JudithJeyafreeda_StressIdent_LT-EDI@EACL2024: GPT for stress identification

Judith Jeyafreeda Andrew
Institut Imagine / Paris, France
PRAIRE / Paris, France
Université Paris Cité / Paris, France
judithjeyafreeda@gmail.com

Abstract

Stress detection from social media texts has proven to play an important role in mental health assessments. People tend to express their stress on social media more easily. Analysing and classifying these texts allows for improvements in development of recommender systems and automated mental health assessments. In this paper, a GPT model is used for classification of social media texts into two classes: stressed and not-stressed. The texts used for classification are in two Dravidian languages: Tamil and Telugu. The results, although not very good shows a promising direction of research to use GPT models for classification.

1 Introduction

Emotion classification is a well-known and highly-utilised application in the field of text analysis. Stress is a form of emotion. Thus, stress classification is a particular case of emotion classification. Stress is defined as a state of imbalance between one's internal demands and his/her ability to meet those demands [5], and is widely regarded as a medical problem (Rastogi et al., 2022). Stress can be a potentially life threatening problem. Stress can be identified from text, facial expressions, videos and audios of people. People express stress in different ways and forms. The field of identification of stress from text is an emerging research area. This is due to the fact that people have started to express themselves more comfortably on social media platforms with their friends and followers. In this work, we explore a method to classify social media text into stressed and non-stressed texts. The data from the task given in (Kayalvizhi Sampath and Rajkumar, 2023) is in the Dravidian languages of Tamil and Telugu.

2 Task Description

The task given in (Kayalvizhi Sampath and Rajkumar, 2023) is a binary classification of social media posts in the languages of Tamil and Telugu. The two labels are "stressed" and "not stressed". The data from (S et al., 2022) are given as separate sets for training, development and testing. Table 1 gives statistics on the number of text statements in each language provided for training, development and testing within the task.

Language	Train	Dev	Test
Tamil	5504	1378	1020
Telugu	5097	1239	1050

Table 1: Data statistics for the classification task

3 Related Work

There have been several studies in the areas of sentiment analysis and emotion classification. Several ML methods have been developed for this purpose. (Jadhav et al., 2019) presents a Bidirectional Long Short-Term Memory (BLSTM) with attention mechanism to classify psychological stress and categorize the tweets based on their hashtag content, which gives the best performance. (Arya and Mishra, 2021) gives a review of all machine learning methods developed within the health sector, their advantages, their limitations and areas for further research. The authors reviewed papers on mental stress detection using ML that used social networking sites, blogs, discussion forums, Questioner technique, clinical dataset, real-time data, Bio-signal technology (ECG, EEG), a wireless device, and suicidal tendency. (Nijhawan et al., 2022) shows the accuracy of each ML model trained specifically for mental illness.

Pre-trained language models like ELMo (Peters et al., 2018), OpenAI GPT (Radford et al., 2018), and BERT (Devlin et al., 2019) have moved Natural Language Processing (NLP) passing into a new era. This has allowed the pre-trained model to play the role of the base, and this can be fine-tuned to

Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 173–176
March 21, 2024

respond to the NLP task. (Asghar et al., 2017) presented a method enhanced by lexicons.

With respect to language of the text, there have been several works in the English language. However, works in Dravidian languages have recently increased. (Chakravarthi, 2022b; Kumaresan et al., 2022; Chakravarthi, 2022a) presents an improvement of word sense translation for under-resourced languages. (Andrew, 2020) uses several Machine Learning algorithms which have been adapted to the task of Multiclass Classification Sentiment Analysis. (Andrew, 2021, 2022) suggests several machine language approaches to classify texts from Code-mixed Dravidian Languages such as Tamil, Telugu, Kannada and Malayalam. (Andrew, 2023) uses a GPT model to perform a classification task on YouTube comments in different Dravidian Languages. Although, the results are not too high, this allows for further research to improve GPT models.

4 Proposed System

In this work, GPT2 is used for classification of social media texts into stressful and non-stressful comments. The model is fine-tuned on the training dataset for each language, thus creating a task specific and language specific model. For the languages Tamil and Telugu, the text to be classified is transliterated to their English language equivalents, this approach has been inspired from (Andrew, 2021) and (Andrew, 2022). The labels for the task are in English language, thus transliteration is not required for this.

Pre-processing: Similar to (Andrew, 2022), a few steps of pre-processing is performed to get the accurate representation of the text.

This involves the following:

- Texts from the Tamil and Telugu languages are transliterated to their English equivalents. Transliteration refers to the method of mapping from one system of writing to another based on phonetic similarity. This transliteration is performed using the *polyglot.transliteration* package in Python.

- The emojis are substituted with the words of the emotion they represent like happy, sad, excited etc.

- The tokenizer from the pretrained GPT2 model is used for tokenization of the transformed text.

GPT models: Generative Pre-trained Transformers (GPT) models are general-purpose language models that can perform a broad range of tasks from creating original content to write code, summarizing text, and extracting data from documents (GPT). Generative Pre-trained Transformers (GPT) are a family of neural network models that uses the transformer architecture. These use a self-attention mechanism allowing to focus on different parts of the input text during the various stages on processing. The value of these models lies in their speed and the scale at which they can operate. In particular, GPT-2 model has 1.5 billion parameters and has been trained on 8 million web pages in a self-supervised fashion. (Radford et al., 2019) provides a detailed description of the model. The model uses internally a mask-mechanism to make sure the predictions for the token i only uses the inputs from 1 to i but not the future tokens. This allows the model to learn the inner representation of the language, which can then be used to extract features for downstream tasks.

GPT models for classification: Although most use cases for a GPT involve text generation operations, recent research has shown that these models can also be fine-tuned for downstream tasks like classification. (Andrew, 2023) has used the GPT2 for classification of Homophobic and Transphobic comments from social media.

As in (Andrew, 2023), the python packages that allow the use of GPT models as in Hugging Face models are used along with other tools like NLTK and TextBlob to allow cleaning of text.

5 Results and Evaluation

The performance of the classification system is measured in terms of macro averaged Precision, macro averaged Recall and macro averaged F1-Score across all the classes (for both sub tasks). The Scikit-learn [1] package is used for this purpose, similar to (Andrew, 2023). The Macro and Weighted results for the task are shown in Tables 2 and 3 respectively. Overall, the results for the Tamil language is better than that for the Telugu language.

Language	M.Precision	M.Recall	M.F1
Tamil	0.459	0.498	0.273
Telugu	0.255	0.247	0.251

Table 2: Results of the task of classifying social media text to stressed and non-stressed. (M. stands for "Macro")

Language	W.Precision	W.Recall	W.F1
Tamil	0.485	0.364	0.202
Telugu	0.293	0.281	0.287

Table 3: Results of the task of classifying social media text to stressed and non-stressed. (W. stands for "Weighted")

6 Conclusion

From Tables 2 and 3, the results on the whole are not too high. This is an interesting result as the number of training data in both languages were similar with the exact same classes for classification. (Andrew, 2021) and (Andrew, 2022) suggest that using IPA substitutes for Dravidian languages works well for certain machine learning approaches, however it might not be the best representation for a transformer based model. Similarly, transliteration might not be the way to go with transformer models as well. Choosing other forms of embedding Dravidian texts could help improve the results.

References

AWS what is gpt? https://aws.amazon.com/what-is/gpt/. Accessed: 2023-12-11.

Judith Andrew. 2020. Judithjeyafreeda@ dravidian-codemix-fire2020:: Sentiment analysis of youtube comments for dravidian languages. In *Forum for Information Retrieval Evaluation*.

Judith Jeyafreeda Andrew. 2021. JudithJeyafreedaAndrew@DravidianLangTech-EACL2021:offensive language detection for Dravidian code-mixed YouTube comments. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 169–174, Kyiv. Association for Computational Linguistics.

Judith Jeyafreeda Andrew. 2022. JudithJeyafreedaAndrew@TamilNLP-ACL2022:CNN for emotion analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 58–63, Dublin, Ireland. Association for Computational Linguistics.

Judith Jeyafreeda Andrew. 2023. JudithJeyafreeda@LT-EDI-2023: Using GPT model for recognition of homophobia/transphobia detection from social media. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 78–82, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Vishakha Arya and Amit Kumar Mishra. 2021. Machine learning approaches to mental stress detection: a review. *Annals of Optimization Theory and Practice*, 4(2):55–67.

Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Maria Qasim, and Imran Ali Khan. 2017. Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. *PloS one*, 12(2):e0171649.

Bharathi Raja Chakravarthi. 2022a. Hope speech detection in Youtube comments. *Social Network Analysis and Mining*, 12(1):75.

Bharathi Raja Chakravarthi. 2022b. Multilingual hope speech detection in English and Dravidian languages. *International Journal of Data Science and Analytics*, 14(4):389–406.

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.

Sachin Jadhav, Apoorva Machale, Pooja Mharnur, Pratik Munot, and Shruti Math. 2019. Text based stress detection techniques analysis using social media. In *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, pages 1–5.

Bharathi Raja Chakravathi Jerin mahiba C Ramya Priya S Bharani Kasinathan Kishore Kumar Ponnusamy Kayalvizhi Sampath, Thenmozhi Durairaj and Charumathi Rajkumar. 2023. Overview of the shared task on stress identification in dravidian languages. In *proceedings of the Third Workshop on Speech and Language Technology for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natuarl Language Processing.

Prasanna Kumar Kumaresan, Rahul Ponnusamy, Elizabeth Sherly, Sangeetha Sivanesan, and Bharathi Raja Chakravarthi. 2022. Transformer based hope speech comment classification in code-mixed text. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 120–137. Springer.

[1] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html

Tanya Nijhawan, Girija Attigeri, and T. Ananthakrishna. 2022. Stress detection using natural language processing and machine learning over social interactions. *Journal of Big Data*, 9(1). Funding Information: Nil. Publisher Copyright: © 2022, The Author(s).

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Aryan Rastogi, Qiang Liu 0004, and Erik Cambria. 2022. Stress detection from social media articles: New dataset benchmark and analytical study. In *International Joint Conference on Neural Networks, IJCNN 2022, Padua, Italy, July 18-23, 2022*, pages 1–8. IEEE.

Kayalvizhi S, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 2022. Findings of the shared task on detecting signs of depression from social media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 331–338, Dublin, Ireland. Association for Computational Linguistics.