SE 491 WEEKLY REPORT 10
11/15/2024 - 11/21/2024

Group number: sdmay25-30

Project title: Explainable AI for source code applications

Client &/Advisor: Arushi Sharma

Team Members:
Manjul Balayar
Kellan Bouwman
Sam Frost
Akhilesh Nevatia
Ethan Rogers

Weekly Rotation
Manjul Balayar
Kellan Bouwman
Sam Frost
Ethan Rogers
Akhilesh Nevatia

Weekly Summary

This week efforts were focused on refining the clustering-alignment workflow, testing implementations, and advancing the integration of clustering into the broader pipeline. Comprehensive documentation was developed, including updates to wikis, READMEs, and code comments, alongside improvements to library structure and adherence to PEP8 style guidelines. Development began on new workflows derived from code concept networks, and research continued on implementing hydra across modules in NeuroXCode. Additionally, the pipeline integration connecting activations, clustering, and labeling was set up, with test scripts developed for real processed activations. These steps collectively contribute to the project's robustness and scalability.

- **Past week accomplishments**

  Manjul Balayar:
  - Worked on clustering-alignment workflow
  - Testing the workflow on a notebook by setting up tests

  Kellan Bouwman:
  - Documentation for wiki's and readme's
  - Code comments and documentation
  - Enforcing Pep8 style guidelines
  - Cleaning library structure
  - Starting development of workflows from code concept net

  Sam Frost:
  - Continued hydra implementation for clustering
  - Research how to implement hydra with other modules in NeuroXCode

  Akhilesh Nevatia:
  - Continued mapping Clustering to new CoCoA directory structure
  - Setting up integrated pipeline (Activations - Clustering - Labeling)
  - Test script for Clustering on real processed Activations

  Ethan Rogers:
  - Write docs for my modules, trying to make them look nice
    - Experimenting with auto-documentation means
  - Looking into CodeConceptNet's LLM datasets for DSPY and LLM training for dataset production

- **Pending issues**

  Manjul Balayar: N/A
  Kellan Bouwman: N/A
  Sam Frost: N/A
  Akhilesh Nevatia: N/A
  Ethan Rogers: N/A

  Individual Time Contributions

Name	Hours This Week	Total Hours
Manjul Balayar	6	59
Kellan Bouwman	5	66
Sam Frost	4	51
Akhilesh Nevatia	4	65
Ethan Rogers	4	61

- o **Plans for the upcoming week**

  Manjul Balayar:
  - ● Focus on showcasing the integrated pipeline in the notebook
  - ● Map clustering-alignment over to new directory, managing the scripts folder in CodeConceptNet

  Kellan Bouwman:
  - ● Pipeline focus
  - ● Example notebooks
  - ● Documentation for examples

  Sam Frost:
  - ● Finish Hydra implementation for clustering
  - ● Continue to next module as needed for project
  - ● Return to implementing unit tests

  Akhilesh Nevatia:
  - ● Hyperbolic Clustering Refactoring
  - ● Clustering Evaluation and metrics
  - ● Clustering directory structure revamp and testing the workflow

  Ethan Rogers:
  - ● Initial implementation of DSPY
    - ○ Run initial tests on breadth and accuracy of dataset creation
    - ○ See if "Intelligence at the Edge of Chaos" can be implemented
  - ● Publish made documentation to the group repository

Summary of weekly advisor meeting

This week Kellan demoed his work on the documentation for the whole NeuroXCode Library. Arushi let us know that NeuroxCode will now be rebranded to CoCoA and pointed us to a specific directory structure in Box we need to mimic to ensure consistency. We have work for the following week lined up including completing the Labeling pipeline, getting the integrated pipeline in a colab notebook for reference, completing the Hyperbolic clustering and comparison of clustering evaluation results. Team plans to have all of the things discussed above completed by the next meeting, and few of the things like the Clustering evaluation results submitted early next week during the break.