Efficient and Controlled Non-Autoregressive Text Generation

Jinghong Chen / Bill Byrne

May 2023

Background

Most modern neural text generation models are auto-regressive (AR) large language models, which means they condition on previously generated tokens to generate the next token sequentially. While effective, AR models have a few intrinsic drawbacks:

1. Generation time scales linearly with the length of generated sequences as the output tokens are generated one-by-one. This can be problematic for response-time critical tasks such as dialogue response generation;

2. It is difficult to control auto-regressive models so that they generate specific words, phrases, and entity names, a crucial ability for many text generation tasks.

In this project the student will explore state-of-the-art non-auto-regressive (NAR) models to achieve faster, more controlled text generation on the Schema-Guided Dialogue [Rastogi et al., 2020] and DART [Nan et al., 2020] dataset for dialogue response generation and data-to-text tasks, respectively. The student will build on an implementation of a state-of-the-art NAR model, the DA-Transformer [Huang et al., 2022], to investigate better training and decoding schemes that improves the quality of NAR text generation.

Familiarity with the Transformer architecture and Pytorch will be helpful. Good Python skills are essential.

References

F. Huang, H. Zhou, Y. Liu, H. Li, and M. Huang. Directed acyclic transformer for non-autoregressive machine translation. In *International Conference on Machine Learning*, pages 9410–9428. PMLR, 2022.

L. Nan, D. Radev, R. Zhang, A. Rau, A. Sivaprasad, C. Hsieh, X. Tang, A. Vyas, N. Verma, P. Krishna, et al. Dart: Open-domain structured data record to text generation. *arXiv preprint arXiv:2007.02871*, 2020.

A. Rastogi, X. Zang, S. Sunkara, R. Gupta, and P. Khaitan. Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8689–8696, 2020.