Proposta de Projeto de TCC
Uso de redes neurais generativas para síntese de voz
Instituto de Matemática e Estatística da USP
Trabalho de Formatura Supervisionado

Estudante: Lucy Anne de Omena Evangelista (lucy.omena@usp.br)

Supervisor: Carlos Eduardo Elmadjian (cadu.elmadjian@usp.br)

03/05/2022

1 Objetivo e escopo geral

O objetivo principal do trabalho de conclusão de curso será oferecer base para a compreensão do que é uma rede neural generativa que síntese voz, de forma a exemplificar aplicações e a caracterizar pilares teóricos necessários para a mesma.

Macrotópicos a serem abordados são:

- Breve introdução teórica à redes neurais
- Redes neurais generativas para síntese de voz e sua arquitetura
- Redes neurais generativas no mercado

2 Introdução

Como continuação da bolsa de iniciação científica que recebeu bolsa CNPq de identificação 120151/2019-7, a presente proposta de TCC tem como objetivo caracterizar redes neurais para síntese de voz. A iniciação científica original é fruto de uma bolsa PIBITI (Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação) integrante do projeto A.D.A. (Assistente Administrativa Avançada), do

grupo de extensão USP Code Lab do IME-USP, que teve como objetivo o estudo de interfaces de voz, campo de estudo de IHC (Interação Humano-Computador).

Durante a iniciação científica foi estudado o processo prático de sintetização de voz a partir de entradas de texto em linguagem natural pois o conhecimento sobre o processo e a manipulação desse tipo de interface é imprescindível para que se tenham meios de adaptar ou mesmo propor novas técnicas de interação por voz em IHC para a proposta do projeto ADA.

Esta proposta de TCC vem de forma a estudar mais profundamente redes neurais generativas que tem como finalidade realizar a síntese de voz. Assim, busca-se explorar aplicações possíveis das mesmas, em conjunto com a base teórica necessária para que seja possível caracterizar essas redes das demais.

3 Motivação

O estudo de redes neurais generativas não se mostra massificado no meio acadêmico, o que torna difícil o estudo das mesmas por parte de estudantes, em particular de estudantes brasileiros pois há pouca literatura em português. Com isso, o TCC tem como objetivo ser porta de entrada para o estudo de redes neurais generativas, abordando o contexto e desenvolvimento do ramo quando se trata do tema de geração de voz.

Todo o ramo teórico e prático se beneficiam quando mais pessoas o compreendem e estudam, seja de forma a avançar a teoria do ramo ou ao facilitar o estudo e aplicação do mesmo ao se descobrirem novas técnicas e tecnologias especializadas.

4 Método

O método utilizado para desenvolvimento deste trabalho é o estudo da literatura e a pesquisa aberta sobre aplicações no mercado. A literatura inicial proposta pra o trabalho se encontra na seção Referências.

5 Referências

[Abu-Mostafa, 2012] Abu-Mostafa, Y. S. (2012). *Learning from data: a short course*.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

[Luong et al., 2015] Luong, M.-T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.

[Patterson and Gibson, 2017] Patterson, J. and Gibson, A. (2017). *Deep learning: A practitioner's approach*. "O'Reilly Media, Inc.".

[Ping et al., 2018] Ping, W., Peng, K., and Chen, J. (2018). Clarinet: Parallel wave generation in end-to-end text-to-speech. *arXiv preprint arXiv:1807.07281*.

[Ping et al., 2017] Ping, W., Peng, K., Gibiansky, A., Arik, S. O., Kannan, A., Narang, S., Raiman, J., and Miller, J. (2017). Deep voice 3: Scaling text-to-speech with convolutional sequence learning. *arXiv preprint arXiv:1710.07654*.

[Shen et al., 2018] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerrv-Ryan, R., et al. (2018). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4779–4783. IEEE.

[Sotelo et al., 2017] Sotelo, J., Mehri, S., Kumar, K., Santos, J. F., Kastner, K., Courville, A., and Bengio, Y. (2017). Char2wav: End-to-end speech synthesis. *Workshop track - ICLR 2017*.

[Wang et al., 2017] Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., et al. (2017). Tacotron: Towards end-to-end speech synthesis. *arXiv preprint arXiv:1703.10135*.