Evaluating Opinion in Romanian Parliamentary Debates

Problem

Over the past decade Romania has seen a surge of convictions for crimes of administrative and political corruption (e.g., bribery): 3131 people from 2010 to 2014, among which top-tier politicians, civil servants, and businesspeople (Romanian Academic Society 2015). In comparison, neighbouring Bulgaria has zero convictions. What has led to such strong judicial activity in Romania? I have interviewed numerous players in Romanian anti-corruption (public prosecutors, judges, defence lawyers, journalists, civil society activists, and former ministers), and most believe that there has been a ratcheting effect in judicial activity: at first only a few, minor figures were convicted, then more important and central figures, then more. By the time politicians awoke to the danger they all faced, the process had gained already significant momentum.

If this is true, then I should find evidence of this in the public discourse employed by Romanian politicians. Over the past decade, I should find politicians increasingly talking about certain institutions, (e.g., the National Anti-Corruption Directorate) and individuals (e.g., specific judges). Additionally, if it is true that all politicians have started to feel the heat, then politicians from all parties should be using increasingly negative language to describe these institutions and people.

If I do find robust evidence of this process, I can start comparing political speeches to other public discourse (e.g., newspaper articles and press releases from NGOs) and see how different institutions have contributed to the overall discourse over time. Ultimately, if my findings reflect the hypotheses of domain experts (i.e., my interviewees), it implies that I can treat dynamics in public discourse as a proxy for political dynamics more broadly, and thus use text analysis to study large-scale shifts in institutional politics that go beyond the expertise of experts in any one institution (e.g., lawyers, journalists, parliamentarians).

Method

I will use text-categorisation to determine whether, in any given document (a transcript of a political speech) the author (politician) expresses positive or negative sentiment toward some entity (another politician, a judge, a state institution, etc.). First, I will extract named entities for each document, and count how many times an entity is mentioned. Then, for each sentence in which the entity is mentioned, a classifier will determine whether the entity is refereed to in negative (-1), positive (1), or neutral terms (0). Finally, for each document I will sum the scores for each entity to create a ranking of entities from most to least positive. Ultimately I will compare rankings across speeches/time to see how a given actor's appraisal of different entities has evolved.

I will hand-code a number of these documents, and will train both Naive Bayes and SVM classifiers, comparing the performance of both against the hand-coded gold standard. My features will be a Romanian-language lexicon of positive and negative terms, unigrams and bigrams.

Data

The data are Romanian parliamentary debates occasioned by a request from the prosecutor's office to lift the judicial immunity of a sitting member of parliament. In Romania no formal judicial proceedings can be initiated against parliamentarians until their immunity is lifted, an act that can

---

1 (Iftene et al. 2011)
2 (Bădică, Colhon, and Şendre 2014)
only by done by a majority vote of their colleagues. Unlike legislative debates, speeches occasioned by a request to lift immunity are likely to contain less formal, legal jargon, be more straightforward, and to display more emotional opinion. They are therefore a good place to begin training a classifier.

**Evaluation**

I will compare the classifiers against the gold standard in terms of accuracy, recall, precision, and the F1 score, in addition to whether or not they give substantively meaningful results.

