Weekly report

Fan Zhang
a1738078

•What you have done for the project during the week individually

1. Did research and figured out how to execute the codesensor tool and published the usable version into my branch in github

2. Created the first data preprocessing example (c file to AST)

3. I have focused on the data preprocessing part and the experience setup part of the target paper and I found they replicated the setting of another paper ” POSTER: Vulnerability Discovery with Function Representation Learning from Unlabeled Projects”. I went through this paper and figured out what settings have been replicated by the target paper.

•What you plan to do for the project next week individually

There are four main functions to replicate this whole project and I am going to figure out how do these functions work and replicate them.

1. ProcessCFilesWithCodeSensor.py file is for invoking the CodeSensor to parse functions to ASTs in serialized format (for detail information and usage of CodeSensor, please visit the author's blog: http://codeexploration.blogspot.com.au/ for more details).
2. ProcessRawASTs_DFT.py file is to process the output of ProcessCFilesWithCodeSensor.py and convert the serialized ASTs to textual vectors.
3. BlurProjectSpecific.py file is to blur the project specific content and convert the textual vectors (the output of ProcessRawASTs_DFT.py) to numeric vectors which can be used as the input of ML algorithms.
4. LSTM.py file contains the Python code sample for implementing LSTM network based on Keras with Tensorflow backend.

•List the issue you encounter and give reasons if applicable

1. Some reference papers need to pay to get access and our university library does not have the resources. So I searched some websites to get the resources. (sci hub…)
2. There are some materials that teach us how to use the tools but are outdated. So I need to spend more time to do some researches.