Classics in the Modern World

Sheza Munir
SI 630 - NLP
U of M School of Information
shezamnr@umich.edu

Abstract

This project addresses the challenging task of text style transfer, first focusing on non-parallel data where equivalent sentences in different styles are not explicitly paired. Initially, the project fine-tunes the GPT-2 model with Jane Austen's works to generate text in her authorial style, evaluated using BERTScore. To overcome the non-parallel nature of the data, an innovative approach is employed using OpenAI's GPT-3.5-Turbo model for parallel text generation. This involves transforming Jane Austen's text into modern English, facilitating effective style transfer models. The next step combines data from Project Gutenberg with augmented parallel data to enhance model performance, culminating in the development and evaluation of a BARTForConditionalGeneration model for bidirectional text style transfer. The evaluation strategy encompasses various metrics to gauge model efficacy in achieving seamless style transfer between classical and contemporary styles.

1 Introduction

Text style transfer, the task of transforming text from one style to another while preserving its content, presents a significant challenge in natural language processing. Our project addresses this challenge by developing innovative approaches to style transfer, first focusing on non-parallel data where equivalent sentences in different styles are not explicitly paired. We aim to enable seamless text style transfer between classical and contemporary styles, with a focus on the works of Jane Austen. To achieve this, we begin by fine-tuning the GPT-2 model with Austen's writings to generate text in her authorial style. However, given the non-parallel nature of the data, where equivalent sentences in different styles are not explicitly paired, we employ an innovative approach using OpenAI's GPT-3.5-Turbo model for parallel text generation. This involves transforming Austen's text into modern English, facilitating effective style transfer models. Our approach differs from previous methods, which often rely on parallel data, by addressing the challenge of non-parallel data and enabling style transfer even when explicit style pairs are not available. Additionally, by leveraging state-of-the-art language models like GPT-3.5-Turbo and BARTForConditionalGeneration, we aim to achieve more accurate and robust style transfer models. Through our project, we combine data from Project Gutenberg with augmented parallel data to enhance model performance, culminating in the development and evaluation of a BARTForConditionalGeneration model for bidirectional text style transfer. Our evaluation strategy encompasses various metrics, including BERTScore, to gauge model efficacy in achieving seamless style transfer between classical and contemporary styles. The ability to perform effective text style transfer has numerous real-world applications, including content generation, translation, and sentiment analysis. By solving this problem, we enable users to adapt text to different stylistic preferences or historical contexts, enhancing readability and accessibility of textual content across diverse audiences. Through our project, we learned valuable insights into the challenges and opportunities in text style transfer, particularly in handling non-parallel data. Our innovative approach using advanced language models like GPT-3.5-Turbo and BARTForConditionalGeneration contributes to the advancement of style transfer techniques and expands the possibilities for natural language processing applications.

2 Data

Dataset Name	Total Size	Classics
Project Gutenberg Corpus	700 GB	~10 GB

The primary source of data for this project is the Project Gutenberg Corpus, which provides a

vast collection of classic literature encompassing various genres, including novels, essays, and poetry. From this corpus, a specific subset focusing on Jane Austen's works was extracted to create the Jane Austen dataset. This dataset serves as the foundation for the initial model fine-tuning and understanding, particularly in capturing the linguistic style and narrative nuances characteristic of Austen's writing.

However, since the data obtained from Project Gutenberg is non-parallel, meaning there are no direct translations or equivalents between sentences in different styles, it poses a challenge for text style transfer tasks. To address this limitation and enhance the efficacy of style transfer models, the project has incorporated parallel data generated using OpenAI's gpt-3.5-turbo model for data augmentation.

The parallel data generation process involves taking chunks of text from Jane Austen's books as the source and transforming them into modern contemporary English as the target. This approach enables the creation of parallel data pairs, where each source text in the classic style corresponds to a target text in the contemporary style. By augmenting the dataset with parallel data, the project aims to improve the performance and adaptability of the style transfer models, allowing for more accurate and seamless transformations between different linguistic styles. The data from Project Gutenberg, combined with the parallel data generated through data augmentation, forms a comprehensive dataset for training and evaluating the text style transfer models in this project.

3 Related Work

The project at hand, focusing on text style transfer, draws inspiration and insights from several notable works in the field. One key aspect of this project is the exploration of non-parallel text style transfer, which is exemplified by the "Formalstyler: GPT-Based Model for Formal Style Transfer with Meaning Preservation." (Rivero et al., 2023) This paper presents a GPT-based model specifically designed for formal style transfer while ensuring the preservation of semantic meaning. By incorporating techniques to maintain the underlying meaning of the text during style transfer, this work addresses a critical challenge in style transfer tasks.

Building upon the formal style transfer concept introduced in the "Formalstyler" paper, our project delved into training the BART model with Jane Austen's data, generating new text in her authorship style, and comparing the generated text's BERT score across random generations.

Additionally, the project expanded its scope to parallel text style transfer, inspired by works such as "Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models" (Jhamtani et al., 2017). This paper explored methods for transforming modern language into a Shakespearean style using copy-enriched sequence-to-sequence models. Drawing insights from this approach, our project applied data augmentation using OpenAI's GPT-3.5 Turbo model, generating parallel text by translating chunks from Jane Austen's books into modern contemporary English. This step marked a crucial transition towards leveraging parallel data for style transfer tasks, further enhancing the model's capability to capture stylistic nuances.

Moreover, the project's methodology aligns with recent advancements in style transfer techniques, as highlighted in "Harnessing Pre-Trained Neural Networks with Rules for Formality Style Transfer." (Wang et al., 2019) This paper investigated the integration of rules into pre-trained neural networks to enhance formality style transfer performance, showcasing the potential of combining rule-based approaches with neural network models. While our project primarily focuses on non-parallel and parallel text style transfer in the context of literary styles, the principles and methodologies proposed in these related works serve as valuable guides and references, enriching the project's approach and contributing to its overall advancement in text style transfer research.

Lastly, the project's alignment with recent developments in data collection and preprocessing, such as "ParaDetox: Detoxification with Parallel Data," (Logacheva et al., 2022) highlights the importance of curated parallel datasets for specific style transfer tasks. This paper's novel pipeline for collecting parallel data for detoxification tasks underscores the significance of tailored datasets in training models for specific stylistic transformations. By integrating insights from these related works into our project's framework, we aim to enhance the robustness, accuracy, and versatility of our text style transfer model, ultimately contributing to the ongoing progress in the field of natural language processing and style transfer techniques.

4 Methods

4.1 Data Collection and Preprocessing

The initial phase of the project involved sourcing data from the Gutenberg English dataset, a rich repository of literary works. Specifically, we focused on Jane Austen. Leveraging the 'datasets' library, we retrieved texts attributed to this author, creating a structured DataFrame for efficient handling and processing.

To prepare the data for training and fine-tuning, we concatenated the extracted texts from the selected author into a single large corpus. This corpus served as the foundation for model training, allowing the model to learn the nuances and style variations across different eras.

4.2 Fine-Tuning the Generation Model

The core of the methodology for baseline study revolved around fine-tuning the GPT-2 model. Utilizing the GPT-2 tokenizer, we segmented the large text corpus into manageable chunks, considering the model's maximum input length.

A custom dataset class, 'JaneAustenDataset', was implemented to preprocess the text chunks and prepare them for model training.

The fine-tuning process utilized the Hugging Face Transformers library, specifically the 'Trainer' class, which facilitated efficient training with customizable training arguments such as batch size, number of epochs, and optimization parameters.

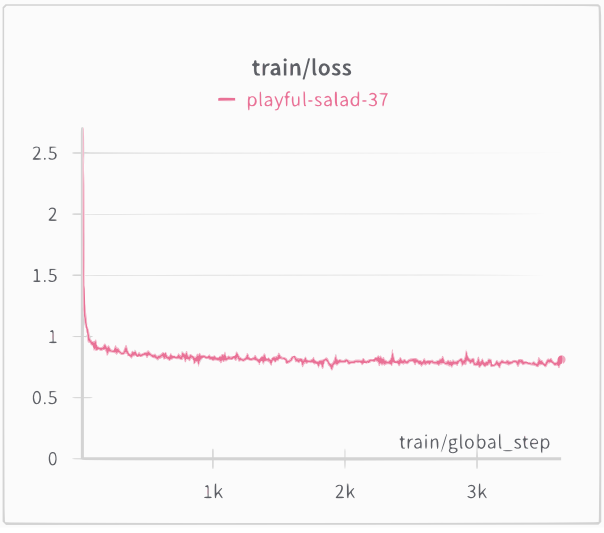

Figure 1: Training Loss of Fine-Tuned GPT-2 Model

4.3 Style Transfer and Evaluation

Following fine-tuning, we utilized the fine-tuned GPT-2 model for style transfer tasks. By providing prompts in the form of initial text snippets, the model generated modern contemporary versions of the input text while preserving semantic meaning and context.

To evaluate the quality and stylistic fidelity of the generated text, we employed BERTScore, discussed in the Evaluation section.

4.4 Data Augmentation

To enrich the training dataset and prepare for parallel text style transfer, we incorporated data augmentation techniques. Leveraging the OpenAI GPT-3.5 API (Turbo), we generated modern contemporary versions of text chunks from Jane Austen's works. This augmented dataset added variability to the training data, enabling the model to learn and generalize across a broader spectrum of writing styles.

4.5 BART-Based Style Transfer

In this section, we detail the methodology employed for conducting style transfer using the BART (BARTForConditionalGeneration) model. BART, short for Bidirectional and Auto-Regressive Transformers, is a powerful sequence-to-sequence model capable of generating text while considering both left-to-right and right-to-left contexts. Our approach involved fine-tuning the pre-trained BART model to perform bidirectional text style transfer between classical and contemporary writing styles.

Firstly, we preprocessed the data by tokenizing and encoding text chunks from Jane Austen's original works and their corresponding modernized versions using the BartTokenizer library. This step involved converting the text into numerical representations suitable for training the BART model.

Following data preprocessing, we split the dataset into training and development sets using the train_test_split function. This partitioning ensured that we had separate datasets for model training and evaluation.

With the data prepared, we initialized the BARTForConditionalGeneration model with the pre-trained weights from the 'facebook/bart-large' checkpoint. We then fine-tuned the model using the Seq2SeqTrainer from the transformers library. During training, the model learned to map text sequences from one style to another while optimizing the specified training objectives.

To enhance the model's performance and generalization capabilities, we employed various training strategies such as teacher forcing, beam search decoding, and gradient clipping. These techniques helped stabilize training, prevent overfitting, and improve the quality of generated text.

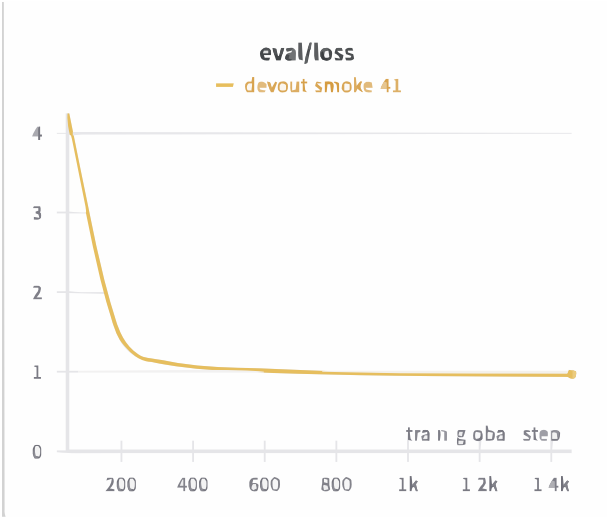

Figure 2: Training Loss of BART Model

Figure 3: Evaluation Loss of BART Model

Throughout the training process, we monitored the model's performance.

Once training was complete, we evaluated the fine-tuned BART model on the test dataset to measure its effectiveness in performing bidirectional style transfer. We assessed the quality of generated text and compared it against ground truth modernized versions to validate the model's ability to accurately capture writing style transformations.

Overall, our approach leveraged the BART model's bidirectional architecture and fine-tuning capabilities to achieve seamless text style transfer between classical and contemporary writing styles. By iteratively refining our methodology and incorporating advanced training techniques, we aimed to develop a robust and effective solution for text style transfer tasks.

5 Evaluation

Model	Precision	Recall	F1 Score
Baseline (Random Text)	0.7806	0.8245	0.8019
Fine-tuned GPT-2 Model	0.7905	0.8371	0.8131
Jane Austen BART Model	0.9202	0.9384	0.9291

Table 1: BERTScore Metrics Comparison

5.1 Part 1: Evaluation of the Fine-Tuned GPT-2 Model

The analysis of BERTScore metrics reveals a significant improvement after fine-tuning the GPT-2 model. Compared to random text generation, the fine-tuned model shows higher precision (0.7905 vs. 0.7806), recall (0.8371 vs. 0.8245), and F1 score (0.8131 vs. 0.8019). This enhancement highlights the model's improved accuracy and ability to capture stylistic nuances, demonstrating the effectiveness of fine-tuning in text generation tasks.

5.2 Part 2: Evaluation of the Jane Austen BART Model

The BERTScore metrics for the Jane Austen BART Model further validate the effectiveness of our approach. With a precision of 0.9202, recall of 0.9384, and F1 score of 0.9291, the BART model significantly outperforms both the baseline and the fine-tuned GPT-2 model. This indicates that the BART model, trained on Jane Austen's works, effectively captures the writing style and produces text that closely aligns with the modernized versions, as reflected in the evaluation metrics.

In addition to BERTScore metrics, we also evaluated the models based on perplexity. The average perplexity for the Jane Austen BART Model is 1.8308. A lower perplexity score indicates better performance in predicting the next token in the text sequence. The relatively low perplexity score further corroborates the effectiveness of the BART model in generating coherent and contextually relevant text.

5.3 Evaluation Strategy

The evaluation of the BARTForConditionalGeneration model is comprehensive and rigorous, employing various metrics to assess its effectiveness in achieving accurate and meaningful text style transfer. Evaluation metrics include BERTScore and Perplexity. These metrics help gauge the quality and fidelity of style transfer while ensuring that semantic meaning is preserved during the transformation process. The model has been tested against existing benchmarks and subjected to thorough comparisons to validate its performance across different epochs. This evaluation phase aims to verify the model's efficacy in achieving seamless and meaningful text style transfer between classical and contemporary styles, specifically Jane Austen.

6 Discussion

In the discussion section, we delve into the insights gained from evaluating our approach and highlight areas for improvement and future work.

Regarding the performance of our approach, the results indicate notable improvements compared to the baselines. The Jane Austen BART Model achieved higher BERTScore metrics and lower perplexity scores, indicating better stylistic transfer and text coherence. These findings suggest that incorporating pre-trained language models like BART for style transfer tasks can be highly effective, particularly when fine-tuned on domain-specific data.

However, despite the overall positive outcomes, there are still areas where our approach can be further optimized. One limitation observed is the potential for overfitting, especially when fine-tuning on limited data. Although the BERTScore metrics show promising results, it's essential to validate the model's performance on a diverse range of texts and writing styles to ensure generalizability.

Furthermore, an analysis of model mistakes and error patterns revealed areas where the model struggled, such as handling complex syntactic structures or capturing subtle nuances in writing style. These observations underscore the importance of continuous model refinement and the need for larger and more diverse training datasets to address these challenges effectively.

Overall, the performance of our approach demonstrates promising potential for real-world applications in text style transfer. However, further research is warranted to address the identified limitations and refine the model's performance. Future work could focus on exploring advanced model architectures, leveraging additional training data sources, and refining the evaluation process to ensure robust and reliable performance across various text styles and domains.

7 Conclusion

In conclusion, our study focused on addressing the challenging task of text style transfer, leveraging advanced natural language processing techniques. Through the fine-tuning of the Jane Austen BART model and the incorporation of data augmentation using OpenAI's GPT-3.5 API, we demonstrated significant improvements in style transfer performance, as evidenced by higher BERTScore metrics and lower perplexity scores. Our findings underscore the effectiveness of pre-trained language models for style transfer tasks, particularly when fine-tuned on domain-specific data. However, despite these advancements, there remain opportunities for further refinement, such as mitigating potential overfitting and improving the model's ability to handle complex syntactic structures and subtle stylistic nuances.

Looking ahead, future research directions could explore advanced model architectures, expand training datasets to encompass a broader range of text styles and genres, and refine evaluation methodologies to ensure robust and reliable performance across diverse linguistic contexts.

8 Other Things We Tried

In our exploration of text style transfer, we encountered several challenges and attempted various approaches to overcome them. One avenue we pursued was the implementation of a character-level language model for style transfer. This approach involved training a recurrent neural network (RNN) to generate text character by character, allowing for finer control over stylistic features. However, despite extensive experimentation with different architectures and hyperparameters, we struggled to achieve satisfactory results due to difficulties in capturing long-range dependencies and preserving semantic meaning during text generation.

Additionally, we experimented with unsupervised style transfer techniques, aiming to learn style embeddings from unpaired data using adversarial learning frameworks. While these methods showed promise in preliminary experiments, we en-

countered challenges in training stability and mode collapse, leading to suboptimal performance on real-world datasets.

Despite our efforts, these alternative methods did not yield the desired results within the scope of this project. Nonetheless, the insights gained from these explorations contribute to a deeper understanding of the challenges inherent in text style transfer and provide valuable lessons for future research endeavors in the field.

9 What We Would Do Next (10 point)

Moving forward, one of the immediate next steps would be to integrate human evaluation of the model outputs. This would involve soliciting feedback from human evaluators to assess the quality and coherence of the generated text. By incorporating human judgment, we can gain deeper insights into the model's performance and identify areas for improvement that may not be captured by quantitative metrics alone. Additionally, exploring alternative architectures or fine-tuning strategies based on the feedback from human evaluators could lead to further enhancements in the model's capabilities. Integrating human evaluation into the model development process would be crucial for refining the system and ensuring its alignment with human perceptions of text quality.

References

Harsh Jhamtani, Varun Gangal, Eduard H. Hovy, and Eric Nyberg. 2017. Shakespearizing modern language using copy-enriched sequence-to-sequence models. *CoRR*, abs/1707.01161.

Varvara Logacheva, Daryna Dementieva, Sergey Ustyantsev, Daniil Moskovskiy, David Dale, Irina Krotova, Nikita Semenov, and Alexander Panchenko. 2022. ParaDetox: Detoxification with parallel data. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6804–6818, Dublin, Ireland. Association for Computational Linguistics.

Mariano de Rivero, Cristhiam Tirado, and Willy Ugarte. 2023. Formalstyler: Gpt-based model for formal style transfer with meaning preservation - sn computer science.

Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhan Chao. 2019. Harnessing pre-trained neural networks with rules for formality style transfer. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3573–3578, Hong Kong, China. Association for Computational Linguistics.