TopFish: Topic-Based Analysis of Political Position in US Electoral Campaigns

Federico Nanni[1], Cäcilia Zirn[1], Goran Glavaš[1,3], Jason Eichorst[2], Simone Paolo Ponzetto[1]
[1] Data and Web Science Group, University of Mannheim
B6 26, DE-68161 Mannheim, Germany
[2] Collaborative Research Center SFB 884, University of Mannheim
L13 15-17, DE-68161 Mannheim, Germany
[3] Text Analysis and Knowledge Engineering Lab, University of Zagreb
Unska 3, HR-10000 Zagreb, Croatia
{federico,caecilia,goran,simone}@informatik.uni-mannheim.de
eichorst@uni-mannheim.de

Abstract

In this paper we present TopFish, a multi-level computational method that integrates topic detection and political scaling and shows its applicability for a temporal aspect analysis of political campaigns (pre-primary elections, primary elections, and general elections). It enables researchers to perform a range of multidimensional empirical analyses, ultimately allowing them to better understand how candidates position themselves during elections, with respect to a specific topic. The approach has been employed and tested on speeches from the 2008, 2012, and the (ongoing) 2016 US presidential campaigns.

1 Introduction

The competition for votes in US elections provides an opportunity for candidates to communicate their positions. Evidence suggests that campaign statements are designed to inform voters of the types of policy a candidate will pursue in legislative (Ringquist and Dasse, 2004) and executive offices (Marschall and McKee, 2002).
Converging on a position, however, is a complicated process. Candidates must not only satisfy the interests of voters in the general election, but also win in primary elections where party identification is shared among candidates and support is ultimately won from informal organizations within the party (Masket, 2009).

Adequately capturing this process, namely the development of candidates' positions and reputations in campaigns is a challenging empirical problem that relies on processing large amounts of political texts. Significant advancements in quantitative methods from the field of natural language processing (NLP) have enabled coarse-grained analyses of texts produced in presidential campaigns (Medzihorsky et al., 2014; Sim et al., 2013; Gross et al., 2013). However, in all of these works positions are analysed based on the content of the whole documents. Put differenly, there is still an empirical gap with respect to fine-grained analysis of politicians positions towards particular topics and how these topically-bounded positions change over time.

In this paper, we present *TopFish*, a computational method that (1) identifies parts of public campaign speeches that correspond to topics of interest and (2) determines candidates positions specifically towards each of these topics. TopFish combines a topical classifier following the idea of our previous work on party manifesto classification (Zirn et al., 2016) and the Wordfish tool (Slapin and Proksch, 2008), which is commonly used for quantitatively estimating candidate positions in political science analyses (Grimmer and Stewart, 2013).

In order to show why there is the need for a more fine-grained position analysis on topic level, we apply TopFish to speeches delivered in presidential election campaigns. In a qualitative analysis, we discuss how candidates' positions do not only vary with respect to topics, but how they also change in different phases of an election campaign. In other words, we show how some topic-based positions of some candidates change from pre-primaries, over primaries, to general election.

The approach we present is weakly-supervised because it depends on an appropriate topic-labeled dataset, yet it does not require any manual annotations for positions themselves. Therefore, it can be easily applied to other types of political texts

Proceedings of the International Conference on the Advances in Computational Analysis of Political Text (PolText 2016),
pages 61–67, Dubrovnik, Croatia, 14–16 July 2016.

such as online discussions or debate transcripts.

2 Related Work

During the last decade, there has been a consistent growth in application of natural language processing (NLP) methods in political science research (Grimmer and Stewart, 2013). Here we cover the most relevant lines of work.

Topic detection in political text. The detection of topics in political documents has been performed adopting unsupervised techniques such as latent semantic analyses (LSA) (Hofmann, 1999) and latent dirichlet allocations (LDA) (Blei et al., 2003) as well as supervised adaptations like Supervised LDA (sLDA) (Mcauliffe and Blei, 2008) and labeled LDA (lLDA) (Ramage et al., 2009). For example, (Quinn et al., 2010) present a method that estimates a hierarchical structure of topics in political discussions, while Balasubramanyan et al. (2012) describe an adaptation of sLDA for studying the topic-based polarization of debates in the US and Gottipati et al. (2013) explore the potential of Debatepedia for determining political topics and positions. Zirn and Stuckenschmidt (2014) propose a method for analyzing and comparing documents according to a set of predefined topics based on lLDA, while Nanni and Fabo (2016) combine entity linking (Rao et al., 2013) and labeled LDA in order to overcome the most common limitation of unsupervised topic modeling techniques, namely the interpretability of the results.

Fully supervised approaches for topic detection have been also performed (see for example Hillard et al. (2008)). However, as these solutions rely on expert knowledge for establishing in advance a set of relevant topics and on annotating a large set of training data, they generally are more time-consuming to build. In contrast, we show that for our approach a small set of annotated data is enough, and we explore the use of external annotated training sources.

Political position scaling. While there has been a long term interest in modelling ideological beliefs using automated systems (see for example Abelson and Carroll (1965)), only in recent years we have seen a growth of advanced computational techniques for performing the task. In 2003, Laver, Benoit and Garry presented Wordscores (Laver et al., 2003), a supervised approach that relies on a set of pre-defined reference texts to determine the position of political documents in space. Inspired by it, in 2008 Slapin and Proksch developed Wordfish (Slapin and Proksch, 2008), a completely unsupervised solution for scaling documents on a single dimension.

The techniques presented above analyse coarse-grained political positions on document level and do not fully exploit the potential of topic-based political scaling.

Text-based analyses of political campaigns. In the last decade, computer-based analysis of political campaigns has attracted the attention of journalists (Silver, 2012) and academics (Foot et al., 2003). Scharl and Weichselbraun (2008) studyied trends in political media coverage before and after the 2004 U.S. presidential election applying NLP methods. Recently, Prabhakaran et al. (2014) studied the topic dynamics of interactions during the 2012 Republican presidential primary debates. Transcriptions of speeches have been employed by Gross et al. (2013) adopting the method presented in (Sim et al., 2013) to study the US 2008 and 2012 campaigns and in particular to test the Etch-a-Sketch hypothesis[1]. We will address the same hypothesis in our qualitiative evaluation part in subsection 4.2.

3 Topic Detection and Scaling

In this section, we describe in detail the two steps of TopFish, which consists of identifying the topics in the speeches and separately scaling the topic-specific positions based on parts of text belonging to a particular topic of interest.

3.1 Identification of topics in speeches

In the first step, our goal is to identify the topics that are discussed in the collected candidate speeches. We decide to use the classification scheme developed by the Comparative Manifesto Project (Volkens et al., 2011), which distinguishes between seven topical domains: *External Relations*, *Freedom and Democracy*, *Political System*, *Economy*, *Welfare and Quality of Life*, *Fabric of Society* and *Social Groups*. We assume that those domains, which are used to capture all topics tackled in party election programs, also correspond to major coarse-grained topics of interest in electoral speeches.

[1] From Mitt Romney's own words: "I think you hit a reset button for the fall campaign [i.e., the general election]. Everything changes. Its almost like an Etch-a-Sketch. You can kind of shake it up and we start all over again."

In order to determine the topics addressed in a political speech, we follow the idea of a classification approach we introduced in Zirn et al. (2016). This classifier, initially designed to annotate topics in political manifestos, extends a local supervised topic classifier with predictions from topic-shift classifiers and topic distribution knowledge in a global optimization framework. The global optimization step, however, is helpful when applied to the manifestos, as they cover many different topics (potentially all seven) and require classification on sentence level. For the speeches, however, we choose to classify text at paragraph level because whole paragraphs most often belong to the same topic because politicians tend to express their arguments coherently. Moreover, as each speech generally focuses on a few specific topics (for example *External Relations* and *Economy*), and does not cover the entire spectrum of topics, we decided that the optimization step used for manifesto classification would be superfluous in this setting. We thus only apply on speeches a local supervised topic classifier, trained on manifestos, that combines lexical with semantic textual similarity features (Zirn et al., 2016).

We train this local classifier on two different datasets and compare their performance on a gold standard of speeches labeled on paragraph level.

Training set: manifestos. We train the classifier on party manifesto programs labeled on sentence level. A sub-part of the training set was annotated manually by human experts, the rest was labeled automatically with the method presented in (Zirn et al., 2016). The advantage of such a domain transfer approach is the fact that we need no manual topic annotations on speeches. The downside is, however, that the language of manifestos might differ from the language used in speeches. In the next section, we quantify the drop in performance due to the domain change.

Training set: annotated speeches. We manually annotated a small part of the presidential election campaign speeches on paragraph level with their categories. We train the above described system on this data and, in the next section, report the results. We explore whether investing human resources for annotating speeches pays off with more accurate classification results.

3.2 Position analysis

In order to determine the positions of politicians based on their speeches on a left-right spectrum, we adopt Wordfish (Slapin and Proksch, 2008), which is widely adopted for such tasks in political science research (Grimmer and Stewart, 2013). This method is designed to take documents as input and estimates their positions on a one-dimensional scale. Our goal is to determine fine-grained positions towards the topics contained in the speeches instead of the overall position of the whole speech. We therefore apply the classifier described in this section to identify the topics within a speech and divide a speech into subdocuments containing the text for a single topic only. Finally, we apply Wordfish to the subdocuments.

4 Evaluation

We first quantitatively assess the correctness of the topic classification on a small manually-labeled evaluation dataset of speeches. Then, in order to assess the quality of our fine-grained political scaling approach, we apply it to speeches of thee presidential election campaigns and do a qualitative analysis of the results.

Gold Standard Annotation We asked two scholars of political science to annotate a subset of 10 speeches from the US presidential election campaigns of 2008, 2012 and 2016. The set comprises samples of seven candidates. Our annotators labeled each of the 779 selected paragraphs one of the 7 topical classes listed in subsection 3.1. The inter-annotator agreement across the seven topical classes is $\kappa = 0.55$, which is only moderate and thus confirms the difficulty of the task.

4.1 Evaluation of Topic Classification

We compare three different settings to classify the topics in the speeches.

Baseline. As a baseline, we apply a Support Vector Machine (SVM) using a simple bag-of-words features on the gold standard performing 10-fold cross validation.

ClassySpeech. We apply the classifier described in 3.1 to our gold standard and perform 10-fold cross validation. We refer to this model as *ClassySpeech* in the following.

ClassyMan. We train the classifier described in 3.1 in a semi-supervised fashion on a set of party

Model	F_1
ClassyMan	36.2
Standard SVM	71.2
ClassySpeech	78.6

Table 1: Topic classification performance, micro F1-score, 10-fold CV (in %)

manifestos. We first train the local topic classifier model on six manually sentence-level labeled manifestos and then use the globally-optimized classifier (Zirn et al., 2016) to label the collection of 466 unlabeled manifestos from the Comparative Manifesto Project (Volkens et al., 2011). We then re-trained the local topic classifier on this set of 466 automatically labeled manifestos. In this setting we did not need to topically label any speeches. We apply the classifier trained on the manifestos resulting (from now on referred to as *ClassyMan*) to our gold standard set of speeches.

The results of the three models are shown in table 1. As it is evident from Table 1, the baseline performs quite well with an F_1-score of around 71% , re-confirming the already well-known efficiency of the simple bag-of-words-based supervised topic classification models. The drop in performance caused by the domain adaptation (i.e., the low performance of the model trained on manifestos) indicates that, even if the topics discussed in electoral manifestos and in political campaigns are the same, the language in which they are convened seems to be significantly different. Finally, the best performance is achieved by the ClassySpeech model, the local topic-classifier trained on a small set of manually labeled speeches. The fact that the ClassySpeech model drastically outperforms the ClassyMan model shows that having little of in-domain annotations (i.e., annotated speeches) matters more than having a lot of annotations on out-of-domain texts (i.e., manifestos).

4.2 Qualitative Analysis of Topic-Specific Positions

Election campaigns are a long and complex process that represents the essence of contemporary democracies. In the United States, the practice of selecting candidates for the presidential elections spans more than a year, being a major focus of American and international media. More specifically, in our work we identify three major phases in the presidential race: a) the pre-primaries, when politicians announce their candidacy for president and begin to establish their positions; b) the primaries: when candidates sharpen their profile in order to win the support of the party; and c) the presidential elections: when party nominees have to satisfy the interests of a spectrum of voters as large as possible.

Dataset preparation. After collecting speeches made by the most prominent Republican and Democrat candidates of the last three general elections (2008, 2012, 2016), we divided them in three temporal groups, namely: before primaries (i.e. before the 1st of January of the election year), primaries (between January and June of the election year) and elections (after June of the election year). Using the ClassySpeech model, we topically annotated all of the collected political speeches at paragraph level. Next, we grouped together all paragraph from the same topic and the same period (e.g. all text from all Barack Obama's primary campaign speeches labeled with topic *External Relations*).

Analysis. In the third step of the analysis we ran Wordfish on the collection of temporally and topically divided speeches. In order to understand the usefulness of our fine-grained analysis (i.e., the combination of the two dimensions – time and topic), we compared the its qualitative results with two different more coarse-grained studies. In the first study, we ran Wordfish on the entire speech collection of each candidate (i.e., without any temporal and topical slicing). In the second study we considered only the temporal dimension, i.e., we excluded the topical division.

As shown in Fig. 1, the two coarse-grained analyses do not add any new knowledge, by re-confirming already well known facts, such as the global position of candidates over the political spectrum and a common trend in political campaigns, namely the convergence to the center of the selected party candidates after the primary race (see McCain in particular).[2] In contrast, the fine-grained temporally and topically sliced analysis proposed in our study enables to dig deeper into the candidate's process of converging on a specific position[3]. As a matter of fact, it presents

[2] To know more about the Etch-a-Sketch Hypothesis and how to automatically detect it, see Gross et al. (2013)

[3] Other analyses can be found at: `https://federiconanni.com/topfish`

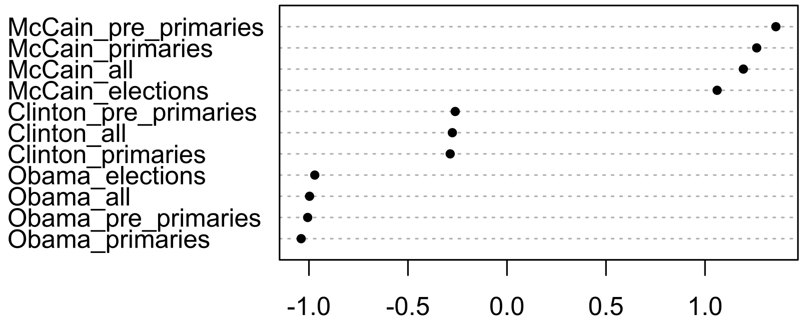

Figure 1: Coarse-grained comparative analyses, using Wordfish.

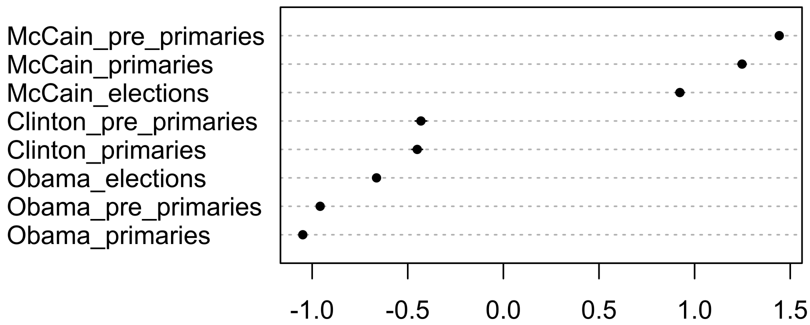

Figure 2: Wordfish position estimates regarding the topic *External Relations*.

a more clear understanding on how candidates have been positioning themselves regarding different relevant political issues, such as *External Relations* (see Fig. 2) and *Welfare and Quality of Life* (see Fig. 3). Additionally, it highlights interesting variations on the established idea of positioning during political campaigns (e.g. the shift to-the-left of Barack Obama presented in Fig. 3) which are completely ignored by a coarse-grained overview on the race.

5 Conclusion

In this paper we presented TopFish, a multilevel computational approach that combines topic detection and political scaling with temporal aspects of political campaigns (pre-primary election, primary election, and general election). We show how this solution enables researchers to perform a range of multidimensional empirical analyses, ultimately allowing them to understand how candidates position themselves during the entire campaign race. The topic-detection method here adopted has been tested against two other solutions, showing its robustness. Additionally, the presented approach has been employed and tested on speeches from the 2008, 2012 and the ongoing 2016 US presidential campaigns, showing its usefulness for examining in a more fine-grained fashion how candidates determine their political space.

Acknowledgments

The authors thank the DFG for Funding under the SFB 884 Political Economy of Reforms C4 project.

References

Robert P Abelson and J Douglas Carroll. 1965. Computer simulation of individual belief systems. *The American Behavioral Scientist (pre-1986)*, 8(9):0_24.

Ramnath Balasubramanyan, William W Cohen, Douglas Pierce, and David P Redlawsk. 2012. Modeling polarizing topics: When do different political communities respond differently to the same news? In *ICWSM*.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

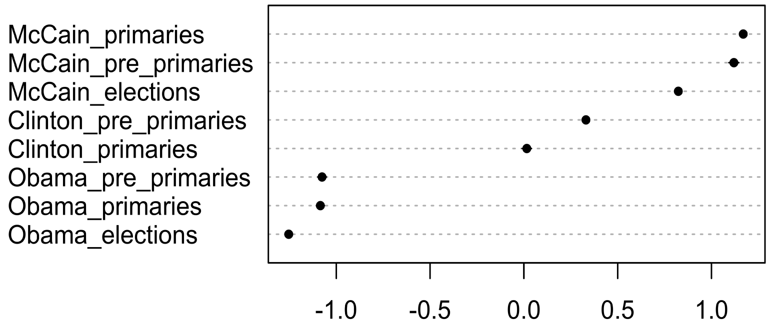

Figure 3: Wordfish position estimates regarding the topic *Welfare and Quality of Life*.

Kirsten Foot, Steven M Schneider, Meghan Dougherty, Michael Xenos, and Elena Larsen. 2003. Analyzing linking practices: Candidate sites in the 2002 us electoral web sphere. *Journal of Computer-Mediated Communication*, 8(4):0–0.

Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang, and Noah A. Smith. 2013. Learning topics and positions from Debatepedia. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1858–1868, Seattle, Washington, USA, October. Association for Computational Linguistics.

Justin Grimmer and Brandon M Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, page mps028.

Justin Gross, Brice Acree, Yanchuan Sim, and Noah A Smith. 2013. Testing the etch-a-sketch hypothesis: A computational analysis of mitt romney's ideological makeover during the 2012 primary vs. general elections. In *APSA 2013 Annual Meeting Paper*.

Dustin Hillard, Stephen Purpura, and John Wilkerson. 2008. Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology & Politics*, 4(4):31–46.

Thomas Hofmann. 1999. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc.

Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(02):311–331.

Melissa J Marschall and Robert J McKee. 2002. From campaign promises to presidential policy: Education reform in the 2000 election. *Educational Policy*, 16(1):96–117.

Seth Masket. 2009. *No middle ground: How informal party organizations control nominations and polarize legislatures*. University of Michigan Press.

Jon D Mcauliffe and David M Blei. 2008. Supervised topic models. In *Advances in neural information processing systems*, pages 121–128.

Juraj Medzihorsky, Levente Littvay, and Erin K Jenne. 2014. Has the tea party era radicalized the republican party? evidence from text analysis of the 2008 and 2012 republican primary debates. *PS: Political Science & Politics*, 47(04):806–812.

Federico Nanni and Pablo Ruiz Fabo. 2016. Entities as topic labels: Improving topic interpretability and evaluability combining entity linking and labeled lda. *To appear in the proceedings of Digital Humanities 2016*.

Vinodkumar Prabhakaran, Ashima Arora, and Owen Rambow. 2014. Staying on topic: An indicator of power in political debates. In *EMNLP*, pages 1481–1486.

Kevin M Quinn, Burt L Monroe, Michael Colaresi, Michael H Crespin, and Dragomir R Radev. 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209–228.

Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D Manning. 2009. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1*, pages 248–256. Association for Computational Linguistics.

Delip Rao, Paul McNamee, and Mark Dredze. 2013. Entity linking: Finding extracted entities in a knowledge base. In *Multi-source, Multilingual Information Extraction and Summarization*, pages 93–115. Springer.

Evan J Ringquist and Carl Dasse. 2004. Lies, damned lies, and campaign promises? environmental legislation in the 105th congress. *Social Science Quarterly*, 85(2):400–419.

Arno Scharl and Albert Weichselbraun. 2008. An automated approach to investigating the online media coverage of us presidential elections. *Journal of Information Technology & Politics*, 5(1):121–132.

Nate Silver. 2012. *The signal and the noise: Why so many predictions fail-but some don't*. Penguin.

Yanchuan Sim, Brice DL Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Jonathan B Slapin and Sven-Oliver Proksch. 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.

Andrea Volkens, Onawa Lacewell, Pola Lehmann, Sven Regel, Henrike Schultze, and Annika Werner. 2011. *The Manifesto Data Collection*. Manifesto Project (MRG/CMP/MARPOR), Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Cäcilia Zirn and Heiner Stuckenschmidt. 2014. Multidimensional topic analysis in political texts. *Data & Knowledge Engineering*, 90:38–53.

Cäcilia Zirn, Goran Glavaš, Federico Nanni, Jason Eichorst, and Heiner Stuckenschmidt. 2016. Classifying topics and detecting topic shifts in political manifestos. PolText.