Visual Impact on Sentiment: Climate Change Tweets Analysis

Pranav Tyagi

August 6th, 2024

1 Introduction

This thesis aims to predict the sentiment, encompassing emotions such as joy, anger, hope, etc., in replies to tweets containing images related to climate change by leveraging state-of-the-art (SOTA) image models. By comparing various approaches, such as zero-shot prediction, multi-modal techniques that integrate textual and visual data, and fine-tuning pre-trained image models, the study aims to identify the most effective strategies for sentiment analysis. Furthermore, the research will propose potential improvements to enhance the accuracy and applicability of these approaches, thereby advancing the understanding of visual impact on public sentiment and contributing to more effective communication strategies.

2 Background

Climate change is one of the most pressing issues of our time, and social media platforms like X (formerly Twitter) are crucial for spreading awareness and shaping public opinion. Tweets often combine text and images to convey messages, and understanding the sentiment these elements evoke is essential for gauging public perception.

The dataset for this research comprises tweets from the "Towards Understanding Climate Change Perceptions: A Social Media Dataset"[5], which includes tweets that contain images, accompanying text, and user reactions (replies).

Sentiment analysis is a well-established field in natural language processing (NLP). Traditional methods include lexicon-based approaches and machine-learning models. Transformer-based models such as BERT[1] have demonstrated high efficacy in encoding textual data, thereby enhancing classification performance. Recently, variants of models like BART[4] e.g. "bart-large-mnli" based on the method proposed by Yin et al.[9] have exhibited superior zero-shot classification capabilities, enabling more comprehensive sentiment and emotion analysis.

On the other hand, image sentiment analysis is relatively less explored. Large CNN-based models such as ResNet[3] and EfficientNet[7] are typically employed to extract image embeddings, which are then used for classification. Recent advancements in computer vision introduced transformer-based models like ViT[2], which aims to replicate the success of transformer in text-based tasks to images. CLIP[6] is a neural network built using ViT-like models in conjunction with a causal model and is trained on image and text pairs. It can be instructed in natural language to perform a task without optimizing for it, which makes it possible to use it to predict emotions without specific training.

Emoset[8] is a large-scale visual emotion dataset annotated with rich attributes containing emotion labels like anger, disgust, awe, etc. This dataset provides valuable data for model training or evaluation or both.

3 Goals and Work Plan

3.1 Goals

- **Sentiment Prediction from Text and Image**: Utilize SOTA text models to predict sentiment from tweet text and replies. Utilize SOTA Image models and multi-modal approaches to predict sentiment from images.
- **Comparison of Approaches**: Compare the performance of zero-shot and multi-modal approaches and identify areas for improvement.
- **Propose Enhancements**: Based on the experimental results, suggest modifications to enhance image-based sentiment prediction approaches.

Additionally, if time permits and the text-based sentiment predictions on tweet replies prove accurate, apply transfer learning by using these predictions as training data to fine-tune the image models.

3.2 Work Plan (Figure 1)

- **Literature Review**: Review literature to identify suitable text and image sentiment analysis models.
- **Sentiment Prediction from Tweet Replies**: Preprocess replies and apply text models. Annotate a subset for evaluation.
- **Sentiment Prediction from Tweet Text**: Preprocess tweet text and apply text models.
- **Sentiment Prediction from Images**: Preprocess images and apply image models. Compare with text predictions.
- **Multi-Modal Sentiment Analysis**: Explore and evaluate multi-modal approaches which integrate visual and textual data.
- **Performance Comparison and Failure Analysis**: Compare approaches and conduct failure analysis.
- **Propose Possible Enhancements**: Suggest possible enhancements to the approaches tested during experimentation.
- **Optional: Fine-Tune Image Models Using Text Predictions**: Fine-tune image models using predictions from text models as labels.

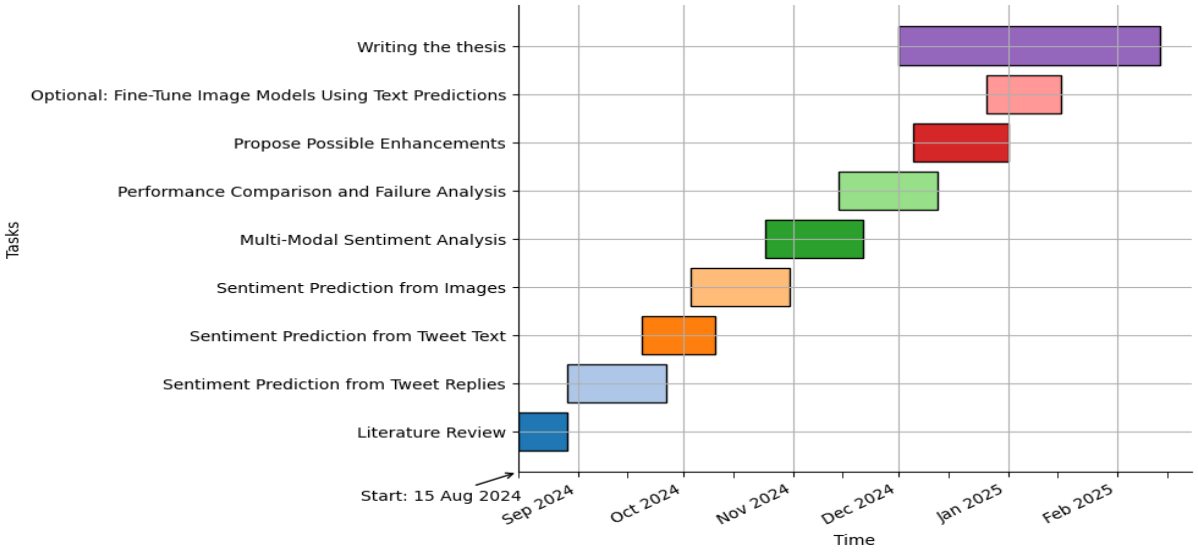

Figure 1: Gantt Chart for Thesis Timeline

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[4] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, 2019.

[5] Katharina Prasse, Steffen Jung, Isaac B Bravo, Stefanie Walter, and Margret Keuper. Towards understanding climate change perceptions: A social media dataset. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023.

[6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.

[7] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.

[8] Jingyuan Yang, Qirui Huang, Tingting Ding, Dani Lischinski, Danny Cohen-Or, and Hui Huang. Emoset: A large-scale visual emotion dataset with rich attributes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20383–20394, October 2023.

[9] Wenpeng Yin, Jamaal Hay, and Dan Roth. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach, 2019.