CS5480: Deep Learning

Project Proposal

Project Title:

Identifying Indian Politician's Stance on National issues

Team members:

Rohan Tondulkar, Bedanta Das, Rajat Saini

Project Objective:

We see politicians using Twitter to express their views on various political issues. Politicians are scrutinized for everything they say on media. Despite tweets being short in length, they are quite vocal about their stance on a particular topic. Also politicians are held accountable for things said on social media. Thus it is important to understand their overall view/sentiment on a particular topic. Our objective is to understand the viewpoint of politicians on various national topics ranging from Demonetisation, G.S.T, black money, various scams, infrastructure projects etc. We aim to find if a politician is in agreement or disagreement on a particular topic/policy.

Proposed Methodology:

1. Dataset collection:

   There is no publicly available dataset for comments of Indian politicians on various topics. Thus we will be using Twitter API to collect tweets on various topics by Indian politicians. Tweets based on twitter handle of various politicians will be extracted. Thus there is no benchmark dataset for this research.

2. Data preparation:
   a. Data labeling:

Using our prior domain knowledge on various topics we will assign labels of 'Agreement' / 'Disagreement' to tweets based on the political party that particular politician belongs to.
Eg: Congress is in disagreement with demonetisation. Thus, all the tweets by Congress leaders on demonetisation will be labeled as 'Disagreement'.

This data will be split into training, validation and test set to analyse results.

b. Data preprocessing:

We will be only considering English tweets for now. Also various unnecessary content like urls, punctuations etc will be removed. Removal of stopwords, stemming etc actions will be applied.

c. Data representation:

Word2Vec format will be used to represent tweets. Also learned embeddings will be used for comparison between models.

3. Experiments:

Use various sequential modelling techniques namely, Recurrent Neural Network Architectures (RNNs) like Long Short-Term Memory (LSTMs), Gated Recurrent Units (GRUs), etc. Also given time, we will also apply attention mechanism.

Comparison of various models will be performed against a standard metric.

4. Results

a. Standard metrics like F1 score, accuracy on the labelled test dataset will be used for comparing performance on various models.
b. Since there is no baseline models, we don't have an estimate on expected performance.

References:

[1] K. Johnson, D. Goldwasser, ‘Identifying Stance by Analyzing Political Discourse on Twitter’, NLP Workshop, EMNLP, 2016.

[2] K. Johnson, D, Jin, D. Goldwasser, ‘Modelling of Political Discourse Framing on Twitter’, Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017).

[3] M. Iyyer, P. Enns, J. Boyd-Graber, P. Resnik, ‘Political Ideology Detection Using Recursive Neural Networks’, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.

[4] D. Maynard, A. Funk, ‘Automatic detection of political opinions in Tweets’, Extended Semantic Web Conference, 2011.