

Daily Log

Monday December 9

Today I was having a bunch of issues with TJ's j26 machine. I was trying to start working with a PyTorch transformers library that's available on GitHub (<https://github.com/huggingface/transformers>). I'd created my environment val1 on the machine and installed the necessary packages onto the environment. From the terminal, they all appeared to install properly. But when I tried to actually import and use them in my Jupyter notebook, I keep getting module not found errors and I can't figure out why. I made sure my Jupyter notebook was running on the same environment (val1) that I did my installation on, but it's still not working. When I do "conda list" from the terminal to see the packages I've installed, the packages I need show up.

Tuesday December 10

Today, I was again trying to fix my environment issues on the cluster. I don't think that my Jupyter notebook is running on my val1 environment, even though I think I configured it to run there.

At home, I was able to fix the issue by running a bash command to force Jupyter to open in the correct environment. Dr. White found the command for me and gave it to me during Parallel. :) Also at home, I worked on installing some of the packages that I knew I was going to need on the environment.

Thursday December 12

With my environment issues finally solved, I was able to get some sample question/answer training data and get a model to start training. It took me a while to get the data into the write folders and download all the necessary packages, because the open-source libraries that I was using (see link to huggingface repository from Monday) require both Tensorflow and PyTorch to be installed, and I had some issues with package versions. It also took me some time to figure out how I should be saving and accessing the training data that I'd saved on my machine. Once I got all those errors figured out, though, the training did start to run and I'm planning on running it to completion next class.

Timeline

Date	Goal	Met
Friday, Dec 20	Finish up my goals from last week that I didn't quite complete yet, and make sure my winter goal and documentation on GitHub for it is all ready to go.	Yes
Winter Break Goal	Implement a basic question answering model based on BERT. Planning on basing it off of previously implemented models and guides.	Yes
Sunday, Jan 19	Train and evaluate a PyTorch question/answering model using the open source huggingface transformers library.	
Sunday, Jan 26	Start tweaking the model made with transformers to increase accuracy by 2%.	-

Reflection

On the whole, I struggled a lot with environments this week, but I'm happy that those issues all appear to be solved now. I finally got all my packages installed with the correct versions in the right locations so that I can start working with my model that uses the language transformer libraries.

I'm glad I moved towards working on a GPU machine rather than using the GPU from Google Colab, though. Over time I found that while my Colab notebooks were fine for running small tests, it wasn't practical as things started getting larger and requiring more external files and packages to be downloaded. Using Colab requires rerunning and reinstalling things every time you reopen an old notebook, which just turned out to be impractical.

Next week, I'm going to be focusing on training my first model that uses transformers and then evaluating it.