Exploring structural features for position analysis in political discussions

Cäcilia Zirn, Michael Schäfer, Michael Strube, Simone Paolo Ponzetto and Heiner Stuckenschmidt
Data and Web Science Group, University of Mannheim, Germany
Heidelberg Institute for Theoretical Studies, Heidelberg, Germany
caecilia@informatik.uni-mannheim.de

Abstract

In the context of the NLP Unshared Task in PoliInformatics 2014, we analyze the structure of FOMC discussions as potential features for position analysis.
We access the length of discussion statements and show that the distinction between long opinionated statements and short spontaneous discussion elements improves the analysis of similarity among speakers. Furthermore, we explore the structure within dialogs by dividing them into subdialogs and representing the subsequence of speakers as graphs. In our web demo, we present visualizations of our analysis including the subgraphs and the similarity among speakers.

1 Introduction

In political science, scholars analyze the positions of persons active in politics. They base their research on various sources like surveys or votes, but also on speeches. Quantitative analysis of texts mostly employs statistical models based on words as in Grimmer and Stewart (2013), Hillard et al. (2008) or Laver et al. (2003). In the context of the NLP Unshared Task in PoliInformatics 2014, we will explore further features that might reveal new insights and ideas for analyzing positions of speakers in political discussions. More precisely, we will investigate the structure of the transcriptions of the Federal Open Market Committee (FOMC) meetings. We focus on the following tasks.

Distinction between statements and discussion elements. Not all of a speaker's utterances actually reveal their position in the same way. We show that it is useful to distinguish between long position statements and short, spontaneous discussion contributions.

Speaker similarity. We compare calculating the similarity between the speakers using their long position statements to the similarity using all their contributions.

Speech subsequence. As a first step in the direction of accessing the dialog structure, we divide the dialog in subdialogs and represent the speaker subsequences, i.e. who is replying to whom.

We visualize the above described features in diagrams that are shown in our web demo[1].

2 FOMC Data

In our work, we analyze the transcriptions of the FOMC meetings between 2005 and 2007. The FOMC is a committee within the central banking system of the US and decides on the target rate. It consists of members of the Federal Reserve Board and Federal Reserve Bank presidents of which 12 members have voting rights while the rest is only allowed to attend and participate in the discussions. Usually, the committee decides with consenting votes, dissenting votes appear rarely. However, the members do have different goals and positions (Havrilesky and Gildea, 1991), (Adolph, 2013). If scholars are interested in the preferences of the FOMC members, they thus rely on analyzing the discussions within the meetings rather than the votes.

3 Distinguishing statements from spontaneous discussion elements

Browsing through the dialogs, we figured out that there are two types of contributions to the discussions. In the first type, the speakers state their opinion, expressing arguments that they have assumably prepared in advance. Following those statements, other speakers ask questions or comment on the speaker's statement; discussions

[1] http://computerlanguste.de/acl2014

might arise. The contributions to those discussions are shorter and seem to be of spontaneous nature. We consider them as the second type. We think that the content of those two types of contributions - **statements** and **discussion elements** - can be of use for different purposes. **Statements** are prepared and reflect the general position of the speaker. According to research in political science, the political position of a speaker is determined by the topics he speaks about (cf. (Grimmer and Stewart, 2013), (Hillard et al., 2008),(Laver et al., 2003)) . The speaker will use the possibility to expand on the topics he considers important.

The shorter **discussion elements** are spontaneous reactions to the discourse contributions of the previous speakers. Rather than the topics the speakers considers important they contain an attitude towards previous speeches: the speaker often expresses agreement or disagreement, as in *"I can see why you assume that, but ..."* or *"To be honest, I don't think"*.

We manually annotated one meeting, classifying each discourse contribution either as statement or as discussion element. The sequence of those contributions and their word length together with their assigned class is shown in Figure 1.

From the diagram, we can see that the threshold between the statements and the discussion elements is around 500 words. We use this number as a shallow heuristic to automatically classify the discourse contributions into the two classes statement and discussion element. For a more advanced approach, a conditional random field could be trained to find statements followed by discussion elements using features like the statement total word count, the differences to word count of surrounding statements, and the speaker before and after the statement being the chairman. For our prototype we just rely on our simple heuristic to label the speaker statements, which correctly labels 98% of the speaker turns.

4 Calculating speaker similarity

Scholars of political science are interested in the positions of committee members and who of them hold similar views. As mentioned in Section 3, the positions are expressed through the topics mentioned in the speeches, which are mainly determined by nouns. We conclude that if two speakers share similar views, they are likely to use the same vocabulary. Therefore, we access the closeness of speakers by calculating the similarities between their speeches. As observed in Chapter 3, a speaker's positions are represented by the longer statements rather then the short dialog elements. As a consequence, we only include those statements in our calculation. In the spontaneous discussion elements, speakers tend to repeat the vocabulary of their previous speakers, for example by phrases like *"I do not agree with your view on unemployment"*, which would erroneously bias our similarity calculation. In natural language, topics are mainly determined by nouns. For our similarity analysis, we filter nouns only and lemmatize them and represent each speaker in each meeting as a word vector. Then, the cosine similarity is calculated for each speaker pair per meeting. In our web demo, we visualize the speaker pair similarities with heat maps. As we do not have a gold standard evaluating the correctness of the similarities, we investigate how stable the similarities of the speaker pairs are across all meetings. Two speakers that are close in one meeting should be close in the other meetings, too, as they are not likely to change their position while being on the committee. For each speaker pair, we calculated the standard deviation of the similarities across all meetings they both attend. They range from zero to 0.37, on average being 0.08. For two thirds of the speaker pairs the standard deviation is below 0.1. Hence, this approach can be considered as being very robust.

To evaluate our hypothesis that the longer statements are more relevant for determining the speakers' positions, we compare the above described results to the similarities calculated using all utterances of a speaker including spontaneous discussion elements. The standard deviations range up to 0.46 with an average of 0.1. For better comparability, we plotted the standard deviations of both experiments in Figure 3 sorted in descending order. We can clearly see that the standard deviations for the similarities calculated using statements only is continuously below the standard deviations based on both utterance types.

4.1 Interpreting similarity results

To get an insight into the results of our similarity analysis, we use a scaling of Reuter's[2] that classifies current FOMC members on a scale from one to five, with the first group being "inflation doves"

[2]http://graphics.thomsonreuters.com/F/10/scale.swf

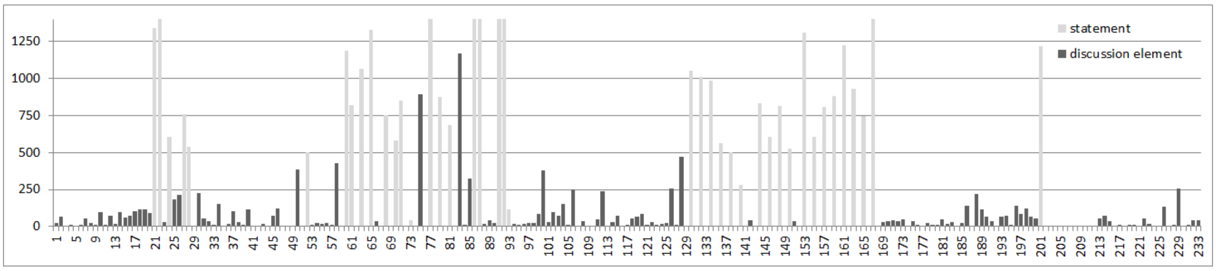

Figure 1: Manually annotated discourse contributions and their lengths (word count).

with no strong bias towards change, and group 5 being "hawks" with somewhat extreme opinions. For a comparison, we use the heatmap showing similarities of the most recent meeting in our data set, which was held on 2007/12/11 (see Figure 2). We put out attention to the speakers that overlap between Reuter's scaling and the members in the 2007 meeting, which are: Yellen (1), Rosengren (1), Dudley (1), Evans (1), Bernanke (2), Pianalto (2), Lockhart (2), Plosser (5), Lacker (5) and Fisher (5). The numbers in brackets gives their group. The most extreme hawks are Plosser, Lacker and Fisher. As can be seen in the heatmap, Fisher seems to have different point of views as most other speakers, with similarity values between 0.22 and 0.52. His highest similarity (0.52) score is for Plosser, who is a hawk, too. The lowest scores are with the Doves Rosengren and Dudley and the dovish Lockhart. Lacker has the highest similarity score with Plosser (0.62), one of the lowest with dove Dudley (0.39). Plosser has the highest similarity scores with Lacker and Fisher. The dove Dudley seems to have different views from all other members having only low similarity scores, the second lowest score however (0.24) is with the dove Fisher. Our observations suggests that there is a correlation between our similarity scores and the dove-hawk-scaling produced by an expert. However, we cannot directly compare the results. For one hand, the data stems from different years. But on the other hand, which is more important, the dove-hawk-scale determines how strong and extreme the views are, not necessarily their position.

5 Identifying subdialogs

We are interested in the structure of the dialogs. As we described in Section 3, we observed that the meetings consist of longer statements in which the speakers express their position followed by discussions. We can thus divide a meeting into subdi-

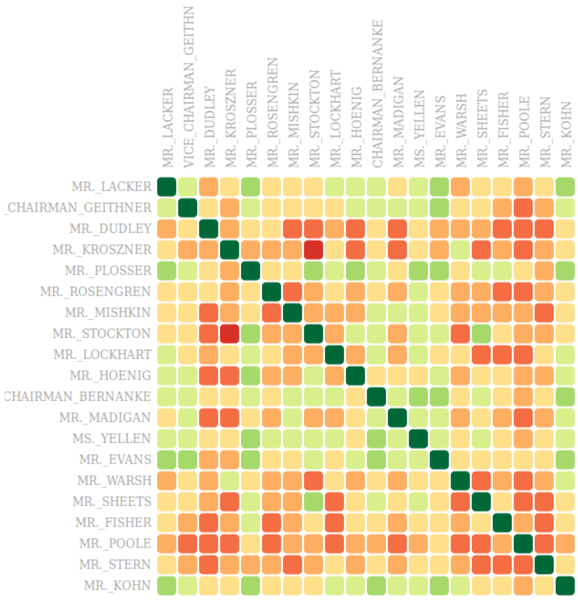

Figure 2: Heatmap showing similarities between speakers within the meeting on 2007/12/11, cosine similarities ranging from dark green ($\geq$ 0.9) to dark orange ($\geq$ 0.1).

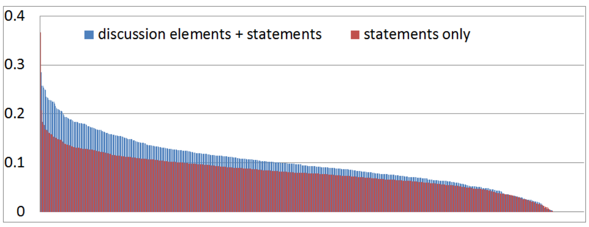

Figure 3: Comparison of the standard deviations of similarities for each speaker pair sorted in descending order calculated on the long statements only vs. calculated on both statements and discussion elements.

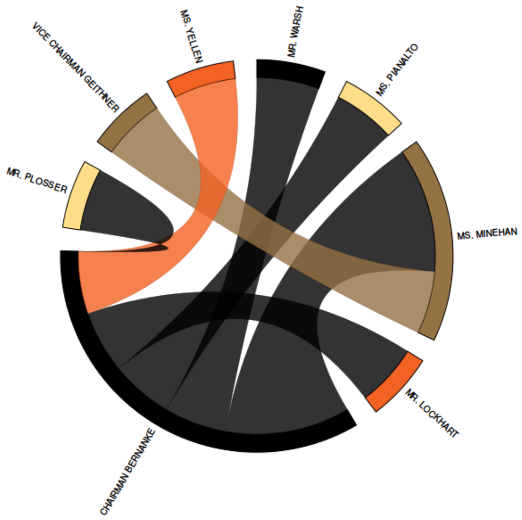

Figure 4: Chord diagram showing a subdialog.

alogs, starting at a statement followed by spontaneous discussion elements. The meetings are automatically divided whenever we find a statement consisting of more than 500 words (see Figure 1). Then, we can analyze the dialog structure: who is replying to whom, who are the active speakers? We visualize the subdialog in chord diagrams, as shown in Figure 4.

A relation between speaker A and B means that speaker B is talking after speaker A. We connect speaker A and B with an edge, while the weight of an edge is the number of those relations occuring within the subdialog. The first phrases a speaker chooses to address another within a discussion reveals essential information about how he agrees with the previous speaker. (Stolcke et al., 2000) refer to such phrases as dialogue acts. Those dialog acts can be used to determine agreement and disagreement between the speakers. For our prototype, we start by visualizing the speaker's adjacency only, and label the edges with all beginning phrases, which we extract by simply choosing the first 5 words of the discourse. The next step will be to extract them in a more sophisticated way and classify them for agreement and disagreement.

A special role is taken by the chairman. On the one hand, he has a functional role and just calls speakers. On the other hand, he is a member of the committee contributing his own views. In the next step of our work, we aim at identifying his statements having a functional role only. Then, we can introduce a direct "speaks after" relation between the previous and the following speaker. Most charts in our web demo clearly reveal the central role of the chairman.

6 Related work

(Abu-Jbara et al., 2012) explored the dialog structure in on-line debates with the goal of subgroup detection. They represented each discussion participant as a vector consisting of the polarity and the target of their opinionated phrases, combining it with the information about who replies to whom. In a final step, they cluster the vectors. They point out that the reply feature needs further investigation since they cannot tell whether speakers tend to agree or disagree when they answer each other.

Exploring agreement and disagreement between speakers is also used by (Thomas et al., 2006). Their goal is to label congressional floor-debate speeches as supporting or opposing the discussed topic. They model speech turns as nodes connected by "same label" relations and then find minimum cuts in the resulting graph. To obtain the relations, they first classify speeches separately with common machine learning techniques as supporting or opposing. Then, they add agreement links between speakers: they extract the context around references by name from one speaker to another, and apply a classifier that was trained on a corpus using consenting votes of the speakers as labels. The agreement links improve their results consistently.

7 Future Work

The focus of our research is on finding speakers with similar opinions. We intend to represent speakers in a multi-graph where each type of edge corresponds to a different feature. One of the features is the similarity of the speakers' statements, as presented in chapter 4. Further important features are agreement and disagreement among the speakers, which can be accessed by further analyzing the subsequent speaker links presented in chapter 5: The beginning of the statements contain dialog acts as described in (Stolcke et al., 2000), which can be classified. More features can be added to the multi-graph such as the political orientation of the speakers, if known. To finally determine the groups of consenting speakers, we can either apply graph clustering methods or find minimum cuts in the graph. Another task that we will address is the special leading role of the chairman.

Acknowledgments

The authors thank the DFG for Funding under the SFB 884 Political Economy of Reforms C4 project. Furthermore, they thank Nicole Rae Berg for helpful advice from a political science perspective, and Michael Dell for assisting with the implementation.

References

Amjad Abu-Jbara, Mona Diab, Pradeep Dasigi, and Dragomir Radev. 2012. Subgroup detection in ideological discussions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 399–409, Stroudsburg, PA, USA. Association for Computational Linguistics.

Christopher Adolph. 2013. *Bankers, Bureaucrats, and Central Bank Politics: The Myth of Neutrality*. Cambridge University Press.

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

Thomas Havrilesky and John Gildea. 1991. The policy preferences of fomc members as revealed by dissenting votes: Comment. *Journal of Money, Credit and Banking*, 1:130–138.

Dustin Hillard, Stephen Purpura, and John Wilkerson. 2008. Computer assisted topic classification for mixed methods social science research. *Journal of Information Technology and Politics*.

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.

Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Comput. Linguist.*, 26(3):339–373, September.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 327–335, Stroudsburg, PA, USA. Association for Computational Linguistics.