

Journal Report 6

10/7/19-10/14/19

Bryan Lu

Computer Systems Research Lab

Period 2, White

Daily Log

Monday, October 7

I quickly added Greek-letter functionality to the part of my code that can identify explicit variables, and reread the original paper to discern what kinds of properties the language needed.

Tuesday, October 8

I figured out what the relations actually were in my problem, and what properties they needed to have, and I started to create a dictionary with detected words mapping to objects and relations.

Thursday, October 10

I created `lexicon.txt`, a file with words to detect for and corresponding objects and relations, and continued to fill in the corresponding relations and objects for various words.

Timeline

9/23	Filter posts that are not standard olympiad geometry problems, and construct a standard lexicon of keywords to look for in a problem both as objects and as relations.	Yes, I created two separate files for various geometrical objects and relations that I found that appeared in my data set.
9/30	Write code that creates a graph structure corresponding to the problem statement, with objects as nodes and relations as edges.	My code definitely is able to identify all the nodes/concepts and what kinds of relations are present in the problem, but not much more.
10/7	Create a logical language to give more structure and properties to the objects and relations detected in the problem.	Not quite the goal I needed to meet, but I figured out the path I need to take going forward and started creating the dictionary I needed to have for my problem.
10/14	Finalize the object identification associated with <code>lexicon.txt</code> , and figure out how to extract the necessary data to pass to the log-linear classifier later.	N/A
10/21	Research how to write code to create a log-linear classifier using <code>scikit</code> , and create some sort of working learning algorithm.	N/A

Reflection

I finally figured out the approach I needed to take to making progress in my project this week. I've begun creating a system for myself – for every word that appears in the text that is geometrically relevant, I'm going to associate with it all the possible objects and relations that can go with it. The objects are the strings that start with a lowercase letter, and the relations look like Java methods and start with an uppercase letter:

```
point point IsPoint
circle circle IsCircle
line line IsLine
ray ray IsRay
segment segment IsSegment
arc arc IsArc
triangle triangle IsTriangle
angle angle IsAngle HasRay
quadrilateral quadrilateral IsQuadrilateral
length LengthOf
polygon polygon IsPolygon

pentagon pentagon polygon IsPentagon IsPolygon
hexagon polygon hexagon
radius RadiusOf
side segment IsSegment
square square IsSquare
rectangle rectangle IsRectangle
trapezoid trapezoid IsTrapezoid
vertex point IsPoint
tangent TangentTo
```

This is admittedly incomplete, but this is what I'm looking for, but for all geometrically-associated words.

I also began to think about creating training data for my machine learning model, which would require me to list what explicit relations are valid for each problem. Looking over my problems/data set once more this week, I found that a lot of them are fairly arbitrary in their configurations, which would make my job to confidently determine which relations are correct for each problem much harder. I'll need to keep this in mind as the end goal for this stage of the process.