Josias Moukpe

Neural Networks

12-02-2022

Project 3 Proposal

Text Generation is a branch of Natural Language Processing (NLP) that predicts the next most likely word given all the previous words. State of the art in Text Generation has recently been the domain of Language Models based on Transformers pretrained on large corpora of text data from various sources on the internet. These pretrained generative transformers are often fine-tuned on a downstream dataset for a downstream task of interest. This project aims to develop a generative language model based on a pretrained language transformer fine-tuned on our two chosen datasets: Merve-Poetry [1] and Rap Lyrics US [2]. We want to obtain a fine-tuned model that can generate original English poetry given an initial prompt and another that can generate original rap lyrics inspired by the styles of contemporary rappers. Both datasets contain English text from various authors and will be processed to extract the bags of words. Those words will then be converted to a designated embedding before being ingested by the generative model during fine-tuning. The pretrained model will be obtained from the Hugging Face library of generative models for English text.

Merve Poetry contains renaissance and modern poetries from subjects such as Love, Nature, Mythology, and Folklore. The dataset is organized into 5 table columns: the poem's content, author, age, and subject type. Rap Lyrics US contains text documents organized by artists. Each text document includes lyrics from the corresponding artist. The final models will take an initial sequence of words (prompt) and generate a sequence from an original poem and rap song. The final architecture and hyperparameters of the model will be determined during the experimentation phase.

I will work solo on this project and handle the dataset, research, and development myself. Any other updates and additional datasets used will be mentioned in the report.

References:

[1] https://huggingface.co/datasets/merve/poetry/viewer/merve--poetry/train

[2] https://github.com/fpaupier/RapLyrics-Scraper/tree/master/lyrics_US

[3]https://huggingface.co/models?language=en&library=tf&pipeline_tag=text-generation&sort=downloads