3110 MS3 Report

Haadi Khan, Domenic Fioveranti, Jiming Chen, Will Bradley

December 12, 2024

Vision

Our vision is to implement a transformer architecture-based neural network trained on high-performing (i.e. highly-upvoted) posts on the popular social media app "Sidechat." The posts will be gathered specifically from Sidechat's Cornell community, which will encourage our model to output Cornell-centric posts. As a test of the model's abilities, our group will make a series of posts to the app using text solely generated from our model, in the hope that the model can mimic an entertaining, human Cornell student and thus garner a positive following. We ran into a minor technical difficulty since MS2, namely our compute power. So, our hope of a high-performing sidechat account has since changed, but the rest of our vision is the same.

Summary

For MS2, Domenic created the initial tokenizer (modeled after Google's wordpiece) for the system. However, we ran into some technical issues with its training stage and decided to abandon it after demoing it in MS2. JJ created the initial transformer stack, including the tokenizer, transformer, and pretraining. Haadi refined the transformer, made it more modular, and created a moderation system for the model. Will optimized the transformer code for pretraining to utilize our limited resources more effectively and wrote tests for matrix. Domenic finished the test suite and built the fully-featured TUI.

A notable change from MS2 is that we decided to abandon the OCaml Torch library. We encountered considerable difficulty with installation for our group. Despite our eventual successful installation, we decided to avoid the burden for our graders, and the required installation of libtorch.

Activity Breakdown

Domenic Fioravanti (dmf252)

- Researched transformer architecture and planned model implementation
- Developed extensive test suite for the project
- Built our refined TUI
- 24 hours

Haadi Khan (hmk68)

- Built the initial driver for the end-user to interact with
- Group planning and role assignment
- Refined transformer model and created a content moderation system
- Aided with test suite + final TUI
- 25 hours

Jiming Chen (jc3579)

- Did data wrangling by finding the Sidechat API and made 5,000 API calls to retrieve 86,980 text posts along with their upvote count, comment count, and timestamps

- Built initial transformer architecture

- Worked on pretraining

- 26 hours

Will Bradley (wjb247)

- Optimized the training process

- Worked with data sets to optimize our model performance

- Made minor contributions to driver

- 25 hours

Productivity analysis

We were efficient during MS3. In our sprints we developed the transformer relatively fast and did a good job overall. The main issue we ran into was compute. Unfortunately, we did not have adequate compute resources to utilize GPUs to speedup our training process which resulted in sub-par performance in the comprehensibility of our eventual model. If we constructed our transformer earlier in the semester we might have been able to do a better job with our project.

That being said, every person in the group pulled their weight and did a good job with their respective responsibilities. We did everything we sought out to do at the beginning of the semester, and our only issue was a lack of compute power for our model's training phase.