CSCI-544 Project Proposal
General Speech Style Translator

Author
Jerry Dong / cejingxi@usc.edu
Jeremy Yang / zhaoxuya@usc.edu
Kexin Zheng / kexinzhe@usc.edu
Yudi Qi / yudiqi@usc.edu
Ziyi Li / zli76976@usc.edu

Abstract

This document contains our initial project proposal in the topic of analyzing and learning speaking and speech styles from political figures. In this part of design, this document focuses on related work, possible data sets and training models, technical approaches, and possible technical challenges we may need to address.

1 Project Domain and Goals

Our project is about the personal style of a paragraph of natural language. First goal is that We are trying to generate a set of paragraphs with a specified type of language. Most of the NLP generative models generate plain text from data without any redundancies, though for humans the language contains a fair number of redundancies to represent different mental states or personal experiences. As part of the natural language, these redundancies should also be able to be generated by a model to mimic human speech with specific mental state and personal experiences. As a result, the paragraph generated by the new model we proposed should look alive as a human writer or speaker.

Moreover, the second goal of this project is to translate a set of paragraphs with their own style to paragraphs with specified style instead of changing their own meaning. Different with the first goal, in order to achieve this goal, first we need to extract information without style with the feature identification method we used in feature extraction, then generate a set of styled paragraphs using the plain text.

2 Related Work

Sequence-to-sequence (Seq2seq) model has become an important part of natural language processing, it is successfully used in text-generation and sequence-labeling domains.

- For the voice aspect, Ribeiro et al. [5] managed to convert styles from several different speakers with data augmentation and successfully solved the problem of cross-speaker style transfer. In addition, Weiss et al. [7] utilized the deep neural network for translating texts from one language to another language without transcribing the speech into text.
- For the image aspect, there are also several studies on style transfer. For example, Kashyap et al. [2] applied the Seq2seq model to describe a painting in the style of Shakespearean prose. They used an English poem as an intermediate and then translated it into Shakespearean style.
- For the text aspect, Priya Dwivedi generated Donald Trump Speeches using the GPT-2 model and had a good performance [1]. It looks similar to our project; however, it is for simply generating random speeches in Donald Trump style. While our project is to transfer existing speeches to Donald Trump style. At present, there is still limited research in translating speeches from one language style to another.

3 Datasets

In this part of design, we will be using existing datasets from multiple open sources including

Kaggle, GitHub, etc. Given the goal of this project is to maintain the author's personal style, we may not use existing standard preprocessing and cleaning steps on the datasets as they may also clean out one's personal style in a text document. As part of this design, we will skip preprocessing and cleaning steps and based on the performance on final training, we will be likely to develop our own procedure to preprocess and clean the data. Below are some possible datasets/models we may use in this project:

- Trump Twitter Archive[6]: This dataset contains all of Trump's tweets since 2009. It was copied in its entirety from the website The Trump Archive who did all the work in periodically scraping Trump's Twitter account until his suspension in 2021. Given that Trump's tweets have a strong representation of his personal speech style, we will use this dataset to learn Trump's use of language and his personal speaking style aside from official speeches.

- Donald Trump Rally Speeches[3]: This dataset contains the full speeches that Donald Trump gave at 35 of his rallies. Ideal for natural language processing and sentiment analysis. This dataset will help us learn Trump's use of language in a more official way compared to his personal tweets and combined with tweets, we will try to find similarities and differences of his personal style based on two datasets to better understand his tones.

- GPT-2 Model[4]: This will be a related training model that we may use in this project. The GPT-2 model was released by Open AI last year. It is a language model that has been trained on 40GB of internet text. This model can be fine-tuned on any text sample to tweak its generation style based on the input text.

4 Technical Challenge

Since currently most natural language processing steps contain word lemmatization, and stemming to reduce inflectional forms, one of the technical problems we will solve is that we cannot perform those word cleaning steps as those described in class on our datasets, since different people may have different writing or speaking styles - they may use their accustomed tense and word class, i.e. some people like to use active voice. After performing word cleaning, the expressions our model learns will change and the generated results will not be as accurate as we expected.

Another technical challenge might be how to evaluate the results we generate. The similarity between the results we generate, and real human beings cannot be evaluated by programs. Different reviewers might also have different opinions on how similar the results we generate to a real human being that we want to mimic. In this part of design, we are still exploring possible solutions for testing and evaluation and one possible solution addressing similar issue is to utilize BLEU (Bilingual Evaluation Understudy) score which is a metric for automatically evaluating machine-translated text. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high-quality reference translations such that we will be able to evaluation the performance of speech translation.

5 Division of Labor

In this project, we will work on two major tasks i.e., research task and implementation task, and each team member will be able to participate in both tasks. In research task, all the team members will be work together and learn different models and algorithms. In implementation task, Kexin, Jerry and Ziyi will be responsible for training and adjusting models and Jeremy, Yudi will be responsible for testing and evaluating our models.

References

[1] P. Dwivedi. Election special: Train a gpt-2 to generate donald trump speeches, 2020. URL https://towardsdatascience.com/election-special-train-a-gpt-2-to-generate-donald-trump-speeches-b66fc3aa92b9.

[2] P. Kashyap, S. Phatale, and I. Drori. Prose for a painting. 2019. doi: 10.48550/ARXIV.1910.03634. URL https://arxiv.org/abs/1910.03634.

[3] C. LILLELUND. Donald trump rally speeches. URL https://www.kaggle.com/datasets/christianlillelund/donald-trumps-rallies.

[4] OpenAI. Better language models and their implications, 2019. URL https://openai.com/blog/better-language-models/.

[5] M. S. Ribeiro, J. Roth, G. Comini, G. Huybrechts, A. Gabrys, and J. Lorenzo-Trueba. Cross-speaker style transfer for text-to-speech using data augmentation. 2022. doi: 10.48550/ARXIV.2202.05083. URL https://arxiv.org/abs/2202.05083.

[6] H. O. TAILS. Trump twitter archive. URL https://www.kaggle.com/datasets/headsortails/trump-twitter-archive.

[7] R. J. Weiss, J. Chorowski, N. Jaitly, Y. Wu, and Z. Chen. Sequence-to-sequence models can directly translate foreign speech. 2017. doi: 10.48550/ARXIV.1703.08581. URL https://arxiv.org/abs/1703.08581.