E E/Cpr E/S E 491 Weekly Report 7

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 implementing placeholder names for variables. John implemented detecting variables in the plugin, and we also made some new entries for the dataset that use these generic names.

Past week accomplishments

John Jago

- Dummy names for classes, methods (plugin)
    - Investigated what would be the most natural way to call methods from classes when writing English statements
    - Implemented preprocessing that replaces class names with myClass1, myClass2, etc. and method names with myMethod1, myMethod2, etc.
    - Added logic to replace original class and method names
    - Refactored much of the replacement action code for readability
- Misc. improvements to plugin
    - Details of the OpenNMT-py server, such as the IP address and model ID, are now read in from a properties file instead of being hardcoded
    - Added information about our project to the README and created a README for the Intellij plugin

Keaton Johnson

- Attempted to expand a script that queried the github repo and code api.
    - Was able to return raw files from github where the provided phrase appears.

- - ○ Searches the top 10 java repos on github.
    - ○ Takes a long time to process and generates a lot of redundant files.
  - ● Looked into other ways to gather code from public repositories.
    - ○ Github seamed to be the only one with a robust API.

Jon Novak

- ● Analyzing method labeling with data mined from github

Matthew Orth

- ● OpenNMT-py Research:
  - ○ Learned and documented the OpenNMT-py architecture and how it works at a deeper level
  - ○ Learned more about RNN, LSTM, and the transformer architecture at a deeper level for eventual modifications of the architecture
  - ○ Determined that we can use PyTorch in combination with OpenNMT-py to eventually create a unsupervised learning classification/translation engine
- ● English Linguistics Research
  - ○ Researched and found several synonym APIs that we could possibly use to allow a more generalized matching for the input natural language statement
    - ■ Learned that a lot of this information is found in NLTK, so we can look into that during the next sprint
- ● Generic Variable, Method, and Class In Dataset
  - ○ Added thousands of data points that replace user-specified variable names with generic variable names to improve translation results

Garet Phelps

- ● Added some generic variable entries to the dataset for up to 50 distinct variables.
- ● Looked into some functionalities of NLTK
  - ○ Tokenization
  - ○ Tagging parts of speech
  - ○ Wordnet synsets

Isaac Spanier

- ● Began to Look and Test John's UI for errors and bugs
  - ○ Pulled Hung's Development Branch and worked off of my own branch from that.

Individual contributions

Name	Contributions	Hours this week	Hours cumulative
John Jago	● Dummy names for classes, methods ● Misc. improvements to plugin	6	30

Keaton Johnson	● Attempted to expand a script that queried the github repo and code api. ● Looked into other ways to gather code from public repositories.	2	19
Jon Novak	● Working on labeling data from mined sources	5	23
Matthew Orth	● OpenNMT-py Research ● English Linguistics Research ● Generic Variable, Method, and Class in Dataset	6	74
Garet Phelps	● Added some generic variable entries to the dataset ● Looked into some functionalities of NLTK	2	18
Isaac Spanier	● Testing the User Interface	2	14

Plans for the upcoming week

John Jago

● Continue testing our plugin's behavior using improved models and adjust preprocessing if necessary

Keaton Johnson

● Research methods for input generalization, dataset mining, and dataset labeling

Jon Novak

● Research methods for input generalization, dataset mining, and dataset labeling

Matthew Orth

● Research methods for input generalization, dataset mining, and dataset labeling
● Finalize running Java print dataset through OpenNMT-py and summarize results

Garet Phelps:

● Research methods for input generalization, dataset mining, and dataset labeling

Isaac Spanier

● Continue to Test and Write Tests for the UI component.

Summary of weekly client/advisor meeting

During this week's meeting, we met with Hung and Professor Janessari. We discussed our progress and initial results on the Java print dataset. We also decided that during our next sprint we will finalize our method for dataset generation, input generalization, and code translation labeling. This research will help us finalize our Design Document and plan for next semester's work.