Journal Report 9
11/11/19-11/18/19
Bryan Lu
Computer Systems Research Lab
Period 2, White

Daily Log

Monday, November 11

Extracted problems from `https://imogeometry.blogspot.com`, a curated website for all olympiad geometry problems on AoPS in the Contest Collections, and performed some string parsing to get rid of unicode characters.

Tuesday, November 12

Randomly generated problems from China, Turkey, and Iranian Team Selection Tests (TSTs), and from the ELMO (a MOP contest), and picked 40 triangle geometry problems from this set of problems.

Thursday, November 14

Began assigning all of the valid relations associated with the the selected problems on Tuesday, and adjusted `lexicon.txt` and `finalrelations.txt` to accomodate new relations I hadn't anticipated.

Timeline

Date	Goal	Met
10/28	Begin writing code to create a log-linear classifier using `scikit`, and finalize the inputs needed for the algorithm.	Yes, I have all of the features I need per sentence.
11/4	Create the log-linear classifier/learning algorithm and the training data, and begin testing.	I've started, but I've realized this is a rather ambitious task because I still need to finish creating my training data.
11/11	Complete a set of about 40 problems to serve as my training data set, with the correct relations.	I have the problems, but putting in the correct relations for these problems is a lot of work.
11/18	Build the model with `scikit`, tweaking previous steps as needed, and finish the necessary test cases.	N/A
11/25	Test and train the logistic model, and see if any methods could be added to improve accuracy.	N/A
Winter Goal	Be able to output a set of possible literals (statements) based on detected relations in the problem.	N/A

Reflection

This week, my lack of test cases from first quarter started to really impact my work this week. Luckily, I sent out a forum post to the community a while ago, asking if anyone knew of another compendium of olympiad geometry problems. Sure enough, the community pulled through and some user on AoPS has been doing exactly that! I took the liberty of ripping some of the problems from his blog (but I'm not sure how I should eventually cite this source...)

This still doesn't account for the fact that I still have to create all of my test cases myself – each problem takes a decent amount of time to complete. Here's one of the problems that I'll be testing with, with the correct relations I have associated with the problem:

Problem: Let Gamma be the circumcircle of acute triangle ABC. Points D and E are on segments AB and AC respectively such that AD = AE. The perpendicular bisectors of BD and CE intersect minor arcs AB and AC of Gamma at points F and G respectively. Prove that lines DE and FG are either parallel or they are the same line.

Relations: IsTriangle(ABC) IsCircle(Gamma) IsCircumcircle(Gamma, ABC) IsSegment(AB) IsSegment(AC) IsPoint(D) IsPoint(E) Equals(AD, AE) Collinear(D, A, B) Collinear(E, A, C) IsLine(DE) IsLine(FG) IsParallel(DE, FG)

This will probably expand to include more information about the perpendicular bisectors being accepted as well.

I'm fairly sure that I can finish these test cases this week, but considering the volume of tests associated with each problem, I might not need all of these to train the model. After all, I need to consider all possible pairs of explicit/abstract variables when testing, as long as they're close enough in the statement/in the same sentence, which is still a decent number of pairs per problem.

I am a bit concerned with my progress towards my winter goal – I think it's reasonable, but if I encounter another major challenge to my process in the building and testing of the logistic model I may not be able finish. Fingers crossed that everything goes smoothly.