Transformer-Based Latent GAN for Authorial Text Generation

Thien Nguyen
University of Texas at Dallas
DangThien.Nguyen@utdallas.edu

Timothy Choi
University of Texas at Dallas
Timothy.Choi@UTDallas.edu

Abstract

This proposal introduces a novel framework for adversarial text generation by integrating Transformer architectures into the Skip-Thought Generative Adversarial Network (STGAN) proposed by Ahamad et al. (2019). While the original STGAN leverages RNN-based Skip-Thought vectors to model sentence-level style, its reliance on sequential processing limits scalability and diversity. We propose replacing RNN components with Transformer autoencoders to improve latent space modeling and decoding efficiency. Our methodology combines pre-trained Transformer encoders (e.g., BART) for sentence embedding generation with Wasserstein GANs using gradient penalty regularization. Experiments will evaluate style preservation and text quality across fantasy literature and customer service dialogue datasets using automated metrics (BLEU, ROUGE) and human assessments. This hybrid approach aims to address mode collapse in traditional GANs while enhancing stylistic consistency through Transformer-powered attention mechanisms.

1 Introduction

Adversarial text generation has advanced significantly through architectures like Skip-Thought GANs (STGANs), which leverage sentence-level embeddings to improve semantic coherence. However, existing RNN-based STGAN frameworks face critical challenges: sequential processing limits its scalability for long texts, word-level style modeling struggles with authorial nuance, and traditional GAN objectives often induce mode collapse in open-domain generation. These limitations hinder practical applications in domains requiring both stylistic consistency (e.g., personalized dialogue systems) and semantic diversity (e.g., creative writing assistance). Our work addresses these gaps by proposing Transformer-STGAN, a novel architecture that replaces RNN components with Transformer autoencoders to enhance latent space modeling while integrating Wasserstein GAN objectives with gradient penalty for stable training. The system accepts unstructured text corpora (e.g., author-specific novels or customer service transcripts) and generates paragraph-length outputs that preserve both the input's stylistic patterns and contextual relevance. Our methodology combines pre-trained Transformer encoders for robust sentence embedding generation, conditional GAN discriminators with dynamic style weighting, and curriculum-based adversarial training to progressively refine stylistic fidelity. Experiments will evaluate performance across literary and technical domains using automated metrics (BLEU-4, ROUGE-L) alongside human assessments of style preservation, with ablation studies analyzing the impact of Transformer depth and attention mechanisms. By bridging adversarial training with modern self-attention architectures, this work aims to establish new benchmarks for controllable text generation in low-resource scenarios where traditional fine-tuning approaches prove inadequate.

2 Related Work

Transformer-based models have dominated the machine learning field in recent years, being utilized in natural language processing tasks such as machine translation, text generation, and speech recognition. Skip - Thought Vectors (Kiros et al., 2015) were created to address the problem of word embeddings and semantics. Skip-thought Vectors had a different approach; instead of word or character-based embedding, Skip-thought Vectors focused on entire sentences. This approach was able to generate(Ahamad, 2019) text by utilizing an encoder-decoder model to predict surrounding sentences in a text corpus.

A problem with early Recurrent neural network models and other variants created like LSTM and GRU is that they suffered from long-range dependency issues and high computational cost. The Transformer model (Vaswani et al., 2023) introduced a self-attention mechanism that significantly improved training efficiency and representation quality. This architecture was the foundation for modern NLP architecture like BERT(Devlin et al., 2019) and GPT(Radford et al.).

One of the most challenging tasks of Natural Language Processing is Text generation with consistently changing models from statistical to deep learning. Seq2Seq models (Sutskever et al., 2014) with attention mechanisms (Bahdanau et al., 2016) have been shown to improve text generation quality. The problem with GAN is that it struggles with stability and discrete output space. TILGAN (Diao et al., 2021) combines a Transformer autoencoder with a GAN in the latent space to enhance text generation performance. By matching the distributions of multitoken sequences in a Transformer latent space, TILGAN captures semantic information at varying sizes, improving local and global coherence in generation.

3 Methodology

1. Transformer Autoencoder for Sentence Embedding:

   We will utilize a pre-trained transformer encoder to generate robust sentence embedding from raw text. Transformers use self-attention mechanisms, which allow parallel processing and often capture global context more effectively. This leads to improved scalability and better latent space modeling.

2. Latent Space GAN with Enhanced Stability:

   Enhances the GAN component by using Wasserstein GAN objectives with gradient penalty (WGAN-GP) to provide smoother gradients and improved training stability. Additionally, techniques such as dynamic style weighting and curriculum-based training are introduced to further refine the generator's ability to produce stylistically consistent and diverse outputs.

4 Experiment Plan

Since the GANs are conditioned on attributes to generate data, both the generator and discriminator are going to be conditioned on the skip-thought encoded vectors. As we are limited on hardware and time, instead of using the entire BookCorpus, we will only use part of the BookCorpus to reduce the training time in order to accomplish a desirable outcome. We will continue using a split of 5/1/1, training/test/validation (Zhu et al., 2015). From the BookCorpus, we are going to use sentences belonging to a certain series and author to have a consistent and structured generated text.

The sentences generated will be used for evaluation on BLEU-n, METEOR, ROUGE, and BertScore. For the RNN-based models, we will use the higher score for evaluation compared to the transformer-based model, as transformer models have better scores. To compare the transformer models with our model, we will use BertScore to see which model can capture semantics and sentence representation.

References

Afroz Ahamad. 2019. Generating text through adversarial training using skip-thought vectors. In *Proceedings of the 2019 Conference of the North*, page 53–60. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2016. Neural machine translation by jointly learning to align and translate. *Preprint*, arXiv:1409.0473.

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.

Shizhe Diao, Xinwei Shen, Kashun Shum, Yan Song, and Tong Zhang. 2021. TILGAN: Transformer-based implicit latent GAN for diverse and coherent text generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4844–4858, Online. Association for Computational Linguistics.

Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2015. Skip-thought vectors. *Preprint*, arXiv:1506.06726.

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

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *Preprint*, arXiv:1409.3215.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention is all you need. *Preprint*, arXiv:1706.03762.

Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 19–27.