E E/Cpr E/S E 491 Weekly Report 3

Intelligent Code Editor

Client & Advisor: Ali Jannesari
sdmay20-46
John Jago – Software Lead
Keaton Johnson – Systems Lead
Jon Novak – Machine Learning Lead
Matthew Orth – Meeting Facilitator
Garet Phelps – Report Manager
Isaac Spanier – Test Lead

Weekly summary

This week was focused on the beginning stages of figuring out how we are going to train the model for the NLP. OpenNMT-py was set up, and trained for the first time, to get a feel for how that works. The UI has been started, and will be more or less finished by the end of the next sprint. Research was done on how we are going to have to create our dataset, if we need to do that.

Past week accomplishments

John Jago
- Implemented functionality in the IntelliJ plugin (in Hung's repo) to replace a selected line of natural language with code that comes from a remote OpenNMT-py server.
- Gained a better idea of what the first iteration of our UI might look like based on the data that the OpenNMT-py server provides and the anyCode project mentioned in the paper "Synthesizing Java Expressions from Free-Form Queries".

Keaton Johnson
- Researched best way to obtain and structure natural language to code dataset
  - Looked into line by line translation vs. comment based translation

Jon Novak
- Researching dataset generation
- Shifting from machine learning to help with the dataset

Matthew Orth
- Research AnyCode format
  - Research potential ways to structure our natural language input
  - Determine different architectures and strategies for translation
- Research Dataset Creation
  - Researched methods for automatically generating a training dataset

- ○ Learned usefulness of manual training in generation process
- ○ Found various resources and datasets to train on
- ● OpenNMT-py Installation and Training on GPU server:
    - ○ Install and setup OpenNMT-py on the GPU server
    - ○ Train the model on the Python code corpora
        - ■ Work through GPU memory errors

Garet Phelps
- ● Researched the creation of a dataset for this project
    - ○ Looked into existing datasets, not much clean data
    - ○ May have to just make our own
        - ■ If we create our own dataset, it would be beneficial to work on one concept at a time, this first attempt at the dataset is just calling the print method.

Isaac Spanier
- ● Researched the creation of the dataset for this project
    - ○ Continued digging into natural language datasets, none that are good for us
    - ○ Creating our own dataset
        - ■ BigQuery seems to a great solution for creating a dataset from scratch
        - ■ Also also us to build off of public datasets

Individual contributions

Name	Contributions	Hours this week	Hours cumulative
John Jago	● Added the line replacement functionality from the proof of concept plugin into the actual plugin project. ● Updated the call to get the translation from an external resource, keeping in mind that we are thinking of using OpenNMT-py ● Wrote unit tests for existing functionality. ● Read the paper "Synthesizing Java Expressions from Free-Form Queries".	2	8
Keaton Johnson	● Researched best way to obtain and structure natural language to code dataset	2	8

Jon Novak	● Assisting with dataset generation research	3	8
Matthew Orth	● Research AnyCode format ● Dataset Generation Research ● OpenNMT-py Installation and Training on GPU server	7	43
Garet Phelps	● Researched the creation of a dataset for this project	2	9
Isaac Spanier	● Researched the creation of the dataset for this project	2	8

Plans for the upcoming week

John Jago
● Create the basic UI functionality (translation button, ordered list, OpenNMT-py integration)

Keaton Johnson
● Research methods for creating a dataset for the Java print method (System.out.println())

Jon Novak
● Research methods for creating a dataset for the Java print method (System.out.println())

Matthew Orth
● Research implementation optimization and test effectiveness on dataset

Garet Phelps:
● Research methods for creating a dataset for the Java print method (System.out.println())

Isaac Spanier
● Research methods for creating a dataset for the Java print method (System.out.println())

Summary of weekly client/advisor meeting

Our regularly scheduled meeting with Dr. Jannessari happened on 10/11/19 at 4pm. All parties were present. We discussed how we should go forward with the UI, and what our approach should be for creating the dataset. We are moving forward into the next sprint with the following goals:
● Complete the base idea for the UI.
● Research/Create a small dataset for print statements
● Research how to optimize the issues with OpenNMT-py, and try the dataset we create.