Simon @ DravidianLangTech-EACL2021: Detecting Hope Speech with BERT

Qinyu Que
School of Information Science and Engineering,
Yunnan University, Yunnan, P.R. China
1309487642@qq.com

Abstract

In today's society, the rapid development of communication technology allows us to communicate with people from different parts of the world. In the process of communication, each person treats others differently. Some people are used to using offensive and sarcastic language to express their views. These words cause pain to others and make people feel down. Some people are used to sharing happiness with others and encouraging others. Such people bring joy and hope to others through their words. On social media platforms, these two kinds of language are all over the place. If people want to make the online world a better place, they will have to deal with both. So identifying offensive language and hope language is an essential task. There have been many assignments about offensive language. Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI 2021-EACL 2021 uses another unique perspective – to identify the language of Hope to make contributions to society. The XLM-Roberta model is an excellent multilingual model. Our team used a fine-tuned XLM-Roberta model to accomplish this task.

1 Introduction

Social media is a way to bring people from different regions closer together. Different people have different cultural backgrounds and different world views. This allows people to have a fierce clash of ideas on social platforms. In these interactions, the way people treat each other is very different. Hate speech offensive speech and offensive speech are not recommended (Chakravarthi et al., 2020; Mandl et al., 2020; Chakravarthi et al., 2021; Suryawanshi and Chakravarthi, 2021). So we also do a lot of work to avoid this, for example, we have a lot of competitions to identify offensive content(GermEval[1], IberLEF[2]). And we welcome those who use the language of hope to encourage people to get back on their feet and renew their strength in the face of hardship. To this end, Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI 2021-EACL 2021 was organized by the event organizers. Hope Speech can help people find ways to realize their dreams(Blasius and Phelan, 1997; Chang, 1998; Youssef and Luthans, 2007; Cover, 2013; Snyder et al., 1991). People can learn and master these techniques to make their lives better. Hope Speech is seen by medical personnel as vital to people's recovery. Hope speech makes the whole social environment more optimistic. Therefore, it is very necessary to detect hope speech. The second paragraph of the paper introduces the work done by scholars in relevant fields, the third paragraph introduces the detailed content of tasks, the fourth paragraph introduces the details of data and the methods of processing data, the fifth paragraph introduces the model used by our team, the sixth paragraph introduces the details and results of experiments, and finally the conclusion we have reached.

2 Related Work

Some works have been done on the classification of hope speech. (Puranik et al., 2021; Ghanghor et al., 2021) analyze the evolving international crisis via a substantial corpus constructed using comments on YouTube videos. As well as detecting offensive content, Hope speech can be classified as a useful tool in its own field. Mathur et al. (Mathur et al., 2018) used transfer learning to complete the task of classifying offensive content of tweets in hot data sets. Their team used a convolutional neural network to pre-train English tweets and then re-train

[1] https://projects.fzai.h-da.de/iggsa/germeval/
[2] http://hitz.eus/sepln2019/?q=node/21

Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, pages 175–179
April 19, 2021.

Indo-English tweets. Kamble and Joshi (Kamble and Joshi, 2018) used three classical deep learning models to complete the classification of English-Hindi tweets. By comparing these three classical deep learning models, they found that the use of specific embedding can improve the representation of the target group, and thus improve the score of F1. Santosh and Aravind (Santosh and Aravind, 2019) used different LSTM(Mathur et al., 2018) models to classify hate speech, one is the sub-word level LSTM model and the other is the hierarchical LSTM model based on phonetic location words. Waseem et al.(Waseem and Hovy, 2016) found the influence of different extra language features combined with N-gram characters on classification tasks.

3 Task Details

Shared task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI 2021-EACL 2021 gives the participants three languages to choose from. The three languages are English, Tamil and, Malayalam. The event organizer provides us with a training data set and a test data set. Through the data set, we find that this is a tripartite task. Our group believes that a lot of work has already been done on the task of classifying English. So our team chose a Tamil sub-task and a Malayalam sub-task.

4 Date Details and Data Preprocessing

4.1 Date Description

In each subtask, the data set contains three tags:

- Hope speech: This tag indicates that the comment is Hope speech.
- Not hope speech: This tag indicates that a comment is Not hope speech.
- Not in intended language: This tag indicates that is not the language used by the subtask.

We can see that the label classification is very clear. The details of the data set are showed in Table 1 and Table 2.

4.2 Data Preparation

The data sets we used were collected from YouTube (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). Before training the dataset, we first clean up the dataset. The cleaned data set is more standardized, and we remove miscellaneous words and symbols, reducing the amount of training. The main work of data set cleaning is as follows:

Label	Train set	validation set
Non hope speech	72.45%	73.27%
Hope speech	19.48%	17.76%
Not Malayalam	8.07%	8.97%

Table 1: Label distribution of Malayalam language subtask

Label	Train set	validation set
Non hope speech	48.71%	49.46%
Hope speech	39.15%	37.51%
Not Tamil	12.14%	13.03%

Table 2: Label distribution of Tamil language subtask

- Deleting the emoticons
- Replacing emojis with words, such as words 'crying' instead of a crying face emoji, and words 'panda' instead of a panda emoji.
- Deleting the URL
- Deleting punctuation marks

5 Model Description

In a recent categorization task competition, many teams using the fine-tuned BERT(Devlin et al., 2019) model did well. Our team used XLM-Roberta(Conneau et al., 2020) as a preprocessing model in the experiment. The XLM-Roberta model is a multilingual processing model trained on processing text in 100 different languages. This fits nicely with the shared task of handling multilingual text. Compared with the BERT model, the XLM-Roberta model has improved various performances. This is due to the increased training volume of the XLM-Roberta model, whose training datasets are several orders of magnitude larger than the Wiki-100 corpus used to train its predecessor. Based on the XLM-Roberta model, we used two fine-tuning methods to accomplish this task. The two approaches are shown in Figures 1 and 2.

5.1 Method 1

Firstly, the pooler output of XLM-Roberta (P_O) of XLM-Roberta model is obtained through XLM-Roberta model. Second, input the output of the last three hidden layers of the XLM-Roberta model

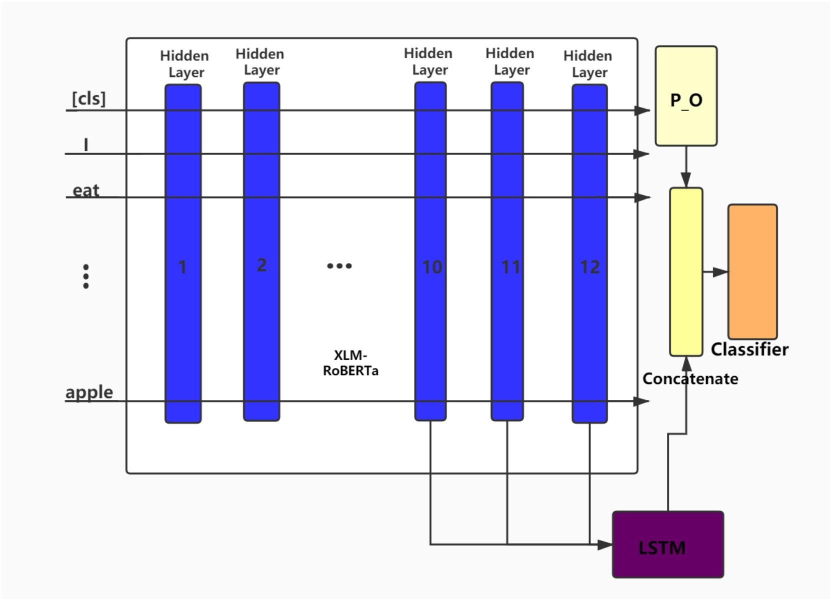

Figure 1: Method 1

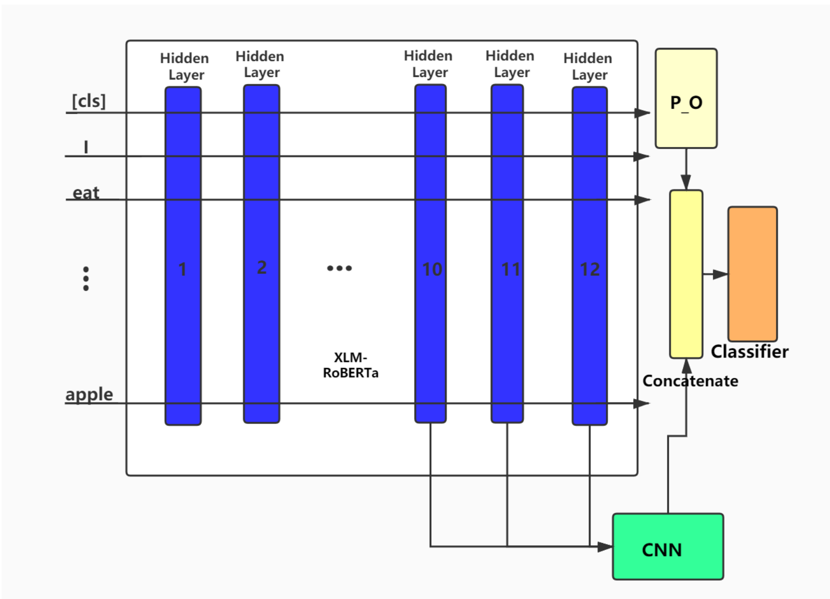

Figure 2: Method 2

Hyper-parameter	Value
dropout	0.5
learning rate	1e-5
epoch	8
per gpu train batch size	32
gradient accumulation steps	8

Table 3: The hyper-parameters

into the LSTM model. Finally, the output obtained from the LSTM model is connected with P_O and put into the classifier.

5.2 Method 2

Firstly, the Pooler output of XLM-Roberta (P_O) of the XLM-Roberta model is obtained through the XLM-Roberta model. The second step is to input the output of the last three hidden layers of the XLM-Roberta model into the CNN(Simonyan and Zisserman, 2014) model. Finally, the output obtained from the CNN model is connected with P_O and put into the classifier.

6 Experiment and Results

In the experimental process, the same hyper-parameter is used in the two methods. Table 3 is the details of the hyper-parameter we used. The training set and validation set used by our team were derived from the training set provided by the event organization after separation using hierarchical K-fold cross-validation. Method 1 performed better on the validation set, so we submitted the results of Method 1. In the Tamil language sub-task, our final F1-score In the official test set is 0.49. In the Malayalam language sub-task, our final F1-score In the official test set is 0.49.

7 Conclusion

This article describes the models and results used by the Simon team in the hope speech classification task. Our results are not good in the leaderboard, which may have something to do with the imbalance of the data set. In the future, we will continue to improve our model to get better results.

References

Mark Blasius and Shane Phelan. 1997. *We are everywhere: A historical sourcebook of gay and lesbian politics*. Psychology Press.

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, M Anand Kumar, John Philip McCrae, Premjith B, Soman KP, and Thomas Mandl. 2020. Overview of the track on HASOC-Offensive Language Identification-DravidianCodeMix. In *Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2020). CEUR Workshop Proceedings. In: CEUR-WS. org, Hyderabad, India*.

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.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Navya Jose, Anand Kumar M, Thomas Mandl, Prasanna Kumar Kumaresan, Rahul Ponnusamy, Hariharan V, Elizabeth Sherly, and John Philip McCrae. 2021. Findings of the shared task on Offensive Language Identification in Tamil, Malayalam, and Kannada. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Edward C Chang. 1998. Hope, problem-solving ability, and coping in a college student population: Some implications for theory and practice. *Journal of clinical psychology*, 54(7):953–962.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Rob Cover. 2013. Queer youth resilience: Critiquing the discourse of hope and hopelessness in lgbt suicide representation. *M/C Journal*, 16(5).

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Nikhil Kumar 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*, Online.

Satyajit Kamble and Aditya Joshi. 2018. Hate speech detection from code-mixed hindi-english tweets using deep learning models. *arXiv preprint arXiv:1811.05145*.

Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2020. Overview of the HASOC Track at FIRE 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German. In *Forum for Information Retrieval Evaluation*, FIRE 2020, page 29–32, New York, NY, USA. Association for Computing Machinery.

Puneet Mathur, Rajiv Shah, Ramit Sawhney, and Debanjan Mahata. 2018. Detecting offensive tweets in hindi-english code-switched language. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 18–26.

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*. Association for Computational Linguistics.

TYSS Santosh and KVS Aravind. 2019. Hate speech detection in hindi-english code-mixed social media text. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pages 310–313.

Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *Computer Science*.

Charles R Snyder, Cheri Harris, John R Anderson, Sharon A Holleran, Lori M Irving, Sandra T Sigmon, Lauren Yoshinobu, June Gibb, Charyle Langelle, and Pat Harney. 1991. The will and the ways: development and validation of an individual-differences measure of hope. *Journal of personality and social psychology*, 60(4):570.

Shardul Suryawanshi and Bharathi Raja Chakravarthi. 2021. Findings of the shared task on Troll Meme Classification in Tamil. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL student research workshop*, pages 88–93.

Carolyn M Youssef and Fred Luthans. 2007. Positive organizational behavior in the workplace: The impact of hope, optimism, and resilience. *Journal of management*, 33(5):774–800.