AI 4 Science - Fall 2023 Final Report

Ayush Karmacharya (psb7wm)

December 2023

1 Overview

During the course of this fall, I spent time working on several different projects. A bit of time was spent preparing for the project to predict future university budgets via student body info; however, most of this was foundational, and the bulk of the work is expected to be completed during the spring next semester. By contrast, most of this semester was spent working on sensitivity analysis related papers, outlined in the COVID-19-age-groups repository in the ML-Sys group. These, and other work performed during this semester are outlined in this report below.

2 Financial Project

As mentioned earlier, the bulk of the work on the financials project is expected to be done during the spring, as we spent most of this semester waiting for the data to be prepared as it met the various university protocols. Outside of meeting with the collaborators briefly, the main thing regarding this project I did do, however, was to make sure I and my PC were both able to obtain the data when it would be ready to be used. To do this, I completed all the necessary forms brought to my attention along with the necessary security modules so that I could be certified to handle the data and get the needed VPN. With this, I'm ready to advance further along with the project come spring.

3 Sensitivity Analysis

My first main contribution to the sensitivity analysis work was in the writing of the poster paper to the AAAI 2023 conference. I was responsible for writing the 'Related Work' section as seen below, on top of also reading over the rest of the paper in order to gain a better understanding of the research at hand and check for any blatant flaws.

Related Work

Numerous past works have explored various sensitivity analysis methods. Utans et al. (1995) introduced a sensitivity analysis technique that determined a saliency measure for each feature variable, which is the resulting training error after its removal from the model. Gasca, Sánchez, and Alonso (2006) measured the relative importance of each input feature and sought to measure redundant features by finding whether any pair of features reached some benchmark value of correlation.Naik and Kiran (2021) took a complex-step perturbation approach (CSPA) to perform sensitivity analysis and determine the relevant features. Ramchandani, Fan, and Mostafavi (2020) explained the feature interactions in the growth classification of cases by randomizing feature values and ranking them based on performance drop.

After this, I focused on reproducing the models and running the sensitivity analysis methods on each model. In doing this, I learned better how to run the code on Rivanna along with the codebase and relevant libraries used such as Captum. I did this by first training the PatchTST model using the entire dataset, after which I added my results to the 'Results' section of the COVID-19-age-groups repository (in which I also created the table for others to add their results as they ran other models, as instructed by Khairul).

The next thing I did was run each of the sensitivity methods on the trained models. I did this for both the PatchTST and Fedformer models. I first ran the methods on the PatchTST model as it was the one I had originally trained; however, we later found that some of the gradient-based sensitivity methods didn't work with PatchTST, causing me to then focus on Fedformer as it was the model we then planned to have our sensitivity analysis paper focus on. Once I completed all this training, I recorded all the results in a shared spreadsheet made by Khairul, allowing for verification that proper results were achieved.

After this point, Khairul brought up the novel mechanism involving performing sensitivity analysis with ground truth versus without ground truth. As such, I then ran each sensitivity model on the Fedformer model with and without ground truth. At this point, Khairul had written most of the draft of the Age Sensitivity paper to be submitted to the AAAI 2024 workshop. As such, I then focused on verifying my results and reading over the paper to build understanding and ensure there were no major flaws within the paper which were missed beforehand. At this point, we are basically at that point of the review process as we prepare for the paper to be submitted for the upcoming deadline.

4 Paper Reading

For the paper reading presentation, I did my presentation on the paper "Acceleration of Large Transformer Model Training by Sensitivity-Based Layer Dropping" by Yujie Zeng et al. This paper was presented at the AAAI 2023 conference. The paper proposes a method to speed up the training of transformer models known as Progressive Layer Dropping (PLD). I chose to make my paper reading presentation on this topic for quite a few reasons. Firstly, the paper itself was from the AAAI conference, a conference to which we had and were planning to submit papers to this semester; thus, in reading it I grew more familiar with the expected writing style of papers submitted to this conference, and also a greater understanding of what topics and trends may be favored by reviewers at the conference. I also thought the paper to be quite relevant in that it covered topics in making the processing of data more efficient. This is especially important in a world we want to process lots of data; we want to accelerate training times while maintaining accuracy, which is exactly what this paper proposed to do. Finally, I found there to be some relevance in that the paper had a focus on transformer models, similarly to the GPCE-COVID project done in the summer and the current sensitivity analysis paper worked on this fall.