

In-class programming session

CPSC 471/571 - March 4, 2024

Session Preparation

Download the VQA V2 validation data, questions, and annotations from [this webpage](#).

Please ensure that you have access to the GPU cluster and have downloaded the dataset. Check that the GPU access is working by importing torch and placing a tensor on the GPU.

```
import torch
z = torch.zeros(3, 3).cuda()
```

This part may take a while, so we ask that you complete this step before the session.

Part 1: Multimodal Interpretability

In this part, you will fine-tune a simple Visual Question Answering (VQA) model. Then, you will apply the Layer Integrated Gradients method for interpretability of the multimodal model. You will use the following [notebook](#) for this part, which contains most of the necessary code.

Steps:

1. Follow instructions to fine-tune the [ViLT](#) model using the GPU cluster. You may need to adapt the file paths accordingly for your particular environment.
2. Answer two questions on the existing code. Then complete the implementation by instantiating an appropriate `LayerIntegratedGradients` object.
3. Generate interpretability visualizations for the three example images.

Part 2: Dataset Exploration

In this part, you will download a dataset for your project and conduct some preliminary analysis.

Steps:

1. Visualize a few examples from the dataset. Do you observe any interesting patterns? Are there any issues that you notice with the dataset that may impact the results?
2. Plot the class distribution. Are there any categories with relatively few examples?
3. *(Extra credit)* Perform additional qualitative/quantitative analyses: feature distribution, clustering, word clouds, correlation analysis, error analysis of baseline models, etc.

Submission Guidelines

Submit your notebook along with your responses in PDF format on Canvas.