

Targeted Writing Style Text Generation ~~Transfer~~

Interim Report for Final Project

By Nathan Harris

CSPB 3832 Natural Language Processing - Summer 2025

1. Project Summary

The problem my project focuses on relates to the homogenization of text online as more people use LLMs to help craft blog posts, social media posts and internet content. For my project, I experiment with generating text in an author's writing style based on writing samples. I plan to use well established authors as well as unknown authors. This NLP problem is called "Text Style Transfer" but I have also seen it called "Authorship Style Transfer." Ideally, writers have a tool that helps them keep track of their own distinct writing style and this project is a stepping stone towards that goal.

2. Data

For experimental purposes I am currently using Moby Dick by Herman Melville via Project Gutenberg. The data is 1,191,789 characters, and 208,458 words.

For preprocessing I have removed the Project Gutenberg Header and Footer. I have segmented the text in to chunks based on paragraphs. I could have chunked it by sentence but I realize that sentence structure is a vital part of writing style.

I take these chunks and transform them using Google Flan T5 large in to neutral style chunks with the prompt "Paraphrase this text for a 5th grade student." (I have tried a few dozen different prompts but this one seems to work best so far).

The final dataset is in hugging face format with input fields as the neutralized text and target fields as the stylized texts as well as the author's name, Herman Melville, attributed to each pairing.

3. Approach

At first I tried to fine tune a model using Moby Dick and it worked well for text generation. However, I soon realized that I needed to be transferring the text into a neutral format and then fine tuning with the stylized/neutral dataset.

I am using Google Colab and Hugging Face Transformers. I took inspiration from the Machine Translation Recipe in Hugging Face tutorials.

My plan is to have 3 Google Colab Notebooks where progress is saved to google drive at the end of each notebook. This way it's convenient to pick up where I left off with the saved files while working on the project.

Notebook 1 "Neutralizer": Input Author's corpus, process and generate neutral text, save as a hugging face dataset in to google drive.

Notebook 2 "Stylizer": Input the dataset, fine tune a pretrained model with the dataset and save the model to use later. Some light testing to make sure it works.

Notebook 3 "Evaluation": Input the model, run and document tests and evaluation metrics.

Right now I am using Hugging Face Google Flan T5 Large for both generating neutral text and converting to stylized texts.

4. Progress So Far

As I mentioned before I tried fine tuning on just the author's text but realized that was only for stylized text *generation* and not style *transfer*.

I have the first two notebooks completed and I have worked with a very small amount of the text just to make sure I wasn't getting any errors.

At this moment the neutralized texts are not great quality and unreliable. The stylizer is also just outputting identical text. I am happy that I have some completed code but I need to work on getting a better dataset with better neutralizing techniques as well as spend more time fine tuning the model to get better *results*.

5. Next Steps

I'd like to generate better neutral texts and train on larger amounts of data to see if my current pipeline is working. I need to also create the third notebook for evaluating the text.

I hope to have the colab notebooks and my power point slides done by the end of the week so that I can record my video presentation and submit on Monday.

6. Questions or Concerns

Any tips on automating the process of generating neutralized chunks of text from stylized chunks of text?