Identifying Opinion-Topics and Polarity of Parliamentary Debate Motions

Gavin Abercrombie and **Riza Batista-Navarro**
School of Computer Science
University of Manchester
Kilburn Building, Manchester M13 9PL
gavin.abercrombie@postgrad.manchester.ac.uk
riza.batista@manchester.ac.uk

Abstract

Analysis of the topics mentioned and opinions expressed in parliamentary debate motions–or proposals–is difficult for human readers, but necessary for understanding and automatic processing of the content of the subsequent speeches. We present a dataset of debate motions with pre-existing '*policy*' labels, and investigate the utility of these labels for simultaneous topic and opinion polarity analysis. For topic detection, we apply one-versus-the-rest supervised topic classification, finding that good performance is achieved in predicting the *policy* topics, and that textual features derived from the debate titles associated with the motions are particularly indicative of motion topic. We then examine whether the output could also be used to determine the positions taken by proposers towards the different policies by investigating how well humans agree in interpreting the opinion polarities of the motions. Finding very high levels of agreement, we conclude that the *policies* used can be reliable labels for use in these tasks, and that successful topic detection can therefore provide opinion analysis of the motions 'for free'.

1 Introduction

In the House of Commons of the UK Parliament, the topics contained in a debate's *motion*–a proposal one Member of Parliament (MP) puts to the other Members of the House–are the focus of opinions expressed during all subsequent speeches. These motions are therefore crucial for understanding the content of MPs' speeches and the opinions they convey.

It is often difficult for people to process debate motions due to the level of domain-specific knowledge related to the language and workings of Parliament they contain. Indeed, these motions are so hard for ordinary citizens to understand that parliamentary monitoring organisations like the Public Whip[1] and They Work for You[2] produce manually written summaries and annotated versions of them, which are written by crowd-sourced volunteers with domain expertise or interest.
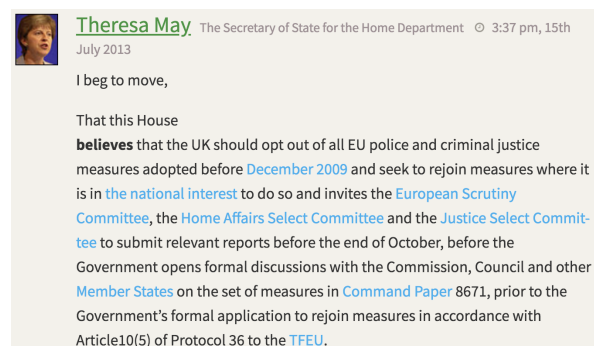

Figure 1: Example of a debate motion as presented on the theyworkforyou.com website. Motions contain information about the topics of debates, as well as the proposer's opinions towards those topics, but can be difficult for human readers to process.

In conducting sentiment analysis of debate speeches, it has been observed that, when formulating these motions, the speakers that propose them themselves express sentiment towards the topics of the motions, and that these motions can act as polarity shifters for subsequent speeches in a debate–that is, depending on the sentiment polarity of a motion, the sentiment polarity of language used in subsequent speeches may be reversed (Abercrombie and Batista-Navarro, 2018). Identification of both the topics and polarity of motions is therefore crucial for any further investigation of debates, and is likely to be a key step in tasks such as sentiment or stance analysis of debate speeches.

Our contributions We create a dataset of UK parliamentary debate motions labelled with both topic

---

[1] http://www.publicwhip.org.uk
[2] https://www.theyworkforyou.com/

and opinion polarity. We then investigate the utility of these labels for two tasks:

1) For motion topic detection, we treat the *policy* labels as topic classes and assess the performance of a multilabel classifier in predicting them.

2) We exploit the fact that the *policies* used as topic labels inherently incorporate information regarding the proposers' opinions towards those policies–that is their policy positions: whether they support or oppose them. We investigate whether, by correctly identifying a motion's *policy* category, we can also determine its position towards the policy in question, in effect obtaining opinion analysis of the motions 'for free'. We compare the output of this approach to human produced opinion polarity labels.

2 Background

2.1 Hansard debate transcripts

Debates in the House of Commons are of the following format: An MP proposes a *motion*, to which other MPs may respond when invited by the *Speaker* (the presiding officer of the chamber), either in support of, or opposition to the motion.

A domain with unique characteristics, the *Hansard* transcripts lie somewhere between formal written language and transcripts of spoken dialogue–they are near-verbatim transcriptions of almost everything that is said in Parliament, although disfluencies are removed and some contextual information (such as the names of the speakers) is added by the parliamentary reporters.

There exist a number of challenges associated with this domain. Here, analysis is complicated by the language employed by politicians, who tend to use: (1) little extreme or overtly polarised (especially negative) language, and (2) a tactical, political use of terminology–for example, policies that may be percieved as negative (such as cuts to services or tax increases) are generally not framed using those terms (Abercrombie and Batista-Navarro, 2018).

Additionally, the format of debates is complex, with manifold topics discussed by multiple participants. Motions may reference various entities, some of which may be described only within other debates or documents referred to in the motion (as in Figure 1).

Finally, the language used is often arcane, with much procedural terminology. In fact, many motions consist entirely of such language, giving lit-tle or no clue as to the topic under discussion (for further details see Section 3).

However, the existence of motions that have been manually labelled with 'policy votes' indicates that it may be feasible to train machine classifiers to conduct a form of motion topic detection. The fact that these labels also encompass policy positions suggests that they could also be used simultaneously for opinion analysis.

2.2 Opinion-topic labelling

Parliamentary monitoring website the Public Whip maintains a list of debates organised under '*policies*.'[3] These are sets 'of votes that represent a view on a particular issue' such as *European Union – For* and *Stop climate change*. Under each of these, members of the public are invited to submit debates–motions with vote outcomes–which match these descriptions.

We make use of these categorisations as labels for supervised topic classification. In many cases, it is not straightforward to determine a motion's *policy* label from the debate title–for example, for the *policy* '*More Powers for Local Councils*', debate titles include 'High Streets', 'Housing', 'Fixed Odds Betting Terminals', 'Local Bus Services'. Similarly, the text of a motion alone does not necessarily reveal its topic, with many motions consisting purely of procedural language, such as 'That the Bill be read a Second time'. As a result, human readers often require access to the title, motion, and sometimes other information found elsewhere in a debate in order to determine the motion's polarity.

While the *policies* represent both a policy topic and a polarised position towards it, this is a reflection of the vote outcome of the debate, not necessarily the position expressed in the motion. For example, if a motion proposed in support of a policy position is rejected, it will be labelled with a *policy* that reflects opposition to that position (see Figure 2).

In a further layer of complexity, the Public Whip also provides motions with a 'policy vote' label–the contributors' assessment of how somebody who supports each policy 'would have voted'–with the additional tags 'majority', 'minority', or 'abstain'. All in all, this means that each label has two potential polarity shifters (the vote
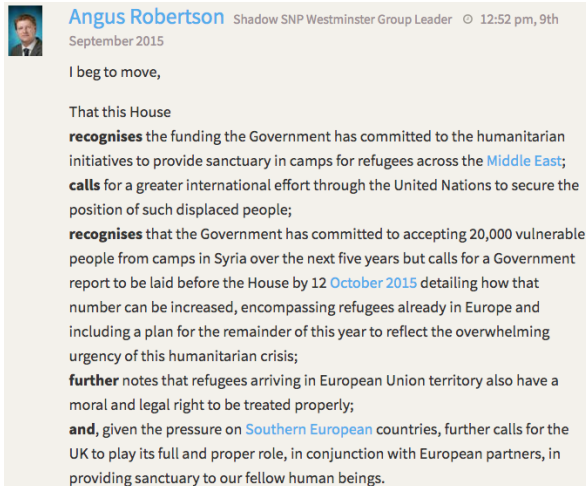
Figure 2: Motion from the *policy* category *Asylum System – More strict*. The motion, from a debate entitled '*Humanitarian Crisis in the Mediterranean and Europe*', opposes the idea of making the asylum system stricter, but the fact that it was rejected by the House, explains why it has been given this label.

outcome and the *policy vote*), which need to be taken into account if the Public Whip *policies* are to be used as labels for opinion polarity analysis.

3 Data

We present a dataset[4] of 592 UK parliamentary debate motions proposed in the House of Commons between 1997 and 2018.[5] We match these with the corresponding *policies* from the Public Whip for use as labels for supervised opinion-topic classification. We therefore include only those motions, which have been classified by *policy* on the Public Whip website. In order to provide sufficient examples to train a classifier we use only those debates for which there exist at least 20 examples per *policy* label. Because, for example, a debate may have been categorised with both the specific *policy* '*Higher taxes on alcoholic drinks*', as well as the more general label '*Increase VAT*' (Value Added Tax), motions may have been included in more than one *policy* category. The final dataset includes 13 different *policy* topic labels, with each applied to a minimum of 24 and a maximum of 129 motions ($\mu = 46.6$). 14 of the motions have two labels, while the remaining 578 have just one.

In addition to the Public Whip's crowdsourced

labels, we provide a second set of manually annotated opinion polarity labels. For these, annotation was conducted by the first author of this paper, who read each example (motion, title, and supplementary information), and applied either *positive* or *negative* labels according to the opinion they perceived to be expressed towards the *policy* in question.

As potential machine classification features, we include the textual content of the motions as well as the following metadata information from the transcripts:

- *motion_speaker_name*: Some MPs are more or less likely to speak on various topics, depending on their interests and position.

- *motion_party*: Party affiliation of speakers is likely to be an indicator of both interest in topics and policy positions.

- *debate_title*: Titles are often, but not always, related to policy vote topics.

- *additional_information*: Information such as the names of relevant documents or explanations of amendments is often included in the transcripts, preceding the motion.

Motions in this dataset broadly follow one of the following three formats:

1. 'That this House {*verb*} {*argument*}; {*verb*} {*argument*}; ... and {*verb*} {*argument*}'
   –where the motion may contain several clauses (see examples in Figures 1 and 2).

2. That the {*legislation*} be now read {a Second/the Third} time.
   –where {*legislation*} is a Bill, Paper etc.

3. ...amendment {*number*}, page {*number*}, line {*number*}, leave out '{'*phrase*'}' and insert '{*phrase*}'.

Motions of type 2 and 3 contain very little topic information, so it may be necessary to make use of cues in the debate title or additional information provided in the transcript in order to determine the topic in such cases.

4 Method

Data pre-processing consisted of removal of stop-words, lowercasing and stemming of textual data, and binarization of metadata information.

In order to detect the topics of debate motions we employ a supervised machine classification approach. For this, we investigate the use of combinations of the following features:

— Textual features: uni-, bi-, and trigrams from the debate titles, motions and supplementary information.
— Metadata features: speaker name and party affiliation.

As some motions have more than one topic label, we apply one-vs-the-rest classification on a randomised 90-10% train-test split of the data. After initial experimentation with a range of algorithms, we apply a multilabel implementation of Support Vector Machine classification.

5 Results

Because we have 13 different classes, and therefore highly imbalanced datasets for each round of one-vs-the-rest classification, we use the F1 score as a performance metric. Strongest performance is achieved using n-gram features from both the debate motions and titles (F1 = 77.0).
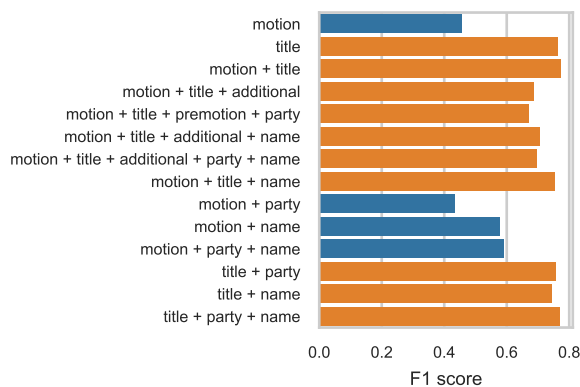

Figure 3: F1 score for different combinations of features used for topic classification. Highest performances (lighter, orange bars) are achieved when textual features derived from the debate titles are included.

Overall, use of the debate titles, with or without metadata features, produces the highest F1 scores, while the addition of other textual features does not generally lead to improvement, and in some cases results in losses in performance.

The motions themselves do not appear to provide particularly useful features for topic detec-

tion. Many consist solely of procedural terms that give no indication of the topics under discussion–such as motion types 2 and 3 (described in Section 3). Indeed, only 121 (20.8%) motions are of the more informative type 1.

Of the metadata features used, speaker name is more indicative of topic than party affiliation. This reflects the fact that each party is represented in most *policy* categories, but that individual MPs tend to be strongly associated with just a few, or in most cases, one single topic related to their particular role–of 234 MPs in the dataset, 163 (69.7%) propose motions on only one *policy*, and only one is represented in more than four.
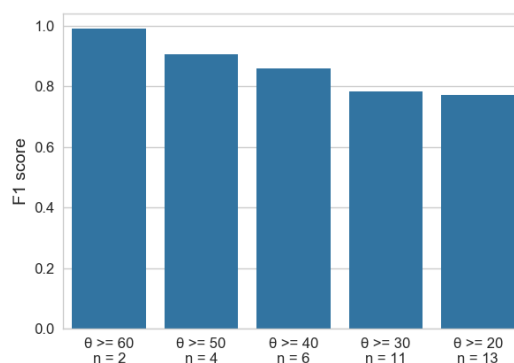

Figure 4: F1 scores for classification with different thresholds (θ) for the minimum number of example motions. As this threshold is lowered, the number of topic classes (n) increases and performance decreases.

As the threshold for the minumum number of examples per *policy* in the dataset is somewhat arbitary, we also test the system with a range of different thresholds. As the threshold decreases and the number of different topic classes increases, the F1 score drops, indicating that it may be challenging to obtain good results with a larger corpus and a greater number of topics (Figure 4).

6 Discussion

6.1 Topic detection

Considering the small number of training examples for each class, reasonable results are obtained using these labels for topic classification. However, it should be noted that many of the *policy* classes in this dataset feature debates with similar or even identical titles, in which cases the classifier is trained and tested on very similar data. While this is a common scenario in Parliament–the same pieces of legislation are debated mutiple times and often revisited year after year–it remains to be seen

how well this system would perform on new, completely unseen examples from future debates.

6.2 Opinion polarity analysis

As Public Whip Policies are created with inbuilt policy positioning, we examine their use as opinion labels by comparing their polarity with the second set of manual annotations. We ignore cases labelled in the Public Whip with the *policy vote* 'abstain' (as these are assumed not to take a position towards the *policy* in question). We then treat the 'majority' motions as being labelled according to the vote outcome–those which were 'approved' by the vote are *positive*, while those which were 'rejected' are *negative*–and the 'minority' labels as polarity shifters–that is, the tag 'minority' reverses the label derived from the outcome, while a 'majority' tag preserves it (see Table 1).

	Outcome	Policy Vote	Opinion
Policy	Approved	'majority'	*positive*
		'minority'	*negative*
	Rejected	'majority'	*negative*
		'minority'	*positive*

Table 1: Interpretation of Policy labels for opinion analysis. For each of its *policy* labels, a motion also has two tags–*outcome* and *policy vote*–that can potentially reverse its opinion polarity.

To examine the utility of the output labels, we calculate inter-rater agreement between these and our own annotations, finding Cohen's kappa (κ) to be 94.2. This represents 'near-perfect' agreement,[6] indicating that the Public Whip's *policies* appear to be reliable labels for opinion position of motions towards the policies in question. Although these results are promising, it should be noted that the system used to interpret motion opinion from *policies* relies on the use of additional, manually applied *policy vote* tags. For use with future, unseen examples that do not have such tags, it would be necessary to reorganise the way that the Public Whip's *policies* are created, splitting those labelled 'majority' and 'minority' into different *for* and *against* Policy categories.

7 Related work

The legislative debates domain has attracted interest from researchers with a variety of backrounds, and there is a considerable body of work related

---
[6]Interpretation of κ: (Landis and Koch, 1977).

to the analysis of both topics and speaker opinion contained in parliamentary and congressional debates, although these tasks have been tackled separately and from differing research perspectives.

For opinion analysis of US congressional debates, the dataset of Thomas et al. (2006) has been widely used (e.g. Balahur et al., 2009; Burfoot et al., 2011), and similar experiments have also been conducted on other legislatures such as the Dutch parliament (Grijzenhout et al., 2010), and the UK House of Commons (Salah, 2014).

Others have utilised similar techniques to facilitate other tasks. For example, Duthie et al. (2016) attempt to identify the 'ethos' of speakers in the UK Parliament, while Li et al. (2017) detect political ideology in those of the US Congress. Meanwhile, political scientists, such as Proksch and Slapin (2010) and Lauderdale and Herzog (2016) have analysed debates to position speakers on a range of scales related to policy and ideology.

While most work on this domain focuses on speeches, ignoring the role of motions in shaping the content of debates, Abercrombie and Batista-Navarro (2018) include analysis of the sentiment expressed in debate motions. However, they do not analyse the topics or identify the targets of sentiment in the motions.

Analysis of the topics contained within legislative debates has primarily focused on topic modelling based on speech content. For example, van der Zwaan et al. (2016) combine topic and political position analysis on Dutch parliamentary speech transcripts, while Zirn (2014) do similar for the German Bundestag. As far as we are aware, there exists no previous work on extracting topics from debate motions.

8 Conclusion

Opinion-topic labels derived from the Public Whip's *policies* can be used to train a classifier to achieve good performance in classifying the *policy* topics of parliamentary debate motions. These categories, which incorporate inbuilt policy position information, also appear to be reliable markers of opinion polarity, suggesting that we can use these labels to simultaneously obtain opinion analysis 'for free'. However, the ability of this approach to deal with new examples from future, unseen debates is uncertain, and it may be advisable to explore unsupervised methods of determining motion opinion-topics.

References

Gavin Abercrombie and Riza Batista-Navarro. 2018. '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)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Alexandra Balahur, Zornitsa Kozareva, and Andrés Montoyo. 2009. Determining the polarity and source of opinions expressed in political debates. In *Computational Linguistics and Intelligent Text Processing. CICLing 2009. Lecture Notes in Computer Science*, pages 468–480. Springer.

Clinton Burfoot, Steven Bird, and Timothy Baldwin. 2011. Collective classification of congressional floor-debate transcripts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1506–1515. Association for Computational Linguistics.

Rory Duthie, Katarzyna Budzynska, and Chris Reed. 2016. Mining ethos in political debate. In *Computational Models of Argument: Proceedings from the Sixth International Conference on Computational Models of Argument (COMMA)*, pages 299–310. IOS Press.

Steven Grijzenhout, Valentin Jijkoun, Maarten Marx, et al. 2010. Opinion mining in Dutch Hansards. In *Proceedings of the Workshop From Text to Political Positions, Free University of Amsterdam*.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 4(1):159–174.

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

Xilian Li, Wei Chen, Tengjiao Wang, and Weijing Huang. 2017. Target-specific convolutional bi-directional lstm neural network for political ideology analysis. In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data*, pages 64–72. Springer.

Sven-Oliver Proksch and Jonathan B Slapin. 2010. Position taking in European Parliament speeches. *British Journal of Political Science*, 40(3):587–611.

Zaher Salah. 2014. *Machine learning and sentiment analysis approaches for the analysis of parliamentary debates.* Ph.D. thesis, University of Liverpool.

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.

Cäcilia Zirn. 2014. Analyzing positions and topics in political discussions of the German Bundestag. In *Proceedings of the ACL 2014 Student Research Workshop*, pages 26–33. Association for Computational Linguistics.

Janneke M van der Zwaan, Maarten Marx, and Jaap Kamps. 2016. Validating cross-perspective topic modeling for extracting political parties' positions from parliamentary proceedings. In *Proceedings of ECAI: 22nd European Conference on Artificial Intelligence*, pages 28–36. IOS Press.