COMP 495 Mentored Research Contract

STUDENT INFORMATION

Name Andrew Li

RESEARCH INFORMATION

Project description:

Bricks is a software system that has been used for numerous semesters in support of teaching intro programming. Students write and submit code to Bricks and have it graded immediately for correctness. Current level of feedback is simply to show students if it was correct or not and indicate what input data values caused erroneous output, leaving the student to figure out where the error(s) might be in the source code. With live TA support, a student could be guided to probable areas of code to work on. With the automated grading, we don't have that level of sophistication in the feedback. This project seeks to increase the sophistication of feedback to a student who writes an incorrect program.

After 8-10 semesters of using Bricks, the instructor has collected a huge number of incorrect programs. This project will seek to use ML techniques to find clusters in those incorrect programs (meaning hopefully members of a cluster all have the same error), and also produce a method for characterizing a new program (when erroneous) as belonging to a particular error cluster. Then the students submitting the newly erroneous program can receive a message more similar to what a live TA would give them.

The project consists of identifying appropriate ML techniques for clustering our data, developing a categorization for new errors, and place in writing the resulting methods into the existing Bricks software system for further experimentation.

Meeting requirements:

Once a week for 1 hour (Tuesdays 1PM)

Reading assignments:

Intro materials to Machine Learning techniques (survey articles and overviews)
Details on specific ML techniques that apply to characterization of program source code

Written assignments:

Generate a report on ML techniques that characterize program source code
Write a presentation on characterizing several ML packages with recommendations for which ones are appropriate for our use case
Experimental reports on application of chosen tool package to the student program data from Bricks learning environment
We hope to produce and submit a conference paper from this work

Software or hardware deliverables: (The deliverables should include the approximate scope and depth of work.)

Numerous small experimental programs to develop ML techniques in the context of finding errors in student programs
Modify Bricks environment code to make use of the results of these experiments.

Other assignments (e.g., presentations):
Final powerpoint presentation summarizing semester's work and experimentation

Assessment Plan:
Weekly meetings will assess the work done on a high pass/pass/low pass basis as well as plan for the next week.

Retention: This contract is to be retained for a minimum of four years. Last update: 30 August 2018