CS 6390 Project Declaration

Jared Amen (u1228357)

February 23, 2021

1 Team Name

My team's name will be "Persuasion Extraction".

2 Team Member

I will be the sole team member of this project.

3 Task

I got my idea for this task from this SemEval task . The task is on the detection of persuasive techniques in texts and images. These texts and images are most appropriately expressed as memes, which are common in their use for persuasion of agenda. Since this is an *information extraction* project, it will be specifically focusing on the **second subtask** of the original task. From the project page:

> **Subtask 2:** Given only the "textual content" of a meme, identify which of the 20 techniques are used in it together with the span(s) of text covered by each technique. This is a multilabel sequence tagging task. The task is the combination of the two subtasks of the SemEval 2020 task 11 on "detecting propaganda techniques in news articles."

My project will be able to complete this subtask to a relatively successful extent.

4 Data

The test dataset will come from the corpus for this task, provided for SEMEVAL-2021: it can be found here.

The gold (training) dataset will also come from the same corpus for this task: it can be found here.

There is also a "dev" dataset, which will come from the same corpus for this task, and will be utilized in my project: it can be found here.

I have mild concerns about the size of the dataset. Similar datasets were used for a related task, so I can expect that my system's performance will be likely in the scope of the related project's leaderboard unless I expand my training dataset.

5 Evaluation

For this project, I will utilize a couple methods to help evaluate my program. First, the given SemEval task provides a scorer which can be found here. I will be utilizing this script to assist with the overall score of my program. Furthermore, I will be developing my own scorer which will give better detail to scoring than that of the built-in counterpart. In the end, my scorer will log F-Score, Precision, and Recall, and will also log notable differences between expected output and actual output.