PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Instituto de Ciências Exatas e de Informática

Abordagens para melhorar a estabilidade e confiabilidade da Rede Adversária Generativa na geração de textos*

Approaches to improving the stability and reliability of the Generative Adversarial Network in text generation.

Bernardo Lara[1]
Felipe Campolina
Gabriela Colen
Marcelo Reis
Raphael Caetano

Resumo

Esse trabalho apresenta uma revisão bibliográfica sobre o assunto "Abordagens para melhorar a estabilidade e confiabilidade da Rede Adversária Generativa na geração de texto"com o fim de que os autores se familiarizem com o estado atual do conhecimento sobre o tema em questão. Além disso, essa revisão ajuda a contextualizar a pesquisa, fornecendo um quadro teórico que sustenta as conclusões e interpretações ao longo do que é apresentado na disciplina, para que possa ser desenvolvido um artigo científico com maior excelência. O tema foi escolhido pelos alunos, uma vez que a Inteligência Artificial está em evidência e gera diversas dúvidas e questionamentos. Foi selecionado o estudo sobre GANs, já que é uma técnica extremamente válida de geração de dados.

Palavras-chave: GANs. Revisão Bibliográfica. Artigo Científico. Inteligente Artificial.

*Artigo apresentado a matéria de Trabalho Interdisciplinar 3.

[1]Alunos do Programa de Graduação em Ciência da Computação, Brasil .

1 INTRODUÇÃO

A geração de textos é uma área de pesquisa multidisciplinar que envolve a aplicação de técnicas de processamento de linguagem natural, aprendizado de máquina e inteligência artificial para criar modelos capazes de produzir texto em linguagem humana. A produção desses textos é uma área de grande interesse em Ciência da Computação e Linguística Computacional, e as Redes Adversárias Generativas (GANs) são técnicas de aprendizado de máquina que tem recebido atenção crescente nesse contexto. As GANs consistem em um modelo generativo e um modelo discriminativo que são treinados de forma adversarial para gerar dados realistas a partir de uma distribuição de dados de entrada.

Embora as GANS tenham mostrado promessa na geração de textos em diversos cenários, como resumos de notícias, diálogos em linguagem natural, tradução automática e geração de descrições de imagens, a construção de textos usando GANs ainda enfrenta desafios significativos relacionados à estabilidade e confiabilidade dos modelos. Esses desafios limitam a qualidade e coerência dos textos gerados, tornando essencial buscar abordagens que superem essas limitações.

Os objetivos dessa pesquisa são analisar os avanços mais recentes em termos de arquitetura de modelo, técnicas de treinamento e métricas de avaliação no contexto da geração de texto com Redes Adversárias Generativas (GANs). Busca-se identificar as abordagens mais recomendadas na literatura para aprimorar a estabilidade e confiabilidade dos modelos GANs na geração de texto, considerando suas aplicações em resumos de notícias, diálogos em linguagem natural, tradução automática e geração de descrições de imagens.

Embora as GANs tenham se mostrado promissoras na geração de texto em diversas aplicações, ainda existem desafios significativos a serem enfrentados, especialmente em relação à estabilidade e confiabilidade do modelo. Esses desafios limitam a aplicação prática das GANs na geração de texto, uma vez que a qualidade e coerência dos textos gerados podem variar consideravelmente. Portanto, uma revisão bibliográfica se faz necessária para explorar as principais abordagens recomendadas na literatura, a fim de superar essas limitações e aprimorar a qualidade dos textos gerados pelas GANs. A partir dessa revisão, será possível identificar oportunidades de pesquisa e direcionar esforços para o desenvolvimento de novas técnicas e metodologias que melhorem a geração de texto com GANs, impulsionando avanços nessa área tão relevante para a linguística computacional e suas aplicações práticas.

2 REVISÃO BIBLIOGRÁFICA

Uma das principais formas de melhorar a estabilidade das GANs é utilizar técnicas de regularização, que ajudam a evitar o overfitting do modelo e melhoram a capacidade de generalização dos resultados. O artigo *Improved Training of Wasserstein GANs* (GULRAJANI et al., 2017) propõe o uso de uma técnica de regularização denominada “penalidade de gradiente”,

que penaliza a norma de gradiente da função discriminadora.

Já o trabalho de *Spectral Normalization for Generative Adversarial Networks* (MIYATO et al., 2018) foi proposto uma técnica de regularização denominada “spectral normalization”, que normaliza a matriz de peso da rede discriminadora. Além disso, um estudo recente sobre *Adaptive instance normalization for generative adversarial networks* (ZHANG et al., 2021) analisa o uso de uma técnica de regularização chamada "Adaptive Instance Normalization (AdaIN)"para melhorar a estabilidade das GANs. Essa técnica permite que a rede geradora manipule estatísticas de instância em diferentes camadas, ajustando a média e a variância de diferentes atributos da imagem.

Outra técnica de regularização para melhorar a estabilidade das GANs é a *Stabilizing Training of Generative Adversarial Networks through Regularization* (ROTH et al., 2017). O objetivo é impedir que a rede geradora crie amostras que sejam muito diferentes das amostras reais, levando a um treinamento instável e resultados de baixa qualidade. Introduzindo-se um termo de penalização no objetivo de treinamento da GAN, que limita a magnitude dos gradientes durante o treinamento. Isso ajuda a reduzir a variância dos gradientes e melhora a estabilidade do treinamento.

Outra forma de melhorar a estabilidade dos GANs é usar arquiteturas mais avançadas, como o Attention GAN (AGAN) ou o Recurrent GAN (RNN-GAN). A GAN proposto em *Adversarial Feature Augmentation for Unsupervised Domain Adaptation* (ZHANG et al., 2019) usa um mecanismo de atenção para guiar a geração de texto, enquanto os RNN-GANs propostos no estudo *Sequence Generative Adversarial Nets with Policy Gradient* (YU et al., 2017) utilizam redes neurais recorrentes para gerar textos com estrutura mais complexa.

Em contrapartida, o artigo *Improved Techniques for Training GANs* (SALIMANS et al., 2016) apresenta novas técnicas arquiteturais e de treinamento, onde o objetivo principal dos autores não é treinar um modelo que atribua alta probabilidade aos dados de teste, nem exigir que o modelo aprenda bem sem usar rótulos. Em vez disso, as técnicas propostas pelos autores visam melhorar a qualidade das amostras geradas pelas GANs e sua capacidade de aprender com exemplos não rotulados adicionais, assim alcançando resultados de ponta em tarefas semi-supervisionadas e na geração de imagens realistas.

Aliás, o uso de treinamentos mais avançadas também tem sido colocado como forma de melhorar a confiança das GANs na criação de textos. O artigo *Spherical Latent Spaces for Stable Variational Autoencoders* (XU et al., 2018) recomenda o uso da técnica de treinamento "self-imitation learning", que incentiva textos mais acessíveis na criação, dificultando assim que o modelo seja muito repetitivo. Já o trabalho *Adversarial Learning for Neural Dialogue Generation* (LI et al., 2017) sugere o uso da técnica de treinamento chamada "adversarial training with denoising autoencoder", que usa um autoencoder para pré-processar o texto, melhorando assim a qualidade do modelo.

Outras abordagens para melhorar a qualidade e a confiança na geração de textos das GANs incluem o uso de modelos preparados para ativar os pesos do modelo do estudo *Proceedings of the AAAI Conference on Artificial Intelligence, Improved TextGAN with Self-Attention-*

Guided Reinforcement Learning (HUANG et al., 2018), a junção de varias GANs, gerando assim textos mais acessíveis e confiáveis (WANG et al., 2018) e o uso de técnicas de pós-processamento para que os textos gerados sejam mais legíveis, explicitados em *Adversarial Network-based Chinese Poetry Generation*, (ZHAO et al., 2017).

3 METODOLOGIA

Primeiramente, a fim de explorar a área de geração de texto com Redes Adversárias Generativas (GANs), foi realizado um levantamento bibliográfico abrangente. Esse levantamento teve como objetivo identificar os principais estudos publicados sobre o assunto. Para garantir a seleção dos estudos mais relevantes e recentes, estabeleceram-se critérios de inclusão e exclusão pré-determinados.

No que diz respeito aos critérios de inclusão, foram considerados estudos que se concentram especificamente na geração de texto com GANs. Além disso, era necessário que os estudos apresentassem resultados empíricos, proporcionando uma base sólida para análise e avaliação. Por fim, foi levado em conta o fato de os estudos estarem publicados em revistas, conferências ou workshops reconhecidos, com revisão por pares, garantindo assim a qualidade dos estudos selecionados.

Por outro lado, foram estabelecidos critérios de exclusão para filtrar os estudos que não atendiam aos requisitos do estudo. Foram excluídos estudos que não eram relevantes para a geração de texto com GANs, ou seja, estudos que se concentravam em outras áreas ou abordagens. Também foram excluídos estudos que apresentavam resultados insuficientes ou não alcançavam um nível de confiabilidade necessário para a análise. Além disso, estudos disponíveis apenas em pré-impressões ou arquivos pessoais foram excluídos, visando obter uma base de dados consistente e confiável.

Após a seleção dos estudos mais relevantes, foram extraídas informações relevantes de cada um deles. Essas informações incluíam as abordagens recomendadas pelos autores para melhorar a estabilidade e confiabilidade das GANs na geração de texto. Para realizar essa tarefa, foi feita uma análise minuciosa dos artigos, buscando identificar as técnicas, metodologias e arquiteturas utilizadas pelos pesquisadores.

Com base nas informações extraídas e na análise e síntese desses dados, foram elaboradas conclusões abrangentes sobre as principais tendências e recomendações presentes na literatura para aprimorar a qualidade e confiabilidade das GANs na geração de texto. Essas conclusões têm como objetivo fornecer uma visão abrangente do estado atual da área, identificando lacunas, desafios e possíveis direções futuras de pesquisa para o avanço da geração de texto com GANs.

4 CRONOGRAMA

O cronograma de atividades adotado para o desenvolvimento deste trabalho foi cuidadosamente elaborado, levando em consideração as etapas necessárias para a realização do estudo e o tempo disponível para sua execução. O presente cronograma foi estruturado no início do curso da disciplina Trabalho Interdisciplinar III: Pesquisa Aplicada, pelos professores responsáveis, e foi seguido de acordo com a tabela abaixo:

Cronograma	
Data	Objetivo
03/02/2023	Início do projeto de pesquisa e estudo de técninas e metodologias
06/03/2023	Levantamento bibliográfico inicial
09/03/2023	Resumos iniciais
31/03/2023	Definição da equipe,problema e objetivo
28/04/2023	Discussão da revisão bibliográfica com os professores
12/05/2023	Apresentação prévia do projeto de pesquisa
09/06/2023	Apresentação final do projeto de pesquisa
16/06/2023	Entrega do pitch
23/06/2023	Entrega do texto final do projeto de pesquisa
03/02/2024	Refinamento e melhoria do projeto de pesquisa

A estruturação do cronograma proporcionou uma organização eficiente das atividades, permitindo um progresso contínuo e o cumprimento dos prazos estabelecidos. A adoção dessa metodologia contribuiu para a qualidade e sucesso do presente projeto de pesquisa.

5 CONSIDERAÇÕES FINAIS

Em suma, a revisão bibliográfica apresentada demonstra que as GANs são uma ferramenta promissora para gerar textos em diferentes cenários e aplicações. No entanto, a construção de textos utilizando essa técnica ainda enfrenta desafios em relação à estabilidade e confiabilidade do modelo. Felizmente, existem diversas abordagens que podem ser empregadas para melhorar esses aspectos.

Uma das abordagens promissoras para melhorar a estabilidade das GANs na geração de texto é o uso de técnicas de regularização. Entre elas, destacam-se a "penalidade de gradiente", que adiciona um termo de penalidade ao treinamento para evitar explosões de gradiente, a "spectral normalization", que normaliza os pesos da rede para controlar a magnitude do gradiente, e a "Adaptive Instance Normalization", que ajusta a normalização de instâncias em tempo real para aumentar a estabilidade durante o treinamento.

O uso de arquiteturas mais avançadas também tem sido explorado como forma de melhorar a estabilidade das GANs na geração de texto. Exemplos incluem o AGAN (Attentional Generative Adversarial Network), que incorpora mecanismos de atenção para direcionar a geração de palavras mais relevantes, e o RNN-GAN (Recurrent Neural Network GAN), que utiliza

redes neurais recorrentes para capturar a dependência temporal das sequências de texto.

Além das técnicas de regularização e das arquiteturas avançadas, abordagens de treinamento mais sofisticadas têm sido propostas para aprimorar a qualidade e confiabilidade dos textos gerados pelas GANs. O "self-imitation learning"é uma técnica que incentiva o modelo a aprender a imitar seus próprios textos gerados anteriormente, enquanto o "adversarial training with denoising autoencoder"combina o treinamento adversarial com um autoencoder para remover o ruído das representações de texto.

Outras abordagens mencionadas na literatura incluem o uso de modelos de linguagem pré-treinados para inicializar os pesos do modelo gerador, a junção de várias GANs para melhorar a diversidade dos textos gerados e técnicas de pós-processamento para melhorar a legibilidade e coerência dos textos.

Em geral, a revisão bibliográfica apresentada mostra que aprimorar a estabilidade e confiabilidade das GANs na geração de texto é um campo em constante evolução. Novas abordagens estão sendo constantemente desenvolvidas e testadas na literatura científica para superar os desafios existentes. É interessante observar também a grande quantidade de materiais orientais sobre o assunto, indicando um forte interesse e contribuição dessa região para o avanço da pesquisa em geração de texto com GANs. Essas descobertas ressaltam a importância de continuar investigando e explorando novas técnicas e abordagens para impulsionar o progresso nesse campo multidisciplinar.

REFERÊNCIAS

GULRAJANI, Ishaan et al. Improved training of wasserstein gans. **Advances in Neural Information Processing Systems**, v. 30, p. 5767–5777, 2017.

HUANG, Kexin; WANG, Zhe; CHEN, Chaoyue. Improved textgan with self-attention-guided reinforcement learning. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2018. v. 32, n. 1, p. 6975–6982.

LI, Jiwei et al. Adversarial learning for neural dialogue generation. In: **Proceedings of the Conference on Empirical Methods in Natural Language Processing**. [S.l.: s.n.], 2017. p. 2157–2169.

MIYATO, Takeru et al. Spectral normalization for generative adversarial networks. In: **International Conference on Learning Representations**. [S.l.: s.n.], 2018.

ROTH, Kevin et al. Stabilizing training of generative adversarial networks through regularization. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2017.

SALIMANS, Tim et al. Improved techniques for training gans. In: . [S.l.: s.n.], 2016.

WANG, Xinyu et al. Topic guided variational autoencoder for neural text generation. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2018. v. 32, n. 1, p. 7242–7249.

XU, Jiaming et al. Spherical latent spaces for stable variational autoencoders. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2018. p. 5829–5839.

YU, Lantao et al. Seqgan: Sequence generative adversarial nets with policy gradient. In: **AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2017. p. 2852–2858.

ZHANG, Hongyi et al. Adaptive instance normalization for generative adversarial networks. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 32, n. 6, p. 2377–2389, 2021.

ZHANG, Yaxing et al. Adversarial feature augmentation for unsupervised domain adaptation. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2019. p. 9519–9528.

ZHAO, Tianyu et al. Adversarial network-based chinese poetry generation. In: **Proceedings of the International Conference on Computer Science and Network Technology**. [S.l.: s.n.], 2017. p. 665–669.