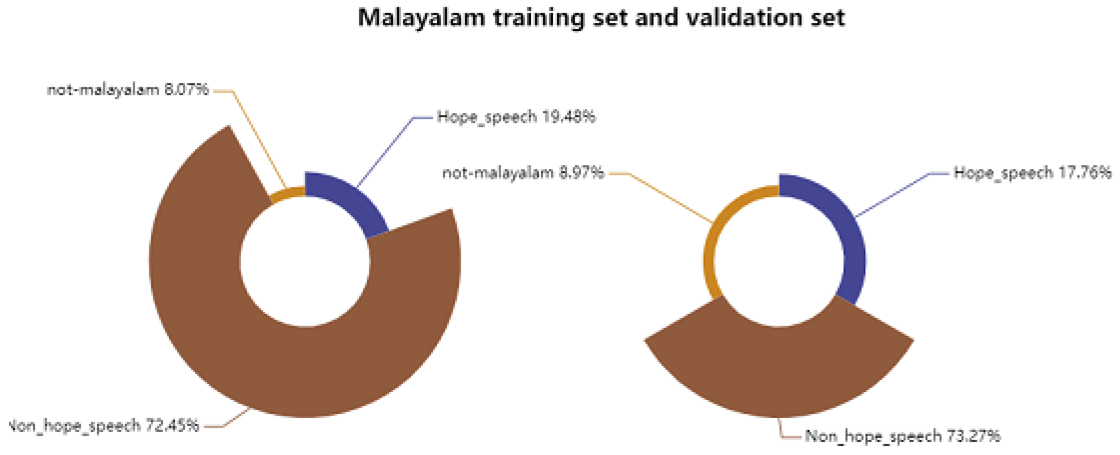

Figure 2: Labels distribution of Malayalam training set and validation set. In the training set, Non hope speech: 72.45%, Hope speech: 19.48%, Not Malayalam: 8.07%. In the validation set, Non hope speech: 73.27%, Hope speech: 17.76%, Not Malayalam: 8.97%.

On three different language data sets, the distribution probability of the number of labels on the training set and the validation set in each language data set is similar. Besides, in the three different language data sets, the data quantity distribution of the three different labels is very unbalanced. In particular, the number of the three labels Not_English (training set 0.1%, validation set 0.07%), Not_Malayalam (training set 8.07%, validation set 8.97%), Not_Tamil (training set 12.14%, validation set 13.03%) is very small in the English, Malayalam, and Tamil data sets. In the data sets of these three languages, the largest value of the labeling ratio in the data set of each language is the Non_Hope_speech label, which accounts for half of the total data volume, or much greater than half of the total data volume.

In addition to the above-mentioned feature of unbalanced data distribution for each label, the task data set also has a big feature that there is some code-mixing language data. Code-mixing language means that a post may be composed of two or more languages. Because users who use social media platforms have different cultural backgrounds and language habits, and posts on social media have no restrictions on grammar and text content structure, the result is a large number of code-mixing languages appearing on social media (Gupta et al., 2018). For us, in addition to data imbalance, code-mixing language is also one of the difficulties we have to face.

3.2 Methods

Based on our previous analysis of the task description and task data, combined with the characteristics of the pre-training language model, the pre-training language model we chose in this task is the XLM-RoBERTa model.

The structure of the XLM-RoBERTa pre-training language model can be seen as a combination of the XLM model and the RoBERTa model (Liu et al., 2019b). The structure of RoBERTa and BERT are all improvements from the Encoder part of the Transformer. Transformer has achieved good results on multiple tasks in the natural language field because of the addition of the Attention mechanism (Vaswani et al., 2017). Compared with Bert, RoBERTa deletes the task of predicting the next sentence in the pre-training stage, and also uses a new dynamic Masking mechanism. In terms of training data sets, XLM-RoBERTa uses a corpus of more than 100 different languages that is larger than the multilingual BERT. Therefore, XLM-RoBERTa's performance in the pre-training phase and certain downstream tasks is better than BERT. Besides, it is also superior to multilingual BERT in terms of cross-language functions.

To alleviate the undesirable effects of code-mixing on the results, we introduce Tf-Idf in our method. Use Tf-idf to weigh with the output of the last layer of XLM-RoBERTa to get a weighted output. Then input this weighted output and the output of the last layer of XLM-RoBERTa into the same Inception block. Use convolution kernels of

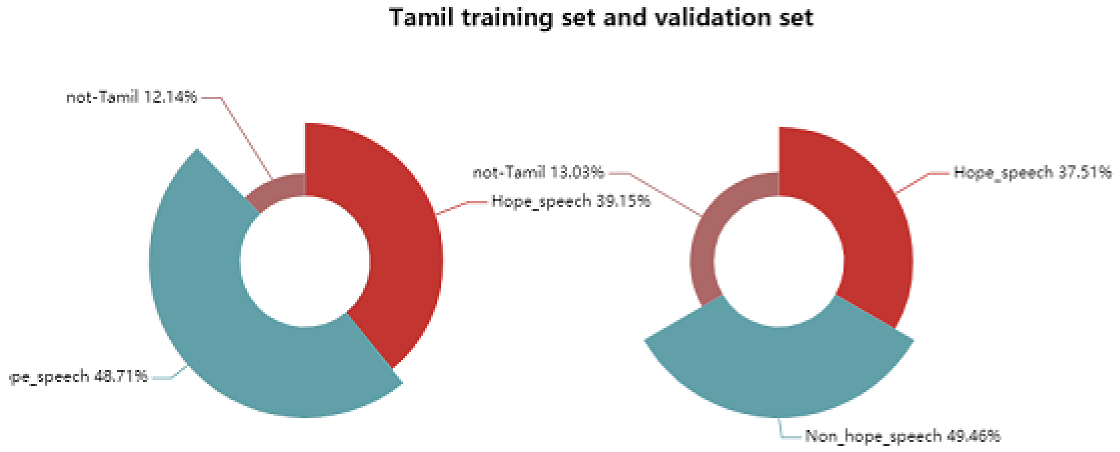

Figure 3: Labels distribution of Tamil training set and validation set. In the training set, Not hope speech: 48.71%, Hope speech: 39.15%, Not Tamil: 12.14%. In the validation set, Not hope speech: 49.46%, Hope speech: 37.51%, Not Tamil: 13.03%.

different sizes in the Inception block to capture semantic information of different lengths. So in this model, the inception block can be seen as a module with shared parameters. Use this module to learn the output of XLM-RoBERTa and the weighted semantic information of Tf-Idf. Finally, the output of the Inception module is obtained and input into the classifier to obtain the prediction result of the model. Compared with the sequence structure of LSTM, CNN can better identify the semantic information of a post on a social media platform. Because the order of some words in YouTube will be changed by various noises. At the same time, CNN structure will have better performance than serialized LSTM in training time. The structure of the model can be seen in Figure 4.

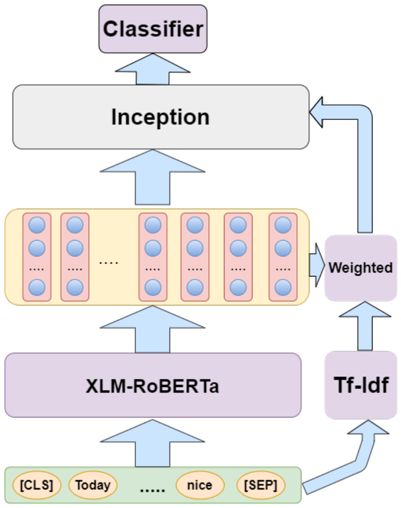

Figure 4: The model structure we used in this task.

4 Experiment and Results

4.1 Data Preprocessing

In terms of data preprocessing, our work is mainly focused on the corresponding Tf-Idf encoding from each language data set using the Tf-Idf algorithm. We merge the training set, and validation set of each language. Then use the Tf-Idf algorithm to get the Tf-Idf code of the sentence. Delete the Tf-Idf codes larger than the maximum sentence length, and perform zero-filling operations on the Tf-Idf codes smaller than the maximum sentence length. Finally, the processed Tf-Idf code is input into the model and combined with the output of XLM-RoBERTa.

4.2 Experiment setting

In the choice of the pre-training model XLM-RoBERTa, we use the XLM-RoBERTa-base pre-training language model composed of a 12-layer Encoder[1]. The structure of Inception we use is the solution implemented by Szegedy et. (Szegedy et al., 2015). Besides, because we are dealing with text data, the Conv1d convolution provided by Pytorch is used in the Inception block[2]. Combined with our previous analysis of the task and data, we choose CrossEntropyLoss in the loss function. We use Radam as the optimizer (Liu et al., 2019a). Because the data set category distribution of each

[1] https://huggingface.co/xlm-roberta-base/tree/main
[2] https://pytorch.org/

Language	F1 Score	Precision	Recall
English	0.93	0.93	0.93
Malayalam	0.85	0.85	0.85
Tamil	0.62	0.62	0.64

Table 1: The result of our model and method on the validation set provided by the task organizer.

Language	F1 Score	Precision	Recall
English	0.93	0.93	0.93
Malayalam	0.84	0.84	0.85
Tamil	0.59	0.61	0.61

Table 2: The results of our model and method on the test set. The score of the test set comes from the ranking list announced by the task organizer.

language is different, we set different hyperparameters for each language task. The parameters used in our experiments are all parameter combinations verified on the validation set.

- **English task**: The learning rate of XLM-RoBERTa is set to 3e-5, and the learning rate of the Inception block and linear classifier is set to 2e-4. The maximum sentence length is set to 65. Epoch and batch size are set to 5 and 32 respectively.

- **Malayalam task**: The learning rate of XLM-RoBERTa is set to 4e-5, and the learning rate of the Inception block and linear classifier is set to 1e-4. The maximum sentence length is set to 70. Epoch and batch size are set to 5 and 32 respectively.

- **Tamil task**: The learning rate of XLM-RoBERTa is set to 3e-5, and the learning rate of the Inception block and linear classifier is set to 1e-4. The maximum sentence length is set to 70. Epoch and batch size are set to 4 and 32 respectively.

4.3 Analysis of Results

The scores of each team on the test set announced by the task organizer team are ranked using weighted average F1-score. At the same time, the scores of the two evaluation indicators Precision and Recall are also announced.

Compare the contents of Table 1 and Table 2, the scores(F1 Score, Precision, Recall) of our model and method on the validation set are the same as those on the test set. By comparing the result score on the validation set with the score result on the test set, it is not difficult to find that there is a gap between the result score on the Tamil validation set and the result score on the test set. The scores of the validation set and the test set of the other two languages are very consistent. Among the three different languages, our model and method ranked first in the English task ranking, second in the Malayalam language task ranking, and third in the Tamil language task ranking.

5 Conclusion

In this task, our team used the XLM-RoBERTa model with cross-language capabilities as the basis and combined it with the Tf-Idf algorithm to complete this task. Our experimental results confirm the effectiveness of our method, and at the same time find the problem of our method and parameter settings on the Tamil language task data set. Our work in the data preprocessing part is not detailed enough. We ignore the influence of stop words, emoticons, and other special symbols on the results. This is a flaw in our work. In future work, we will make up for the shortcomings of our work in this task and improve our methods. Besides, we will continue to pay attention to the progress of related work on low-resource language communities, and the progress of hope speeches in the field of natural language processing.

References

Mohit Bhardwaj, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2020. Hostility detection dataset in hindi. *arXiv preprint arXiv:2011.03588*.

Bharathi Raja Chakravarthi. 2020. HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53, Barcelona, Spain (Online). Association for Computational Linguistics.

Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. Findings of the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

BR Chakravarthi, R Priyadharshini, V Muralidaran, S Suryawanshi, N Jose, E Sherly, and JP McCrae. 2020. Overview of the track on sentiment analysis for dravidian languages in code-mixed text. In *Working Notes of the Forum for Information Retrieval*

Evaluation (FIRE 2020). CEUR Workshop Proceedings. In: CEUR-WS. org, Hyderabad, India.

Edward C Chang. 2017. Hope and hopelessness as predictors of suicide ideation in hungarian college students. *Death studies*, 41(7):455–460.

Naganna Chetty and Sreejith Alathur. 2018. Hate speech review in the context of online social networks. *Aggression and violent behavior*, 40:108–118.

Jessica Clement. 2020. Number of global social network users 2010-2023. *Retrieved April*, 30:2020.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Bahareh Eslami, Adrienne H Kovacs, Philip Moons, Kyomars Abbasi, and Jamie L Jackson. 2017. Hopelessness among adults with congenital heart disease: Cause for despair or hope? *International journal of cardiology*, 230:64–69.

Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4):1–30.

Deepak Gupta, Pabitra Lenka, Asif Ekbal, and Pushpak Bhattacharyya. 2018. Uncovering code-mixed challenges: A framework for linguistically driven question generation and neural based question answering. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 119–130.

Silvia C Hernandez and James C Overholser. 2020. A systematic review of interventions for hope/hopelessness in older adults. *Clinical gerontologist*, pages 1–15.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. 2019a. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Shervin Malmasi and Marcos Zampieri. 2017a. Detecting hate speech in social media. *arXiv preprint arXiv:1712.06427*.

Shervin Malmasi and Marcos Zampieri. 2017b. Detecting hate speech in social media. *arXiv preprint arXiv:1712.06427*.

Shriphani Palakodety, Ashiqur R KhudaBukhsh, and Jaime G Carbonell. 2019. Hope speech detection: A computational analysis of the voice of peace. *arXiv preprint arXiv:1909.12940*.

Juan Carlos Pereira-Kohatsu, Lara Quijano-Sánchez, Federico Liberatore, and Miguel Camacho-Collados. 2019. Detecting and monitoring hate speech in twitter. *Sensors*, 19(21):4654.

Francisco Rangel, Anastasia Giachanou, Bilal Ghanem, and Paolo Rosso. 2020. Overview of the 8th author profiling task at pan 2020: Profiling fake news spreaders on twitter. In *CLEF*.

Nematollah Ranzadeh and Shahnaz Arsh Akmal. 2020. The representation of thoughts of death in relation to hope and hopelessness in malakut. *Literary Text Research*.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Chenguang Wang, Mu Li, and Alexander J Smola. 2019. Language models with transformers. *arXiv preprint arXiv:1904.09408*.

Marcos Zampieri, Preslav Nakov, Sara Rosenthal, Pepa Atanasova, Georgi Karadzhov, Hamdy Mubarak, Leon Derczynski, Zeses Pitenis, and Çağrı Çöltekin. 2020. Semeval-2020 task 12: Multilingual offensive language identification in social media (offenseval 2020). *arXiv preprint arXiv:2006.07235*.