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

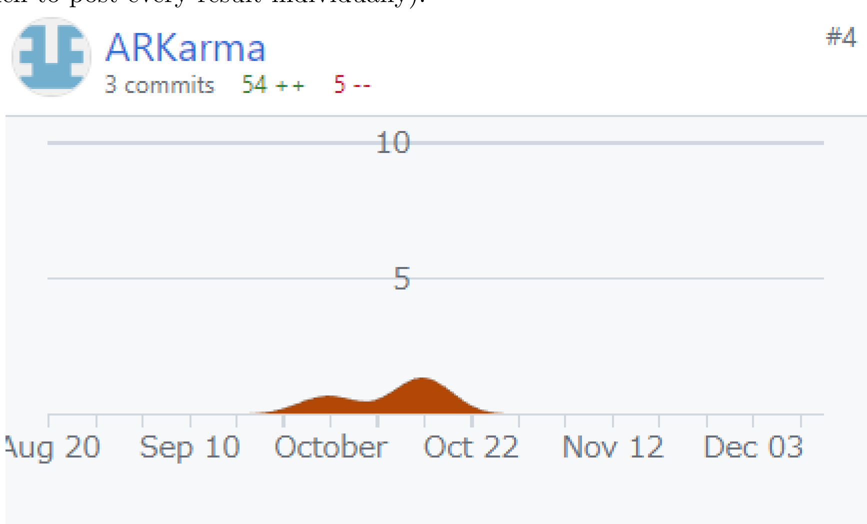
Numerous past works have explored various sensitivity analysis methods. Utans et al. (1995) introduced a sensitivity analysis technique that determined a saliency measure for each feature variable, which is the resulting training error after its removal from the model. Gasca, Sánchez, and Alonso (2006) 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 (The results of which can be seen in a table attached at the end of this section), 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.

Below here, a few images of the work done can be seen. The first shows some of my commit history in the COVID-19-age-groups repository. The second shows a table from the repository that includes the results of training the total dataset on each model. I added this table to the readme, along with adding the results for the PatchTST model. The third image is to demonstrate my initial reproduction of applying the interpretation methods to the trained Fedformer model, which was recorded in a spreadsheet provided by Khairul (which itself is where I took the image from). The fourth image is a demonstration of my reproduction of my interpretation without ground truth using the Morris Sensitivity methods. I did each sensitivity method in a separate file, so I also had the image afterwords show all of these files in Rivanna (as I figured it was a bit much to post every result individually).



Results

Test Results

The following table shows the test results for the COVID-19 dataset, calculated daily at each US county.

Model	mae	rmse	rmsle	r2
Autoformer	33.701	188.330	1.826	0.457
DLinear	29.525	174.950	1.425	0.524
FEDformer	31.490	180.450	1.659	0.499
PatchTST	33.174	183.647	1.530	0.469
TimesNet	34.354	191.920	1.604	0.415
Transformer	34.818	192.720	1.601	0.410

Local evaluations

Metrics	feature_ablation	occlusion	augmented_occlusion	deep_lift	lime	integrated_gradients	gradient_shap	morris_sensitivity
mae	0.125	0.125	0.047	0.172	0.094	0.203	0.188	0.125
rmse	0.147	0.147	0.077	0.238	0.117	0.25	0.242	0.153
ndcg	0.975	0.975	0.982	0.969	0.98	0.961	0.964	0.976
normalized_mae	0.034	0.034	0.032	0.091	0.033	0.088	0.09	0.032
normalized_rmse	0.045	0.045	0.04	0.132	0.038	0.115	0.126	0.038
normalized_ndcg	0.943	0.943	0.996	0.893	0.969	0.889	0.892	0.964

Global ranks

Age group	feature_ablation	occlusion	augmented_occlusion	deep_lift	lime	integrated_gradients	gradient_shap	morris_sensitivity	Cases
UNDER5	5.673		5.65	6.135	5.67	6.717	6.401	6.066	3.57
AGE517	12.084		13.419	6.778	14.497	6.019	7.232	14.493	14.48
AGE1829	16.077		17.174	7.477	17.388	7.547	8.466	17.722	24.59
AGE3039	16.668		15.977	15.18	15.403	16.96	16.112	14.312	19.33
AGE4049	13.247		12.993	6.796	12.215	6.826	4.741	13.346	14.1
AGE5064	16.259		16.341	46.92	17.466	39.644	44.744	17.736	15.89
AGE6574	12.716		10.933	8.114	10.193	13.134	9.582	9.77	5.07
AGE75PLUS	6.554		7.513	2.6	7.169	3.154	2.723	6.556	2.98

Global scores

Age group	feature_ablation	occlusion	augmented_occlusion	deep_lift	lime	integrated_gradients	gradient_shap	morris_sensitivity	Cases
UNDER5	8		8	7	8	6	6	8	7
AGE517	5		4	6	4	7	5	3	4
AGE1829	3		1	4	2	4	4	2	1
AGE3039	1		3	2	3	2	2	4	2
AGE4049	4		5	5	5	5	7	5	5
AGE5064	2		2	1	1	1	1	1	3
AGE6574	6		6	3	6	3	3	6	6
AGE75PLUS	7		7	8	7	8	8	7	8

```
Starting experiment. Result folder scratch/FEDformer_Total.
Use GPU: cuda:0
fourier enhanced block used!
modes=32, index=[0, 1, 2, 3, 4, 5, 6]
fourier enhanced block used!
modes=32, index=[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
 fourier enhanced cross attention used!
modes_q=10, index_q=[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
modes_kv=7, index_kv=[0, 1, 2, 3, 4, 5, 6]
adding time index columns TimeFromStart
added time encoded known reals ['month', 'day', 'weekday'].

Train samples 2001454, validation samples 87976,             test samples 87976, last samples 1253658
637 days of training, 14 days of validation data,             14 days of test data and 385 of data after test start.

Fitting scalers on train data
Loading dataset from ./dataset/processed/Total/train.pt
Loading dataset from ./dataset/processed/Total/val.pt
Loading dataset from ./dataset/processed/Total/test.pt
loading best model from scratch/FEDformer_Total/checkpoint.pth
Interpretations will be saved in scratch/FEDformer_Total/interpretation
Running morris_sensitivity from 2023-11-13 11:11:53.126795
Experiment ended at 2023-11-13 11:23:42.832253. Total time taken 0:11:49.705458.
  metric  area       comp      suff
0    mae  0.05   4.938901  8.160419
1    mae  0.10   7.546891  6.710694
2    mse  0.05   5.899421  9.894899
3    mse  0.10  12.412353  5.908436
```

Fedformer_deeplift_wo_ground.out	1.29 KB	12/1/2023 1:44:26 PM
Fedformer_gradientshap_wo_ground.out	1.29 KB	12/1/2023 1:51:36 PM
Fedformer_featureablation_wo_ground.out	1.3 KB	12/1/2023 2:01:33 PM
Fedformer_featurepermutation_wo_ground.out	1.29 KB	12/1/2023 2:11:11 PM
Fedformer_integratedgradients_wo_ground.out	1.3 KB	12/1/2023 2:19:32 PM
Fedformer_lime_wo_ground.out	1.28 KB	12/1/2023 6:42:25 PM
Fedformer_occlusion_wo_ground.out	1.29 KB	12/1/2023 6:52:11 PM
Fedformer_augocclusion_wo_ground.out	1.29 KB	12/1/2023 7:02:00 PM

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.