VTUBGM@LT-EDI-2023: Hope Speech Identification using Layered Differential Training of ULMFit

Sanjana Kavatagi[1] and **Rashmi Rachh**[1] and **Shankar Biradar**[2]
[1]Department of Computer Science and Engineering,
VTU,Belagvi, Karnatka, India
[2]Department of Computer Science and Engineering,
IIIT, Dharwad, Karnataka, India
kawatagi.sanjana@gmail.com
rashmirachh@gmail.com

Abstract

Hope speech embodies optimistic and uplifting sentiments, aiming to inspire individuals to maintain faith in positive progress and actively contribute to a better future. In this article, we outline the model presented by our team, VTUBGM, for the shared task "Hope Speech Detection for Equality, Diversity, and Inclusion" at LT-EDI-RANLP 2023. This task entails classifying YouTube comments, which is a classification problem at the comment level. The task was conducted in four different languages: Bulgarian, English, Hindi, and Spanish, with our team participating in the English subtask. The suggested model was developed using layered differential training of the ULMFit and received a macro F1 score of 0.48, placing third in the competition.

1 Introduction

The exponential growth of internet usage has led to a significant rise in the prominence of social media platforms. In recent times, these platforms have evolved to become the primary means of communication for millions of individuals worldwide(Biradar and Saumya, 2022; Shankar Biradar and Chauhan, 2021). As the popularity of social media continues to soar, people are increasingly sharing their thoughts, opinions, and experiences on various issues. This widespread adoption of social media has revolutionized the way individuals communicate and express themselves. Platforms like Facebook, Twitter, Instagram, and YouTube have become integral components of our daily lives. These platforms allow users to express their thoughts, ideas, and feelings to a global audience. In recent times, social media has emerged as a powerful tool for social and political activism, as users leverage these platforms to raise awareness about various social issues and rally support for causes.

Social media has also become a thriving environment for the dissemination of hate speech and fake news, which can spread rapidly across the globe. This harmful content can be targeted toward individuals, groups, religions, or political parties (Biradar et al., 2021; Chakravarthi et al., 2022). Extensive research has been conducted to address this issue and develop strategies to identify and prevent such material's publication and rapid spread on the internet.

While numerous individuals are involved in promoting fake news and hate speech, several groups and individuals are actively working to create and spread positivity and hope. These efforts have surfaced in various social media campaigns to counteract online platforms' negativity and polarisation. Hope is a multifaceted and complex emotion that holds immense importance for human well-being. It encompasses an optimistic outlook on life, enabling us to endure challenging circumstances, nurture the belief that things can get better, and have faith in ourselves and others. Hope has been linked to numerous positive outcomes, such as enhanced physical and mental health, heightened resilience, and increased motivation.

To promote research in the identification of hope speech on social media platforms, the organizers of the "Hope Speech Detection for Equality, Diversity, and Inclusion" shared task at LT-EDI-RANLP 2023 offer an opportunity for researchers to develop machine learning and deep learning models capable of effectively classifying hope speech and non-hope speech within the provided dataset. This initiative encourages the exploration of innovative approaches to accurately distinguish between hopeful and non-hopeful content, thus contributing to the advancement of understanding and harnessing the power of hope in online discourse. In order to construct a model for classifying hope speech and non-hope speech, our team employed the ULMFit

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion,
pages 209–213, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_032

model, an LSTM-based model that was fine-tuned specifically for the task of hope speech detection. Our model achieved third place among the participating teams with a macro F1 score of 0.48.

The remainder of the paper is structured as follows: Section 2 presents a comprehensive literature review, highlighting key studies and research related to the topic. Section 3 explains our proposed model, outlining its architecture and training methodology. The results obtained from our model are discussed in Section 4. Finally, Section 6 concludes the paper by summarizing the key findings and contributions and outlines potential trajectories for future research in this field.

2 Literature Review

While numerous researchers have recently focused on identifying offensive and fraudulent content in social media (Biradar et al., 2022), only a limited number of studies have been conducted on identifying hope within individuals' opinions expressed on social media. In an environment permeated by toxicity, violence, and discrimination, hope speech is a powerful tool that assists and encourages countless needy individuals. It serves as a beacon of hope for those grappling with personal or professional challenges, including issues related to health, finances, relationships, and more (Chakravarthi, 2022). Identifying and automatically detecting such statements can play a crucial role in facilitating their widespread dissemination. By detecting and recognizing these statements, we can amplify their reach and impact, inspiring and motivating the individuals who create them for the betterment of others. This, in turn, fosters a positive and supportive environment where hope speech can serve as a catalyst for positive change, resilience, and empowerment (Palakodety et al., 2019). In the realm of hope speech detection, data is collected by performing web crawling techniques. Researchers have extensively investigated and conducted studies focused on extracting comments and posts from popular social media platforms, including Twitter, Facebook, and YouTube. These platforms serve as rich sources of valuable data for analyzing and understanding hope speech in online discourse (Marrese-Taylor et al., 2017; Muralidhar et al., 2018).

In a pioneering effort, (Chakravarthi, 2020) developed a comprehensive dataset for hope speech encompassing multiple languages such as English, Malayalam, and Tamil. This dataset served as the foundation for the shared task called HopeEDI, designed to promote research and exploration in the field of detecting hopeful and encouraging content within social media. HopeEDI aimed to encourage advancements in understanding and leveraging the power of hope speech in online environments by providing a platform for studying and analyzing such content.

(Dave et al., 2021) directed their attention towards harnessing classic machine learning classifiers, specifically logistic regression (LR) and support vector machine (SVM), for the purpose of categorizing text into hope speech and non-hope speech categories. They utilized TF-IDF and character n-gram techniques for feature extraction to achieve this. By leveraging these advanced classifiers and feature extraction methods, their approach demonstrated promising results in accurately identifying and distinguishing between hope speech and non-hope speech texts. This study exemplifies the effectiveness of employing machine learning techniques for the task of hope speech detection. A novel method that employed an ensemble approach for the classification of hope and non-hope speech was introduced by (Kumar et al., 2022). Their approach involved utilizing both character-level and word-level TF-IDF embedding techniques to extract meaningful features from the text data. By combining these two levels of embeddings in an ensemble model, the authors demonstrated an effective approach for accurately categorizing text into hope and non-hope speech categories. This methodology showcased the potential of leveraging multiple embedding techniques in tandem to improve classification performance in the context of hope speech detection. (Balouchzahi et al., 2021) introduced a strategy that involved extracting features from the given dataset using TF-IDF and syntactic n-grams. They further trained a neural network model and a voting classifier using LR, eXtreme Gradient Boosting (XGB), and MLP algorithms. Additionally, in the context of the HopeEDI shared task, a BERT model was trained specifically to identify hope speech in English. This BERT model showcased remarkable performance, achieving an average F1 score of 0.92. This approach highlights the effectiveness of utilizing advanced language models for hope speech detection, yielding impressive accuracy and classification performance results.

In some of the studies, researchers have utilized transformer models for the identification of the hope speech. The embeddings for the text data were extracted using the BERT (Bidirectional Encoder Representations from Transformers) model, as introduced by (Dowlagar and Mamidi, 2021). The BERT model was utilized with an embedding length of 768, and a sub-word level tokenizer was employed to tokenize each sentence. Specifically, the bert-base-multilingual-cased model was chosen for this method. A Convolutional Neural Network (CNN) was developed to classify the sentences into their respective categories. The rectified linear unit (ReLU) activation function was used in the CNN design, which is a popular choice for nonlinear transformations. Using BERT embeddings and the CNN architecture, this method is intended to effectively identify texts as hopeful or non-hopeful. In the domain of hope speech detection, this combination of BERT embeddings with CNN architecture shows the possibility for accurate classification results.

(Gowda et al., 2022) proposed an innovative method for effectively categorising minority groups by combining resampling approaches with 1D-Convolutional Neural Networks (CNN) combined with Long Short-term Memory (LSTM). This model seeks to solve the widespread issue of imbalanced datasets, in which minority groups are frequently underrepresented. To address this issue, the researchers used resampling techniques to improve minority class representation in the training data, resulting in a more balanced distribution. The model design includes a 1D-CNN with LSTM layers, which allows the network to detect local and temporal relationships in the input data. This combination enables the model to learn patterns and features from resampled minority class data successfully.

2.1 Task and dataset description

The dataset used in this study was obtained from the shared task 'Hope Speech Detection for Equality, Diversity, and Inclusion - LT-EDI-RANLP 2023'(Chakravarthi, 2020). The organizers of the task shared a dataset compiled from YouTube comments. The objective of the task was to classify the provided YouTube comments into two categories: 'Hope Speech' and 'Non-Hope Speech.' The task encompassed four different languages: Bulgarian, English, Hindi, and Spanish. However, our proposed model focuses exclusively on the English language.

Label	Training set	Validation set
Hope_speech	1562	400
Non_hope_speech	16630	4148
Total	18192	4548

Table 1: Distribution of data

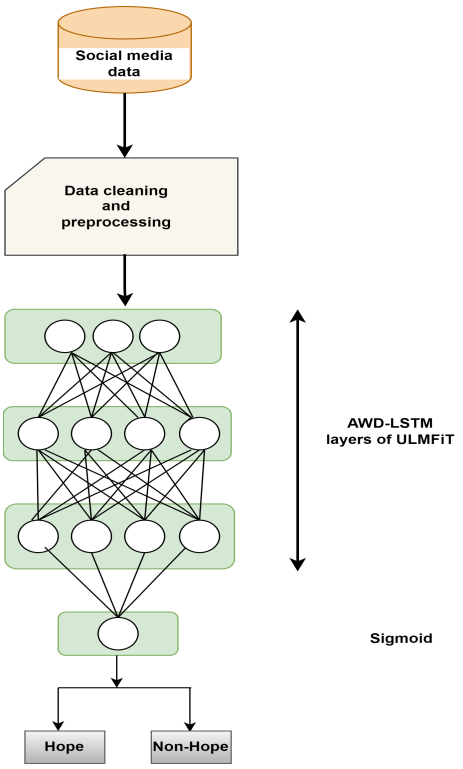

Figure 1: Proposed model

The English dataset, outlined in Table 1, comprises two primary columns: 'Text' and 'Labels.' The 'Text' column contains comments scraped from YouTube, while the 'Labels' column consists of two labels: 'Hope_Speech' and 'Non_Hope_Speech.' The dataset provided by the organizers exhibits an imbalance towards the 'Non_Hope_Speech' class.

3 Methodology

This section presents a comprehensive description of the proposed model, which is outlined in two subsections: Data Preprocessing and Model Description.

3.1 Data pre-treatment

The data provided by the organizers of the task is obtained from YouTube comments. In order to im-

Hyperparameters	Values
Batch size	64
Dropoutvrate	0.3
Learning rate	Best LR is selected by applying grid search between start_LR=slice(10e-7, 10e-5) end_ LR=slice(0.1, 10)
No of epochs	2 for first two layers, 3 for last layer

Table 2: Hyper-parameter tuning

prove the accuracy of the classification, it is crucial to remove the noise present in the data. YouTube comments often contain numerical data, punctuation, emoticons, hyperlinks, and URLs, which are not necessary for classification. Therefore, these elements are removed from the text. Stop words, which do not contribute significantly to the classification, are also eliminated. The entire text is converted to lowercase to avoid redundancy and ensure consistency. Word lemmatization is done to convert all forms of the word into their root word.

3.2 Model description

This section describes the model and training strategy submitted for the shared task "Hope Speech Detection for Equality, Diversity, and Inclusion- LT-EDI-RANLP 2023." Our proposed model utilizes the pre-trained language model ULMFiT from the fast.ai [1] library for classifying hope speech and non-hope speech. ULMFiT is a Long Short-Term Memory (LSTM) based model consisting of multiple layers of Average Weight Dropped (AWD) LSTM (Average Stochastic Gradient Descent weight dropped LSTM) stacked on top of each other. This model has shown promising results in various natural language processing tasks (Howard and Ruder, 2018; Azhan and Ahmad, 2021).

A layered differential training approach was employed to adapt the ULMFiT model for hope speech identification. This approach, which has been successful in computer vision problems, is applied to solve the hope speech problem. In layered differential training, the last three layers of the ULMFiT model are sequentially freezed and unfreezed. This process allows us to fine-tune the model's weights specifically for hope speech data. The last layer is trained slightly longer compared to the previous layers to prevent overfitting and retain valuable information. Finally, the features extracted from the ULMFiT model are passed through a sigmoid layer for classification. Table 2 provides the hyperparameters used in model training. These hyperparameters have been selected based on extensive experimentation and trial runs to optimize the model's performance. The proposed model demonstrates promising capabilities in identifying hope speech by leveraging the ULMFiT language model and employing layered differential training. The architecture of the proposed model is illustrated in Fig 1.

[1] https://www.fast.ai/

4 Result and Discussion

The organizers of the shared task "Hope Speech Detection for Equality, Diversity, and Inclusion-LT-EDI-RANLP 2023" evaluated the performance of the models using the macro-F1 score. Our team submitted a single run with the ULMFit model. Table 3 presents the top-performing models along with their macro-F1 scores. Notably, our proposed model achieved a macro-F1 score of 0.48, which is highlighted in bold in the table. This result showcases the effectiveness and competitiveness of our approach in accurately identifying hope speech in the English language.

Team Name	MF1	Rank
Tercet_English	0.50	1
ML_AI_IIITRanchi	0.50	1
Ranganayaki	0.49	2
VTUBGM	**0.48**	**3**
IIC_Team	0.47	4
MUCS_run2	0.44	5

Table 3: Top performing models

5 Conclusion and future enhancements

In the shared task "Hope Speech Detection for Equality, Diversity, and Inclusion" at LT-EDI-RANLP 2023, our team, VTUBGM, proposed a model built upon the ULMFit framework. We employed a layered differential approach to fine-tune the model specifically for classifying Hope and

Non_Hope speech. The proposed model achieved a macro-F1 score of 0.48 on the dataset provided by the organizers. As a result, our team secured 3rd rank in the competition. In this work the language considered is English, further, the proposed approach can be used to identify hope speech in other low-resource and code-mixed texts.

References

Mohammed Azhan and Mohammad Ahmad. 2021. Ladiff ulmfit: a layer differentiated training approach for ulmfit. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*, pages 54–61. Springer.

Fazlourrahman Balouchzahi, BK Aparna, and HL Shashirekha. 2021. Mucs@ dravidianlangtech-eacl2021: Cooli-code-mixing offensive language identification. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 323–329.

Shankar Biradar and Sunil Saumya. 2022. Iiitdwd@ tamilnlp-acl2022: Transformer-based approach to classify abusive content in dravidian code-mixed text. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 100–104.

Shankar Biradar, Sunil Saumya, and Arun Chauhan. 2021. Hate or non-hate: Translation based hate speech identification in code-mixed hinglish data set. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 2470–2475. IEEE.

Shankar Biradar, Sunil Saumya, and Arun Chauhan. 2022. Fighting hate speech from bilingual hinglish speaker's perspective, a transformer-and translation-based approach. *Social Network Analysis and Mining*, 12(1):87.

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. 2022. Hope speech detection in youtube comments. *Social Network Analysis and Mining*, 12(1):75.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, Subalalitha Cn, John McCrae, Miguel Ángel García, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José García-Díaz. 2022. Overview of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 378–388, Dublin, Ireland. Association for Computational Linguistics.

Bhargav Dave, Shripad Bhat, and Prasenjit Majumder. 2021. Irnlp_daiict@ lt-edi-eacl2021: hope speech detection in code mixed text using tf-idf char n-grams and muril. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 114–117.

Suman Dowlagar and Radhika Mamidi. 2021. Edione@ lt-edi-eacl2021: Pre-trained transformers with convolutional neural networks for hope speech detection. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 86–91.

Anusha Gowda, Fazlourrahman Balouchzahi, Hosahalli Shashirekha, and Grigori Sidorov. 2022. Mucic@ lt-edi-acl2022: Hope speech detection using data re-sampling and 1d conv-lstm. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 161–166.

Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.

Abhinav Kumar, Sunil Saumya, and Pradeep Roy. 2022. Soa_nlp@ lt-edi-acl2022: an ensemble model for hope speech detection from youtube comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 223–228.

Edison Marrese-Taylor, Jorge A Balazs, and Yutaka Matsuo. 2017. Mining fine-grained opinions on closed captions of youtube videos with an attention-rnn. *arXiv preprint arXiv:1708.02420*.

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.

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*.

Sunil Saumya Shankar Biradar and Arun Chauhan. 2021. mbert based model for identification of offensive content in south indian languages. In *Working Notes of FIRE 2021-Forum for Information Retrieval Evaluation (Online). CEUR*.