Sentiment Analysis Models – Learning Python Libraries

AP Research & Senior Research Project

Ojas Chaturvedi

https://github.com/ojas-chaturvedi

This week, my focus has been on learning major Python libraries that will be useful for my project. First, I have been working on learning PyTorch, a famous open-source machine-learning library that provides a high-level interface for building and training deep-learning models. There are many alternatives to it, but the main reason I chose PyTorch was for its seamless GPU acceleration, beneficial for training large models or processing large datasets, and its easy integration with other Python libraries. Specifically, to understand how PyTorch can help me with my implementation of sentiment analysis models, I have been following an NLP-specific tutorial on the **PyTorch docs**. I have also been tinkering with TensorBoard, a visualization toolkit specific to machine learning. TensorBoard will help me create graphs with my data for 2 main reasons: for me to make sure that the models are training well, and so that I can create graphs with my final results to add to my paper and presentation for viewers to understand. Since TensorBoard can be implemented with PyTorch, I have been reading its **documentation** on how to use the library.

Meanwhile, I have also made a small change to my web scraper script. I decided to also include 2 key sections per legislation that will be useful in generating more insights into the field of firearm regulation: Date of Introduction and Party of Sponsors. With the Date of Introduction, I will be able to create a graph of legislation sentiment scores over time to understand more about how the framing of the issue by legislators has changed over time. If time allows, I can take this to the next step and research major firearm events at specific time points to hypothesize a correlation between these major events and legislator's framing. I also included the Part of Sponsors to see if sentiment scores remain similar per Party, which can highlight the stance the Party takes on this issue of firearm legislation.

As for my internship at the ASU Data Mining and Reinforcement Learning Lab, we have made a lot of progress toward our project, which is focused on adding natural language constraints to Safe Reinforcement Learning (which is described in my second blog post titled 'Data Collection - API Exploration And Web Scraping Strategies'). After training a model to optimally navigate a fellow team member's environment, which is filled with constraints, as one of our baseline examples, I have started brainstorming some ideas for adding the 'natural language' constraints part to the environment. Unfortunately, I cannot explain the specifics of my plan, but I can give the idea briefly: my idea is to use LLMs to take the natural language constraints and turn them into code for the environments. With an increased use of LLMs

like ChatGPT or Llama, many developers have started creating plugins to help specifically with coding. However, this still is a very new concept and current famous plugins are very generalized in terms of the programming language and concepts. In addition, the code can sometimes be verbose or memory/space inefficient. After searching through Github and the ChatGPT plugins store, I have found a couple of plugins/LLMs focused towards RL, but are in the very beginning stages and still very inefficient. I am currently waiting for this idea to get approved by the Professor I am working with, and then will get started on implementing this idea. Meanwhile, I have been working with a fellow team member to start a literature review of the field for our paper and to find more baseline experiments to train various models with constraints.

In the upcoming week, I plan to continue working on finishing working on my classification script for the legislative texts. Unfortunately, I could not work on it this week, as I wanted to start working on some models to generate some preliminary results for practice presentations. I will still be working on the models, but I want to finish the classification script as well so that I can be completely done with my initial data collection and only have the results from the sentiment analysis models left. In the meantime, I will be completing my work for my internship and AP Research class, as always.