CS372 Project Proposal:
How Does a President Talk Their Economy into Success?

Youngin Lee 20170473_Youngin Lee
Haksun Son 20170340_Haksun Son
Sejun An 20160351_Sejun An
Akkanit Pornpattanadul 20180815_Akkanit Pornpattanadul

1 Problem Description

Presidential speeches are an important indicator of a government's policies and what they think is important and should be focused on. However, it is hard to quantify the effort a president is willing to put into a certain topic just by listening to each speech.

This project aims to build a model that can quantify the emphasis on a certain topic, then affirm that this emphasis is carried onto real life. Specifically we aim to examine the emphasis on the American economy within presidential speeches, then compare the output of the model to real-life economic indicators such as growth rate, GDP and trade. Economics is a good field to observe, as there are many indicators that measure the growth.

2 Related Work

A closely related field for our project would be economic event detection in text. Economic event detection is applied to news articles, in many cases, to predict and analyze stock markets. Traditional methods are rule based and involve hand-crafted rules and semantic parsing, as Hogenboom et al. (2010) has done in their pipeline. In more recent studies, machine learning is used more widely to eliminate the labor-heavy stage of constructing rules manually. Approaches by Jacobs et al. (2018) and Lefever and Hoste (2016) construct the problem of event detection as classification problems and use supervised machine learning approaches, of which SVM has shown the best performance.

3 Dataset

Our dataset will be comprised of two main components; the corpus of presidential speeches and the economic indicators.

The transcripts of presidential speeches for U.S. presidents can be accessed on the American Presidency Project website, managed by Woolley and Peters (2018), which comprehensively archives documents and transcripts related to U.S. presidents. Thus, we can also access other types of related document should the need for a bigger corpora arise.

The reports for the economic indicators of the American Economy can be accessed at the World Bank Open Data website, managed by World Bank Organization (2020), in excel format. This website provides various economic indicators such as GDP, imports and inflation for many countries from 1960 to 2018.

4 Methodology

Our main approach to quantifying the emphasis on economics in our corpus will be investigating the frequency of words related to the economy. A basic approach would be drawing from a pre-constructed lexicon and counting the frequency of those words; in the case where these measures are not sufficient, we can use more sophisticated measures. Common rule-based method for such detection involve semantic parsing and using WordNet as lexical semantics source, while methods using machine learning include learning keywords using support vector machines.

After we construct a model to quantify the emphasis, we can validate this number by performing linear regression against the economic indicators in the second part of our dataset. To increase the accuracy of the linear regression, we can use various solvers such as L-BFGS and SAG. If the size of the corpora is too small, k-fold cross-validation can ensure the integrity of the results.

We also plan to conduct an experiment to compare the Republican Party and the Democratic Party in terms of the emphasis on economics in their speeches. We plan to use linear regression separately on the two parties and observe any differences from the original models. We also plan to look for differences in the distribution, as in the mean and the variance, in the two parties.

5 Evaluation Plan

We can evaluate our results by the goodness of fit between our metric and economic indicators, measured by the p-values and the R^2 values.

References

Frederik Hogenboom, Alexander Hogenboom, Flavius Frasincar, Uzay Kaymak, Otto van der Meer, Kim Schouten, and Damir Vandic. 2010. Speed: A semantics-based pipeline for economic event detection. In *Conceptual Modeling – ER 2010*, pages 452–457, Berlin, Heidelberg. Springer Berlin Heidelberg.

Gilles Jacobs, Els Lefever, and Véronique Hoste. 2018. Economic event detection in company-specific news text. In *Proceedings of the First Workshop on Economics and Natural Language Processing*, pages 1–10, Melbourne, Australia. Association for Computational Linguistics.

Els Lefever and Véronique Hoste. 2016. A classification-based approach to economic event detection in Dutch news text. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 330–335, Portorož, Slovenia. European Language Resources Association (ELRA).

J. Woolley and G. Peters. 2018. Welcome to the american presidency project: The american presidency project.

World Bank Organization. 2020. World bank open data.