SE 4920 STATUS REPORT 4
2/28/2025 - 3/13/2025

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

This period, we accomplished several things. We addressed UI bugs caused by dash formatting and researched deployment options for the web app. Efforts also focused on adapting and running the hyPHC repository locally, modifying the training script and PyTorch model to process custom embeddings. Temporary outputs were generated, and visualization code was developed to illustrate cluster branches within connected components trained in hyperbolic space. Additionally, research was conducted on implementing phrase-level activations for improved explainability, exploring architectural modifications to support this functionality, and collaborating on a small code integration.

Last period accomplishments

Manjul Balayar:

-

Sam Frost:

- Fixed UI bugs caused by dash formatting
- Research how we will deploy web app

Akhilesh Nevatia:

- Adapted and ran the hyPHC repository locally, modifying the training script and in-built PyTorch model to process our custom embeddings (processed_activations).

- Generated temporary outputs and developed visualization code to illustrate cluster branches within the connected components extracted from the model trained in hyperbolic space.

Ethan Rogers:

- Looked into existing solutions for phrase level activations as opposed to token level activations, which is the primary means of explainability currently
- Sought different ways for the current architecture to give was to phrase-level capabilities
- Spoke with Akhilesh about a small code piece to be added to a function of mine

Rayne Wilde:

- Worked on containerizing the non-technical explanation pages and docs
- Finished library cleaning and repo restructuring

- **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		14
Sam Frost	4	17
Akhilesh Nevatia	6	25
Ethan Rogers	5	16
Rayne Wilde	4	15

Plans for the upcoming period

Manjul Balayar:

-

Sam Frost:

- Continue improving UI/UX
- Assist in deployment of web app

Akhilesh Nevatia:

- Run a more dense file to see better output and validate clusters being generated by the hyperbolic clustering approach
- Better visualization and labeling of clusters
- Documentation and deciding how we may want to integrate the same into our library.

Ethan Rogers:

- Peruse the Colab shared by Arushi for phrase level activations
- Look to provide decent results that are comprehensible using phrase-level activations, perhaps visualize the neurons through existing means
- Update Colab accordingly

Rayne Wilde:

- Library documentation
- Readme and wikis for continued development of the project

Summary of weekly advisor meetings

Our weekly advisor meetings were primarily focused on showing the progress we have made with the web app, as well as the visualizations being generated. We are on pace to finish our deliverables in time for the faculty panel/end of semester presentations, so no major course correction is necessary. We also used the time to ask and answer technical questions related to our work.