1 Background

In public debates, there are certain well-defined tactics and rhetorical strategies that participants employ to gain an advantage over opponents. Debates of this nature are commonly used in the political sphere to compare candidates before the general public. In the U.S., there are no debates more significant than those of the presidential election. Under the guidance of Professor Dragomir Radev, this project aims to use natural language processing (NLP) and machine learning (ML) techniques to analyse the rhetoric employed by participants of U.S. presidential debates. NLP is a subset of artificial intelligence and computational linguistics that concerns itself with the analysis of text and speech through algorithmic and statistical methods. There has already been development in the field of NLP that conducts sentiment analysis of passages and the rhetorical impact of speeches, and NLP is a field that is growing in scope and significance.

The project will identify and categorise the different rhetorical strategies that are used by the candidates, and examine the reactions of their opponents to each of the strategies. The initial focus of the study will be on the most recent cycle of U.S. elections, but the process should be generalisable to all historical U.S. presidential debates.
2 Technical Specifications

This project will mainly utilise Python to implement NLP and ML procedures to analyse the use of rhetorical strategies in political speeches.

The project will begin with finding accurate and verbatim transcripts of the televised debates. Fortunately, major U.S. political speaking events are well-documented and the transcripts are freely available to the public (see appendix). We will then need to find a corpus that relates to U.S. political debates and debating strategies. A prospective corpus that may be used is CORPS, the CORpus of tagged Political Speeches, that was designed by Strapparava and Stock to examine persuasive language in political debates.

We will also find a database of common debating strategies that we can use to tag the elements of the speech. Public speaking and debate are activities that are common throughout the U.S. and there are various resources which provide guidance and classification on common debate strategies. Using those elements, we can tag elements of candidates’ speeches that correspond to certain demarcated speaking strategies and, from the context, examine the opposition’s reactions to these strategies. We can then determine the favoured debating strategies and weaknesses of particular candidates.

The project will run on an input that consists of the transcript of the U.S. presidential debates and a pre-tagged corpus. Initially, we will seek to analyse one-on-one debates, but then we can switch focus to multi-candidate debates, the kind of which is prevalent in party primary debates prior to awarding nominations.

3 Deliverable

The output of this project will consist of a report that contains the analysis of a recent U.S. presidential debate. The output will also consist of a corpus that has tagged bodies of speeches to their respective debating strategies as well as source code to allow for this analysis to be replicated for other U.S. political debates.
4 Prospective timeline

- From 02/08 to early March:
  - Do research on relevant NLP techniques, collect data for corpora and inputs

- From March to Early April:
  - Begin writing program and extracting results

- From April onwards
  - Conduct analysis of results and begin preparing report

- By 04/28:
  - Deliver final project output along with project report
5 Appendix


- The paper in which Strapparava and Stock describes their CORPS political data corpus can be found here: [http://dx.doi.org/10.1080/19331680802149616](http://dx.doi.org/10.1080/19331680802149616).