ZYJ@LT-EDI-EACL2021:XLM-RoBERTa-Based Model with Attention for Hope Speech Detection

Yingjia Zhao
Yunnan University/ Yunnan, P.R. China
zyj1309700118@gmail.com

Xin Tao
Yunnan University / Yunnan, P.R. China
taoxinwy@126.com

Abstract

Due to the development of modern computer technology and the increase in the number of online media users, we can see all kinds of posts and comments everywhere on the internet. Hope speech can not only inspire the creators but also make other viewers pleasant. It is necessary to effectively and automatically detect hope speech. This paper describes the approach of our team in the task of hope speech detection. We use the attention mechanism to adjust the weight of all the output layers of XLM-RoBERTa to make full use of the information extracted from each layer, and use the weighted sum of all the output layers to complete the classification task. And we use the Stratified-K-Fold method to enhance the training data set. We achieve a weighted average F1-score of 0.59, 0.84, and 0.92 for Tamil, Malayalam, and English language, ranked 3rd, 2nd, and 2nd.

1 Introduction

With the development of network media, people can express their opinions on the Internet at any time, among which there will be some hope speech. Hope speech will encourage people to stand firm in their beliefs and follow the path of their goals. At the same time, comments or posts containing hope speech can often contain factors such as Equality, Diversity, and Inclusion. They can also provide support, reassurance, suggestions, inspiration, and insight, etc., which can change people's emotions from negative to positive. The proportion of comments and posts on media platforms that contain hopeful statements can reflect the current emotional tendencies of online communities. As the proportion of words containing hope increases, the corresponding posts conveying negative emotions such as pessimism and despair will decrease, thus promoting social harmony. Because the net-work media platform needs to support multiple languages and faces users with different backgrounds, the multi-lingual, multi-modal, and non-standard writing style of online media posts makes it a very challenging task to effectively detect whether the comment or post belongs to the hope speech.

Our team participated the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI 2021- EACL 2021 (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). The goal of this task is to identify whether a given comment contains hope speech or not. This is a comment/post level classification task. The task is divided into three sub-tasks based on three languages, namely English, Tamil, and Malayalam. Given a Youtube comment, the systems submitted by the participants should classify it into 'Hope speech', 'Not hope speech' and 'Not in intended language'.

In this task, we use XLM-RoBERTa as the base model to extract features and combine 12 layers of output with an embedding layer to do the classification task. In order to train the weight of each layer, we add an attention mechanism for the model. During the training, we process the training data through the Stratified-K-Fold method, each data set is trained to get the different weights, and finally, the best-predicted value is obtained through a voting mechanism. The rest of our paper is structured as follows. Section 2 describes the related work. Methods are described in Section 3. Experiments and results are described in Section 4. The conclusion is drawn in Section 5.

2 Related Work

While hope speech detection is a relatively new task, it can be categorized as a problem of sentiment analysis (Bakshi et al., 2016) and text classification. A great deal of work has been done by many researchers and practitioners from in-

dustry and academia. Turney (2002)proposed a simple unsupervised learning (Barlow, 1989) algorithm for classifying reviews as recommended or not recommended. Basiri et al. (2020)proposed an Attention-based Bidirectional CNN-RNN Deep Model (ABCDM) for polarity classification of long comments and short tweets, two independent bidirectional LSTM and GRU layers are used to extract past and future context by considering temporal information flow in both directions. Chen and Ke proposed a convolutional neural network-regional long short-term memory (CNN-RLSTM) that combines CNN and regional LSTM. At the sentence level (Kim and Hovy, 2004), the emotional information of the whole sentence is retained through CNN network, and through regional LSTM effectively distinguish between different target emotional polarity. Chen et al. proposed a multi-head attention-memory network (MNHMA) based on hierarchy for aspect-based sentiment analysis, which makes full use of the long-term semantic relationship of a given aspect term (Manandhar, 2014) in a sentence to reduce the loss of aspect information.

Wang et al. (2014) demonstrated the effectiveness of ensemble learning in sentiment analysis on ten publicly available datasets. Akhtar et al. (2017) proposed a cascade framework based on feature selection and classifier ensemble (Avidan, 2007). Particle swarm optimization algorithm is used for feature-based sentiment analysis. Because of the powerful performance of BERT (Devlin et al., 2018), it is natural to use it for sentiment analysis. Malhotra et al. proposed an efficacious transfer learning based ensemble model for sentiment analysis.

3 Methodology and Data

3.1 Data description

We make a statistic based on the data set provided by the organizer. We count the proportions of 'hope speech', 'not hope speech', and 'not in integrated language' in the training set and verification set. To make an intuitive observation, the statistical situation of three language datasets, English, Tamil, and Malayalam, are presented in one single table. As shown in Table 1.

3.2 XLM-RoBERTa with attention

In addition to the English dataset, this shared task also involves the training and classification of low-resource languages such as Tamil, and Malayalam.

	Label	Train	Dev
English	Hope speech	1960	272
	Not hope speech	20778	2569
	not-English	22	2
Malayalam	Hope speech	1668	190
	Not hope speech	6205	784
	not-Malayalam	691	96
Tamil	Hope speech	6327	757
	Not hope speech	7872	998
	not-Tamil	1961	263

Table 1: Train and Validation datasets description.

Therefore, XLM-RoBERTa, a pretraining enhancement model based on multiple languages, is undoubtedly the best choice. To further improve the performance of the model, we add a specific network structure based on XLM-RoBERTa. Since the output of each layer of XLM-RoBERTa contains different information, we combine the 12 layers of its output with the embedding layer to do the classification task. To adjust the weight of each layer, the attention mechanism is used to train the weight of each layer. Finally, the weighted sum of all layers is input into the classifier for the classification task, as shown in Figure 1.

4 Experiment and results

4.1 Data preprocessing

Since many words and symbols on social media are useless and sentence expression is not standardized enough, in order to extract useful features more effectively and reduce interference during training the model, we need to clean the text in the training data set with NLTK (Loper and Bird, 2002) tool. We perform the following preprocessing steps.

- **Remove Emoticons.** Text data contains a lot of emoticons. Emoticons increase the number of unknown words and reduce training effects. Emoticons are replaced by emotional words with their own meanings.

- **Convert Abbreviations.** All abbreviations are converted to complete parts. For English datasets, this helps the machine understand the meaning of words (e.g. "there's" becomes 'there' and 'are').

- **Remove Stop Word.** For English datasets, the stop words are meaningless words, such as'
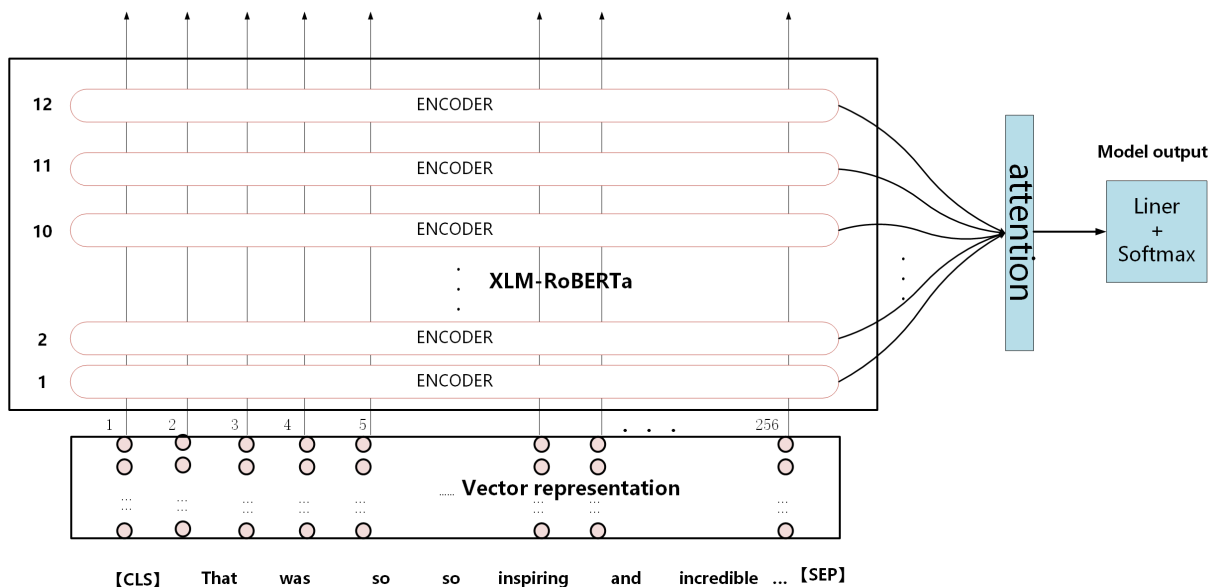
Figure 1: Schematic overview of the architecture of our model

is' /' our '/' the' /' in '/'at' etc. They don't add much meaning to the sentence. Single stop words are very frequent words. To reduce the number of vocabularies we have to deal with, and thus reduce the complexity of subsequent programs, we need to clear out stop words.

- **Stemming and Lemmatization.** For English datasets, to further simplify the textual data, we can standardize the different variations and inflections of words. Stemming extraction is the process of reducing a word to its stem or root. For example 'branching'/' branched '/' branches', can all be reduced to 'branch'. This helps reduce complexity while retaining the basic meaning of the word. For example, the suffixes 'ing' and 'ed'can be discarded;'ies' can be replaced by 'y' and so on. This may result in an incomplete word stem, but only as long as all forms of the word are reduced to the same stem. So they all have the same fundamental meaning.

4.2 Stratified-K-Fold method

Before training, we use the Stratified-K-fold method to process the dataset. The core of Stratified-K-Fold is to ensure stratified sampling. It needs to ensure that the ratio of all kinds of samples in test sets and training sets is the same as in the original data set, namely k-fold crossover segmentation. In the initial training set, each type of data is divided into k sub samples, one single subsample is reserved as the data of validation model, and the

maximum sequence length	learning rate
256	2e-5
Stratified-K-Fold number	**batch size**
5	4

Table 2: Details of the parameters

other k-1 samples are used for training. Finally, we get the final output from k results based on the voting (Onan et al., 2016) method.

4.3 Experiment setting

In this experiment, XLM-RoBERTa-base pre-trained model is used. On the basis of the pre-training model, we add the attention mechanism to adjust the structure of the whole model. Then we began to train the model. The batch size is set to 4, the max sequence length is set to 256, the learning rate is set to 2e-5, and the Stratified-K-Fold number is set to 5. As shown in Table 2.

4.4 Results

The results submitted by teams are evaluated with a weighted average F1-score. According to the leaderboard provided by the organizer, our team's F1-score is 0.59, ranked 3rd place for the Tamil language. For the Malayalam language, our team's F1-score is 0.84 ranked 2nd place, and for the English language, our team's F1-score is 0.92 ranked 2nd place. As shown in Table 3.

120

	Best F1-score	Our Precision	Our Recall	Our F1-score	Rank
Tamil	0.61	0.59	0.59	**0.59**	3rd
Malayalam	0.85	0.84	0.84	**0.84**	2nd
English	0.93	0.92	0.93	**0.92**	2nd

Table 3

5 Conclusion

In this paper, we present our work in the task of hope speech detection for Tamil, Malayalam, and English language. Our team uses a model based on XLM-RoBERTa. This model uses an attention mechanism to adjust the weight of each output layer, makes full use of the information extracted from each layer to complete the classification task, and the Stratified-K-Fold method is used to enhance the training data. Finally, the model we submit achieves significant performance for all three languages. For future work, we will further optimize our model to achieve better results.

References

Md Shad Akhtar, Deepak Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2017. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems*, 125:116–135.

Shai Avidan. 2007. Ensemble tracking. *IEEE transactions on pattern analysis and machine intelligence*, 29(2):261–271.

Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. 2016. Opinion mining and sentiment analysis. In *2016 3rd international conference on computing for sustainable global development (INDIACom)*, pages 452–455. IEEE.

Horace B Barlow. 1989. Unsupervised learning. *Neural computation*, 1(3):295–311.

Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Erik Cambria, and U Rajendra Acharya. 2020. Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis. *Future Generation Computer Systems*, 115:279–294.

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.

Ke Chen and Wende Ke. A hierarchical neural model for target-based sentiment analysis. *Concurrency and Computation: Practice and Experience*, page e6184.

Yuzhong Chen, Tianhao Zhuang, and Kun Guo. Memory network with hierarchical multi-head attention for aspect-based sentiment analysis. *Applied Intelligence*, pages 1–18.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 1367–1373.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Shivani Malhotra, Vinay Kumar, and Alpana Agarwal. Bidirectional transfer learning model for sentiment analysis of natural language. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–21.

Suresh Manandhar. 2014. Semeval-2014 task 4: aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*.

Aytuğ Onan, Serdar Korukoğlu, and Hasan Bulut. 2016. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Systems with Applications*, 62:1–16.

Peter D Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.

Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. 2014. Sentiment classification: The contribution of ensemble learning. *Decision support systems*, 57:77–93.