Investigating Political Herd Mentality: A Community Sentiment Based Approach

Anjali Bhavan*
Delhi Technological University
anjalibhavan98@gmail.com

Rohan Mishra*
Delhi Technological University
rohan.mishra1997@gmail.com

Pradyumna Prakhar Sinha*
Delhi Technological University
pradyumna014@gmail.com

Ramit Sawhney
Netaji Subhash Institute of Technology
ramits.co@nsit.net.in

Rajiv Ratn Shah
MIDAS, IIIT-Delhi
rajivratn@iiitd.ac.in

Abstract

Analyzing polarities and sentiments inherent in political speeches and debates poses an important problem today. This experiment aims to address this issue by analyzing publicly-available Hansard transcripts of the debates conducted in the UK Parliament. Our proposed approach, which uses community-based graph information to augment hand-crafted features based on topic modeling and emotion detection on debate transcripts currently surpasses the benchmark results on the same dataset. Such sentiment classification systems could prove to be of great use in today's politically turbulent times, for public knowledge of politicians stands on various relevant issues proves vital for good governance and citizenship. The experiments also demonstrate that continuous feature representations learned from graphs can improve performance on sentiment classification tasks significantly.

1 Introduction

One of the key aspects of a functional, free society is being presented with comprehensive options in electing government representatives. The decision is aided by the positions politicians take on relevant issues like water, housing, etc. Hence, it becomes important to relay political standings to the general public in a comprehensible manner. The Hansard transcripts of speeches delivered in the UK Parliament are one such source of information. However, owing to the voluminous quantity, esoteric language and opaque procedural jargon of Parliament, it is tougher for the non-expert citizen to assess the standings of their elected representative. Therefore, conducting stance classification studies on such data is a challenging task with potential benefits. However, the documents tend to be highly tedious and difficult to comprehend, and thus become a barrier to information about political issues and leanings.

Sentiment analysis of data from various relevant sources (social media, newspapers, transcripts, etc.) has often given several insights about public opinion, issues of contention, general trends and so on (Carvalho et al., 2011; Loukis et al., 2014). Such techniques have even been used for purposes like predicting election outcomes and the reception of newly-launched products. Since these insights have wide-ranging consequences, it becomes imperative to develop rigorous standards and state-of-the-art techniques for them.

One aspect that helps with analyzing such patterns and sentiments is studying about the interconnections and networks underlying such data. Homophily, or the tendency of people to associate with like-minded individuals, is the fundamental aspect of depicting relationships between users of a social network (for instance). Constructing graphs to map such complex relationships and attributes in data could help one arrive at ready insights and conclusions. This comes particularly useful when studying parliamentary debates and sessions; connecting speakers according to factors like party or position affiliations pro-

* Indicates equal contribution.

Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 281–287
Florence, Italy, July 28 - August 2, 2019.

vides information on how a speaker is likely to respond to an issue being presented. Attempts to analyze social media data based on such approaches have been made (Deitrick and Hu, 2013).

2 Related Work

The analysis of political content and parliamentary debates have opened an exciting line of research in recent years and has shown promising results in tasks of stance classification (Hasan and Ng, 2013) and opinion mining (Karami et al., 2018). A large part of the work initially concentrated on legislative speeches, but the focus has shifted to social media content analysis in recent times. This shift in focus has been particularly rapid with the proliferation of social media data and research (**?**Shah and Zimmermann, 2017; Shah et al., 2016b; Mahata et al., 2018; Shah et al., 2016c,a).

Lauderdale and Herzog (2016) presented their method of determining political positions from legislative speech. The datasets were sourced from Irish and US Senate debates. Rheault et al. (2016) examined the emotional polarity variations in speeches delivered in the British parliament over a hundred years. They observed a correlation between the variations in emotional states of a particular period of time and the national economic situation. Thomas et al. (2006) studied the relationships between segments of speeches delivered in the Congress and the overall tone: of opposition or support. A significant amount of research exists on the political temperament across social media websites like Facebook and Twitter. Stieglitz and Dang-Xuan (2012) studied the relationship between the inherent sentiment of politically relevant tweets and the retweet activity. Ceron et al. (2014) proposed methods for determining the political alignments of citizens and tested them on French and Italian-context datasets. Many new findings based on the contemporary political landscape continue to be developed and presented. Wang and Liu (2018) analyzed US President Donald Trump's speeches delivered during his 2016 election campaign. Rudkowsky et al. (2018) proposed the usage of word embeddings in the place of the traditional Bag-of-Words (BOW) approach for text analysis, and demonstrated experiments on Austrian parliamentary speeches. There have been some approaches to model interactions among members of a network to help in the task of sentiment analysis. Moreover, there have been applications that extract information about each user by representing them as a node in the social graph and creating low dimensional representation usually induced by neural architectures (Grover and Leskovec, 2016). Mishra et al. (2018) and Qian et al. (2018) use such social graph-based features to gain considerable improvement in the task of abuse detection in social media. However there has been no work done to model the interaction between the members of the Parliament for the task of stance classification.

For studying transcripts of speeches delivered in the House of Commons in the UK Parliament, Abercrombie and Batista-Navarro (2018b) curated a dataset consisting of parliamentary motions and debates as provided in the Hansard transcripts, along with other information like party affiliations and polarities of the motions being discussed. This was followed by carrying out studies on the dataset and developing a sentiment analysis model which also demonstrated the results of motion-independent (one-step) and motion-dependent classification of polarities Abercrombie and Batista-Navarro (2018a). This dataset is used for further analysis in our experiments.

3 Dataset

In the UK, transcripts of parliamentary debates are publicly available along with information related to *division votes* as well as manually annotated sentiment labels. To investigate the effectiveness of our pipeline, experiments were conducted using the *HanDeSeT* dataset as created by (Abercrombie and Batista-Navarro, 2018b). The dataset consists of 607 politicians and their speeches over various motions, with a total of 1251 samples. The speeches are divided into five utterances, and other features such as *Debate ID, Debate title, Motion subject with polarities: manual annotation and ruling-opposing-based, Motion and Speaker party affiliations, Speech Polarities: manual and vote-based, Rebellion percentage.*

Sentiment polarity is present in both speeches and motions. Hence labels are provided for motion polarities as well. Two label types are provided for motions: a manually-annotated one predicting positive or negative polarity, and a government/opposition one decided as follows: if the speaker who proposes the motion belongs to the ruling government, the polarity is positive; if the speaker belongs to the opposition then the polarity is negative. Two label types are provided for speeches as well: one manually-annotated, and the other a speaker-vote label extracted from the division related to the corresponding debate.

4 Methodology

The models described in Abercrombie and Batista-Navarro (2018a) extracted n-gram features (uni-grams, bi-grams, tri-grams, and their combinations) from the utterances for sentiment classification. The stance-based relationships between the members are modeled, and their effectiveness is analyzed. This study aims to develop on the limitations of using only text-based features and by doing so present a sound, coherent model for sentiment classification for parliamentary speeches.
The methodology consists of the following subsections: *preprocessing*, to describe the initial data preprocessing methods undertaken; *feature extraction*, which discusses the feature sets used for our model, and *model description* and training, to elaborate on our model and training procedures.

4.1 Preprocessing

The dataset was preprocessed for further analysis. This was required so unnecessary words; characters etc. could be removed and not add further noise to the dataset. The text was lower-cased, and all punctuation marks and other special characters were removed. Following this, stopword removal was done using NLTK. Finally, a few custom stopwords specific to the parliamentary procedure were removed. These were taken from Abercrombie and Batista-Navarro (2018b). Finally, the utterances were concatenated and prepared for feature extraction and model training.

4.2 Feature Extraction

4.2.1 Textual Features

Various textual features were extracted for classification and normalized using the L2 norm. These are listed below.

- *TF-IDF*: Term Frequency-inverse Document Frequency (TF-IDF) features were extracted from n-grams (upto 3) in the text. N-gram features are immensely useful for factoring in contextual information surrounding the components of a text (whether characters or words) and are widely used for text analysis, language processing, etc.

- LDA-based topic modeling: Topic modeling is used to derive information related to the underlying ”topics” contained in a text. In order to extract such topic-based features from the utterances, the Latent Dirichlet Allocation (LDA) (Blei et al., 2003) model was used. The probability distribution over the most commonly occurring 30 topics was used as features for each speech.

- *NRC Emotion*: The NRC Emotion Lexicon (Mohammad and Turney, 2013) is a publicly available lexicon that contains commonly occurring words along with their affect category (anger, fear, anticipation, trust, surprise, sadness, joy, or disgust) and two polarities (negative or positive). The score along these 10 features was computed for the utterances.

4.2.2 Graph-based features

For our analysis, two graphs were constructed from the dataset. The graph consists of nodes that represent the members who participate in the proceedings of the Parliament. The edges among the members are conditioned upon their accord or discord on debates regarding policies. Two members of the same or varying political parties either agree on a policy or differ on it. Therefore, the two graphs are constructed.

- *simGraph*: In order to model the similarity on stances among members, $G_{sim}(v, e)$ is a weighted undirected graph induced on the dataset with vertices v corresponding to the members m of political parties where an edge e between two vertices v and u is defined as $weight(e) = \mid f(v) \cap f(u) \mid$ where $f(v)$ is the set of stances taken by the member that is represented by node v.

- *oppGraph*: Similarly, to model the differences among the members, $G_{opp}(v, e)$ is induced on the dataset such that an edge e between two vertices v and u is defined as $weight(e) = \mid (f(v) \setminus f(u)) \cap (f(u) \setminus f(v)) \mid$ where $f(v)$ is the set of stances taken by the member that is represented by node v.

node2vec: To obtain community based embeddings, feature representations were generated using *node2vec* (Grover and Leskovec, 2016).

Table 1: Statistical properties of constructed graphs

Properties	**Values**	
	simGraph	oppGraph
Number of nodes	607	607
Number of edges	5,431	2,893
Density of graphs	0.0295	0.0157
Average weight	1.047	1.037

node2vec is similar to word2vec (Mikolov et al., 2013b) and uses the same loss function to assign similar representations to nodes that are in the context of each other. To obtain the context of a node, node2vec samples a neighborhood for each of the nodes by constructing a fixed number of random walks of constant length. The traversal strategy for these random walks is determined by the hyper-parameters Return Parameter p and In-out Parameter q which have the ability to moderate the sampling between a depth-first strategy and a breadth-first strategy. The return parameter p controls the likelihood of immediately revisiting a node in the walk, while the in-out parameter q allows the search to differentiate between inward and outward nodes.

Formally, given a graph $G = (V, E)$, we learn a function $f : V \rightarrow \mathbb{R}^d$ that maps nodes to feature representations where d is the dimension of the representation. In order to do so, for every node $u \in V$, we define a neighbourhood $N_S(u) \subseteq V$ is generated using the sampling strategy S.

The skip-gram model (Mikolov et al., 2013a) is then employed to maximize the following objective function:

$$\max_f \sum logPr(N_s(u)|f(u)). \qquad (1)$$

Combining Graph Embeddings: To combine embeddings generated for each member in the two graphs, a dense neural network was used. The embeddings were projected onto a linear layer and fine-tuned upon the classification task. The penultimate layer of the model was used as the graph embedding corresponding to each user.
The network consisted of two input layers for the two embedding sets, followed by single dense layers with hidden layer size 16 and activation ReLU. These two layers were then combined, and the resultant combination passed through two dense layers (layer size 16, activation ReLU), before being passed through a final dense softmax layer. The network was optimized using Adam, and trained over 20 epochs with batch size 64.

4.2.3 Other features

Of all the feature sets explored in Abercrombie and Batista-Navarro (2018a), the feature set all the meta-features had the best results consistently across all the three models. Hence, we used these in addition to our textual and community-based graph features. The meta-features consisted of speaker party affiliation, debate IDs and motion party affiliation.

4.3 Baseline models

The original experiments consisted of 3 models for classification: a one-step model and two two-step models. We consider the two-step models as our baselines, which are described below.

- manAnnot: a two-step model in which motion polarity classification is first performed based on manually-annotated positive or negative sentiments, corresponding to model 2a in the original experiments;
- govAnnot: a two-step model in which motion polarity classification is first performed based on government or opposition labeling, corresponding to model 2b in the original experiments.

In the case of the two two-step models, the dataset is divided into two parts based on the predicted polarities. These two divided datasets are then used for training and classification separately. Two classifiers were used in both the steps: Support Vector Machine (SVM) with the linear kernel and Multi-Layer Perceptron (MLP) with 1 hidden layer containing 100 neurons.

4.4 Proposed model

In the original experiment, the best results were obtained from the two-step models with the MLP classifier. A similar two-step approach is followed here as well, with MLP as the chosen classifier. The network consists of 1 hidden layer with 100 neurons.

5 Experiments

Experiments on two models are presented:

1. manAnnot: here, the dataset is divided into two parts based on predicted motion polarity from manually annotated labels.

Table 2: Observations for manModel

Feature Combinations	without graph-based features				with graph-based features			
	Acc.(%)	Prec.	Recall	F1	Acc.(%)	Prec.	Recall	F1
TF-IDF+meta	89.38	0.897	0.887	0.884	92.26	0.920	0.917	0.917
LDA+meta	86.34	0.875	0.839	0.850	92.34	0.930	0.902	0.915
NRC+meta	86.43	0.860	0.859	0.858	92.25	0.932	0.903	0.916
TF-IDF+LDA+meta	88.59	0.885	0.867	0.874	91.70	0.915	0.910	0.912
TF-IDF+NRC+meta	88.73	0.896	0.867	0.879	91.94	0.918	0.914	0.914
LDA+NRC+meta	85.86	0.861	0.828	0.842	**92.66**	**0.938**	**0.905**	**0.920**
TF-IDF+LDA+NRC+meta	90.89	0.908	0.900	0.908	91.78	0.917	0.909	0.919

Table 3: Observations for govModel

Feature Combinations	without graph-based features				with graph-based features			
	Acc.(%)	Prec.	Recall	F1	Acc.(%)	Prec.	Recall	F1
TF-IDF+meta	90.25	0.902	0.900	0.898	**92.72**	**0.927**	**0.924**	**0.923**
LDA+meta	88.42	0.879	0.880	0.877	92.09	0.917	0.923	0.918
NRC+meta	86.91	0.875	0.848	0.858	92.73	0.927	0.919	0.921
TF-IDF+LDA+meta	89.06	0.887	0.882	0.883	92.33	0.920	0.925	0.920
TF-IDF+NRC+meta	88.97	0.892	0.883	0.885	92.80	0.923	0.911	0.914
LDA+NRC+meta	86.35	0.864	0.851	0.855	92.41	0.923	0.922	0.920
TF-IDF+LDA+NRC+meta	89.22	0.896	0.876	0.883	92.33	0.920	0.922	0.918

2. govAnnot: Here, the dataset is separated into two parts based on the speaker's affiliation: if the speaker presenting the motion belongs to the ruling government, then the motion polarity is positive, or otherwise negative.

The hyperparameters (for each of the feature sets and the classifier) were tuned using grid search. L-BFGS (Liu and Nocedal, 1989) was used for optimization in the neural network. Model training and evaluation was carried out using stratified 10-fold cross-validation. Stratification was performed to account for the slight imbalance in the dataset. Two types of labels are presented in the dataset: vote-based and manually-annotated. We use the manually-annotated labels for our experiments.

For the graph-based features, a grid search was performed which yielded the following parameters for generating embeddings:

- simGraph: p = 10, q = 1, walk length = 15, number of walks = 15, window size = 10. The feature vector obtained from these parameters yielded an accuracy of 79.51%.

- oppGraph: p = 0.1, q = 10, walk length = 5, number of walks = 10, window size = 10. The feature vector obtained from these parameters yielded an accuracy of 69.53%.

6 Results and Discussion

Table 2 and Table 3 present the results on the two models respectively. The values of accuracy, precision, recall, and F1-score are presented on feature sets with and without graph-based features.

In the case of both models, the usage of graph-based features outperforms the results obtained without using them. The difference is large in the case of the feature set comprising of LDA, NRC, and meta-features in the model with manually-annotated labels: the F1 scores obtained with and without graph features differ by 7.8%.

It can be observed that by using graph-based features The baselines for both have been surpassed by using graph-based features along with the other textual and meta-features. Our best results for manAnnot are obtained by using the combination of LDA, NRC, and graph-based features along with meta-features. The best results for govAnnot are obtained by using the combination of TF-IDF and meta-features along with graph-based features.

7 Conclusion

We presented a method for sentiment analysis of parliamentary debate transcripts, which could go a long way in helping determine the position an elected representative might assume on issues of great importance to the general public. The experiments were carried out on the Hansard parliamentary debates dataset (Abercrombie and Batista-Navarro, 2018b). We performed experiments on a variety of textual analysis methods (e.g. topic modeling, emotion classification, n-grams), and combined them with community-based graph features obtained by representational learning on the dataset using node2vec. Our results surpass the state-of-the-art results using both govAnnot and manAnnot. Also, the F1 and accuracy values of the models using graph-based features are higher than those without graph-based features, the difference being considerable in some cases. This gives sufficient demonstration for the ability of representational learning to enhance performances on tasks like sentiment analysis.

8 Future Work

Future work in this area could involve the following aspects:

- Application of the proposed approach to tasks other than sentiment classification, for instance analysis of mental health and suicide ideation on social media.
- Constructing different graphs and analyzing other training and feature extraction methods for enhancing performance and deriving better inferences.
- Application of the proposed approach for analyzing data in different contexts; an example could be the analysis of the recently-conducted elections in India.
- Extend the proposed methodology to other problems (Mahata and Talburt, 2015; Mahata et al., 2015a,b) based on social media.

References

Gavin Abercrombie and Riza Batista-Navarro. 2018a. 'aye'or'no'? speech-level sentiment analysis of hansard uk parliamentary debate transcripts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, pages 33–40.

Gavin Abercrombie and Riza Theresa Batista-Navarro. 2018b. Identifying opinion-topics and polarity of parliamentary debate motions. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 280–285.

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

Paula Carvalho, Luís Sarmento, Jorge Teixeira, and Mário J Silva. 2011. Liars and saviors in a sentiment annotated corpus of comments to political debates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 564–568. Association for Computational Linguistics.

Andrea Ceron, Luigi Curini, Stefano M Iacus, and Giuseppe Porro. 2014. Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to italy and france. *New media & society*, 16(2):340–358.

William Deitrick and Wei Hu. 2013. Mutually enhancing community detection and sentiment analysis on twitter networks. *Journal of Data Analysis and Information Processing*, 1(03):19.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM.

Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356.

Amir Karami, London S Bennett, and Xiaoyun He. 2018. Mining public opinion about economic issues: Twitter and the us presidential election. *International Journal of Strategic Decision Sciences (IJSDS)*, 9(1):18–28.

Benjamin E Lauderdale and Alexander Herzog. 2016. Measuring political positions from legislative speech. *Political Analysis*, 24(3):374–394.

Dong C Liu and Jorge Nocedal. 1989. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Euripides Loukis, Yannis Charalabidis, and Aggeliki Androutsopoulou. 2014. An analysis of multiple social media consultations in the european parliament from a public policy perspective.

Debanjan Mahata, Jasper Friedrichs, Rajiv Ratn Shah, and Jing Jiang. 2018. Detecting personal intake of medicine from twitter. *IEEE Intelligent Systems*, 33(4):87–95.

Debanjan Mahata and John R Talburt. 2015. A framework for collecting, extracting and managing event identity information from twitter. In *ICIQ*.

Debanjan Mahata, John R Talburt, and Vivek Kumar Singh. 2015a. From chirps to whistles: Discovering event-specific informative content from twitter. In *Proceedings of the ACM web science conference*, page 17. ACM.

Debanjan Mahata, John R Talburt, and Vivek Kumar Singh. 2015b. Identification and ranking of event-specific entity-centric informative content from twitter. In *International Conference on Applications of Natural Language to Information Systems*, pages 275–281. Springer.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Pushkar Mishra, Marco Del Tredici, Helen Yannakoudakis, and Ekaterina Shutova. 2018. Author profiling for abuse detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1088–1098.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Council, Canada*.

Jing Qian, Mai ElSherief, Elizabeth M Belding, and William Yang Wang. 2018. Leveraging intra-user and inter-user representation learning for automated hate speech detection. *arXiv preprint arXiv:1804.03124*.

Ludovic Rheault, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. 2016. Measuring emotion in parliamentary debates with automated textual analysis. *PloS one*, 11(12):e0168843.

Elena Rudkowsky, Martin Haselmayer, Matthias Wastian, Marcelo Jenny, Štefan Emrich, and Michael Sedlmair. 2018. More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, 12(2-3):140–157.

Rajiv Shah and Roger Zimmermann. 2017. *Multimodal analysis of user-generated multimedia content*. Springer.

Rajiv Ratn Shah, Anupam Samanta, Deepak Gupta, Yi Yu, Suhua Tang, and Roger Zimmermann. 2016a. Prompt: Personalized user tag recommendation for social media photos leveraging personal and social contexts. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 486–492. IEEE.

Rajiv Ratn Shah, Yi Yu, Suhua Tang, Shin'ichi Satoh, Akshay Verma, and Roger Zimmermann. 2016b. Concept-level multimodal ranking of flickr photo tags via recall based weighting. In *Proceedings of the 2016 ACM Workshop on Multimedia COMMONS*, pages 19–26. ACM.

Rajiv Ratn Shah, Yi Yu, Akshay Verma, Suhua Tang, Anwar Dilawar Shaikh, and Roger Zimmermann. 2016c. Leveraging multimodal information for event summarization and concept-level sentiment analysis. *Knowledge-Based Systems*, 108:102–109.

Stefan Stieglitz and Linh Dang-Xuan. 2012. Political communication and influence through microblogging – an empirical analysis of sentiment in twitter messages and retweet behavior. In *2012 45th Hawaii International Conference on System Sciences*, pages 3500–3509. IEEE.

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*, pages 327–335. Association for Computational Linguistics.

Yaqin Wang and Haitao Liu. 2018. Is trump always rambling like a fourth-grade student? an analysis of stylistic features of donald trumps political discourse during the 2016 election. *Discourse & Society*, 29(3):299–323.