SE 4920 STATUS REPORT 6
4/4/2025 - 4/17/2025

Group number: sdmay25-30

Project title: *Explainable AI for source code applications*

Client &/Advisor: *Arushi Sharma*

Team Members:
Manjul Balayar
Sam Frost
Akhilesh Nevatia
Ethan Rogers
Rayne Wilde

Period Summary

Over the past period, we focused on both finalizing core system components and expanding our project's usability through improved visualization and deployment workflows. We set up the UI framework for our web application's visualizations and implemented interactive elements such as a dendrogram and sunburst diagram, along with backend endpoints for layer selection and cluster navigation. On the documentation side, we completed major updates to the design document, including a new section that outlines the architecture and function of the web application. We also advanced the hyperbolic clustering training process, addressing compute limitations by preparing to migrate our workflow from Google Colab to the Nova HPC cluster. Additionally, we developed a CI/CD pipeline to automate the containerized deployment of the web application and enhanced supporting documentation to ensure clarity and maintainability moving forward.

Last period accomplishments

Manjul Balayar:
- Setup the UI framework for the visualizations in webapp
- Setup up endpoints for layer selection and cluster navigation
- Implemented dendrogram and sunburst

Sam Frost:
- Finished updated sections for design document
- Mapped out section for web app in document

Akhilesh Nevatia:

- Worked on training script improvements for hyperbolic clustering
- Ran into compute issues on collab, switching to nova cluster, requested for access

Ethan Rogers:
- Finished first iteration of phrase level extraction using tree-sitter
- Allows for functions, classes, loops, and expressions to be captured
- Provide all necessary parsing info to activations extraction function

Rayne Wilde:
- Developed CI/CD pipeline for auto build / deployment of webapp container
- Worked on documentation

- **Pending issues**

Manjul Balayar: N/A
Sam Frost: N/A
Akhilesh Nevatia: N/A
Ethan Rogers: N/A
Rayne Wilde: N/A

Individual Time Contributions

Name	Hours This Period	Total Hours
Manjul Balayar	7	21
Sam Frost	8	29
Akhilesh Nevatia	5	34
Ethan Rogers	10	30
Rayne Wilde	5	26

Plans for the upcoming period

Manjul Balayar:
- Work on linkage matrix transformation for visualizations
- Implement different chart types for webapp

Sam Frost:

- Continue polishing design document
- Finish sections for web app

Akhilesh Nevatia:

- Get familiar with using nova
- Using nova access test the optimized training script and run the pytorch model on nova cluster
- Ideally have results for any one layer of hyperbolic embeddings to evaluate performance of current changes made to forked hyPHC repo and present to Arushi

Ethan Rogers:

- Finalizing phrase level activations
- Properly parsing phrase hierarchies for focus on specific token subgroups
- Adding on to existing activation extraction with per-string inputs for on-the-fly computation of code snippets

Rayne Wilde:

- Finish Hosting Issues for webapp
- Deploy updates to repo layout and design
- Deploy static webpages for information

Summary of weekly advisor meetings

In our most recent advisors meetings, we presented progress on both the backend and frontend components of the project. We demonstrated the initial version of the web application's visualizations, including the dendrogram and sunburst, and discussed our approach to interactive layer and cluster navigation. We also reviewed the updated design document and outlined the planned section for detailing the web app. Our advisor provided feedback on the UI functionality and supported our decision to transition hyperbolic clustering training to the Nova HPC cluster due to Colab limitations.