Using Argumentative Semantic Feature for Summarization

by

Venkata Krishna Rohit Sakala, Manish Shrivastava

in

5TH INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATION
(*ICSC-2019*)

Report No: IIIT/TR/2019/-1

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
March 2019

Using Argumentative Semantic Feature for Summarization

Sakala Venkata Krishna Rohit
IIIT Hyderabad
India
Email: rohit.sakala@research.iiit.ac.in

Dr Manish Shrivastava
IIIT Hyderabad
India
Email: m.shrivastava@iiit.ac.in

***Abstract*—The last decade has witnessed digitization of many government organization's data. Text summarization of the political discourse, particularly the parliamentary proceedings is relatively a lesser explored area of research. In this paper, we investigate the role of semantics especially theory of argumentation in debate summarization and use it to design a semi automatic pipeline for generating these summaries. The proposed approach considers the topic-relevance, argumentative nature, sentiment and context features. We test our approach on the dataset of debates mined from Lok Sabha, the elected house of representatives in India. Our proposed methodology and pipeline show significant improvement over the high performing popular systems for ROUGE-1, ROUGE-2 and ROUGE-L metrics.**

I. Introduction

As democracy seeps deeper into humanity, parliaments across the world are using the internet to propagate its debates and proceedings. This data is vastly used by think tanks, journalists and people to understand the dynamics of their democracy. These intricate and verbose debates need comprehensive summaries.

Each proposed legislation is debated hotly and contested by the members of the house. They provide various arguments in support of their statements and the arguments are refuted by opposition party members with their counter-arguments. This discourse is particularly tricky for standard NLP techniques because of the use of sarcasm, interjections and allegations [1]. Thus, our dataset is ideally suited for a task such as this.

A. Background

Wikipedia's definition of a summary is that it is a brief set of sentences of a text such as an article, thesis, conference proceeding or debate and is often used to help the reader quickly ascertain it's purpose. The process of generating summary through computational tools is called automatic summarization. According to Wikipedia, there are popular approaches to automatic summarization (1) Extractive Summarization (2) Abstractive Summarization. Extractive methods work by selecting a subset of existing words, phrases, or sentences in the original text to form the summary. In contrast, abstractive methods build an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a human might express. In this paper, we generate extractive summary through our pipeline.

Argumentation is an inter disciplinary research field which involves philosophy, communication science, logic, linguistics, psychology, and computer science. We consider argumentation as a semantic feature for summarization. Argument mining is relatively a new area of research in Natural Language Processing. Researchers have defined different argument schema and proposed computational models for it [2] [3]. We use the early proposed argumentation schema in our paper. [4]. In order to evaluate precisely, if argumentative theory acts as an important feature in the generation of extractive summarization, we do not opt for computational model instead we use the dataset annotated with the argument schema. [5] experimented the overlap between extractive summarization and argument mining to check if techniques used in summarization support argument mining. In our research, we work on the opposite side i.e to find out whether argument mining can help in efficient extractive summarization or not ?

1) Argumentation Schema: There are numerous proposals for modeling argumentation. We focus on monological models in this paper as they address the internal micro structure of arguments in a text and are well suited for developing computational methods. They focus on the function of argument components,the links between them, and the reasoning type [2] . In the schema that we have used for annotation of the dataset, an argument is composed of two entities mainly a claim and a premise. A claim is either supported or attacked by at least one premise. The claim is the central component of an argument. It is a controversial statement that should not be accepted by the reader without additional support. The discourse structure of arguments is modelled by the argumentative relations. The argument components are related and constitute the structure of argumentative discourse [6]. We use two relations (1) supports and (2) attacks in our schema. An argument diagram is a node-link diagram whereby each node represents an argument component, i.e. a claim or a premise and each link between the nodes represents a directed argumentative relation. These nodes and links together form a graph for a debate which have multiple connected components. The theory and the schema are well explained in [2].

The novelty of our pipeline is that we use this argument schema annotated graph for generating extractive summary in contrast to the algorithms proposed previously [7] [8] [9] in the online debate summarization.

The contributions of this paper are that we verify if argumentative schema acts as an important feature in the generation of extractive summarization in the case of parliamentary debate summarization. NMF [10] topic model performs better than other topic modelling techniques in the task of summarization as stated in [11]. This research also encourages other researchers to apply argumentative theory in various NLP tasks.

The rest of the paper is organized as follows: Section 2 describes related work; Section 3 gives a detailed description of our approach; Section 4 describes the results and Section 6 gives the conclusion and future work.

II. Related Work

Textual summarization has seen pioneering research in the recent years. The research and techniques varied depending on the typology of the textual context.

A. Document Summarization

[12] worked on two unsupervised methods for opinion summarization: first method being scoring sentences based on topic and document relevance and the second being a graphical approach to model the dialogue structures to find salient utterances in conversations.

B. Opinion Summarization

[13] worked on opinion summarization for product reviews. [7] and [12] published on the notion of summarization in online debates forums. They generated extractive summaries by giving scores to dialogue units using topic, document, sentiment and context relevance features. [8] developed a point-based summarization technique, where in a point is formed using various frames of dependency graph. [9] developed on [8]'s idea and improved the system by adding more filters such as topic relevance and point curation methods.

C. Debate Summarization

[14] has tackled online debate summarization challenge using clustering techniques such as term-based clustering and X-means. [15] used HAN (Hierarchical Attention Networks) [16] for thread summarization using annotated dataset of 600 threads of Trip Advisor. The model developed by them learns sentence representation effectively by attending to important words and also learns thread representation by attending to important sentences in the thread using encoders. By this, we conclude that summarization using argumentative theory is still unexplored area especially in the case of debate summarization.

D. Parliamentary Proceedings

[17] gives an overview of the parliamentary records and corpora from countries of Europe with a focus on their availability through their website. [18] created a highly multilingual parallel corpus of European parliament and demonstrated that it is useful for statistical machine translation. [19] worked to find out connections between different committees in the parliament house using network theory. [20] addressed if opinion mining techniques can be used on Congressional debates. [21] worked on creating a debate graph where in speakers are the nodes and exchanges between the speakers are edges for Hansard debates (England Parliamentary Debates) through which one can cluster members into groups. [21] worked on calculating sentiment polarity using machine learning approaches and SentiWordNet [22]. [23] worked on claim detection from UK political debates using linguistic and speech features. [24] worked on detecting perspectives in UK political debates using a Bayesian modelling approach. With this, we conclude, not much research has been done in the area of summarization of parliamentary debates.

III. Approach

Our pipeline contains 4 stages typically which are divided into sub sections below. In each stage, sentences are filtered based on the logic of that stage and then the selected sentences are given as input to the next stage. The input to the pipeline will be the annotated debate text and the length variable. The output will be an extractive summary of the length specified in the input. We also show how sentences of a snippet are filtered in each stage of the pipeline to be included in the summary.

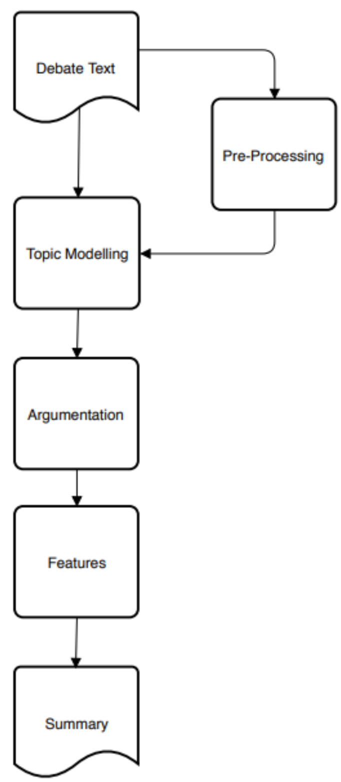

Fig. 1. System Pipeline with all stages.

A. Debate Text

As specified, we use the dataset of debates from Lok Sabha, the elected house of representatives of the Indian Parliament. The data that we used is part of the dataset developed by

[25]. The dataset has been annotated according to the schema explained by an expert using brat, an annotation tool [1]. The annotation has been done at the clause-level. We have randomly used 5 debates which are the discussions on various bills introduced in Lok Sabha house of Indian Parliament.

Consider the input as a snippet as shown below in which the premises are shown in italics and claims in bold text. In this particular snippet the sentences (2,3), (2,4), (4a,5) have support relations.

> (1) The Goods and Services Tax was introduced by a legislative process by which we first amended the Constitution of India. (2) *But then because of the provisions of Article 370, the Constitution Amendment was applicable to the whole country except to the State of Jammu & Kashmir.* (3) **Therefore, Jammu & Kashmir, under their Constitution, had to separately undergo a legislative process.** (4) **Jammu & Kashmir went through that legislative process and finally passed a Resolution in their State Assembly** and (4a) *then brought in a relevant legislation by virtue of which they passed their own State GST Act.* (5) However, **corresponding changes had to be made in the Central GST Act and the Integrated GST Act.**

B. Pre-Processing

This stage is only done for the purpose of topic modelling. We did not want to the topics to include stop words, prepositions or conjunctions. So , we pre-process the text in which we remove stop words, stem words so that all words are in their lemmetized form and then remove prepositions, conjunctions using NLTK python library.

C. Topic Modelling

Topic modelling is used in many summarization techniques as a filter and proves to be reliable as mentioned in [9] [7].The processed data is then passed through various topic models. If a topic word is present in the sentence, then only the sentence is selected for the next stage.

The topic models we use are:

1) **Non-negative Matrix Factorization** Non negative matrix factorization (NMF) approximates a non negative matrix by the product of two low-rank non negative matrices. Since it gives semantically meaningful result that is easily interpretable in clustering applications, NMF has been widely used as a clustering method especially for document data, and as a topic modeling method [26].
2) **Latent Dirichlet Allocation** LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities. [27].
3) **TextRank** TextRank is an unsupervised graph based ranking model for keyword and sentence extraction. TextRank does not require deep linguistic knowledge, nor domain or language specific annotated corpora, which makes it highly portable to other domains, genres, or language [28].
4) **Latent Semantic Analysis** In the case LSA, the core idea is to generate two matrix one document-topic matrix and another topic-word matrix and do dimensional reduction using SVD. These matrices are used for fetching topics. [29].

At this stage, the snippet contains only (2),(3),(4),(4a) and (5) sentences as only those sentences have the topic words. For this snippet, we consider NMF topic modelling.

> (2) *But then because of the provisions of Article 370, the Constitution Amendment was applicable to the whole country except to the State of Jammu & Kashmir.* (3) **Therefore, Jammu & Kashmir, under their Constitution, had to separately undergo a legislative process.** (4) **Jammu & Kashmir went through that legislative process and finally passed a Resolution in their State Assembly** and (4a) *then brought in a relevant legislation by virtue of which they passed their own State GST Act.* (5) However, **corresponding changes had to be made in the Central GST Act and the Integrated GST Act.**

D. Argumentation

In this stage, we describe how we filter the sentences after topic modelling. As explained in the introduction, each debate forms a graph structure with nodes being claims and premises, and links being the relation between them. Figure 3 and 4 are examples of how a connected component looks like. A debate will have many connected components.

Here, we are faced with the problem of which node/sentence is to be included in the summary. We develop a heuristic in which we first consider sentences or nodes which have both the relations i.e support and attack in a connected component. If there aren't enough connected components which have both attack and support relations, we consider the sentences in the connected component which have support relations. We used this approach as humans consider both the negative and positive aspects of a topic while writing summaries.

At this stage, the snippet consists of only (2), (3) and (4) sentences.

> (2) *But then because of the provisions of Article 370, the Constitution Amendment was applicable to the whole country except to the State of Jammu & Kashmir.* (3) **Therefore, Jammu & Kashmir, under their Constitution, had to separately undergo a legislative process.** (4) **Jammu & Kashmir went through that legislative process and finally passed a Resolution in their State Assembly.**

[1] http://brat.nlplab.org/

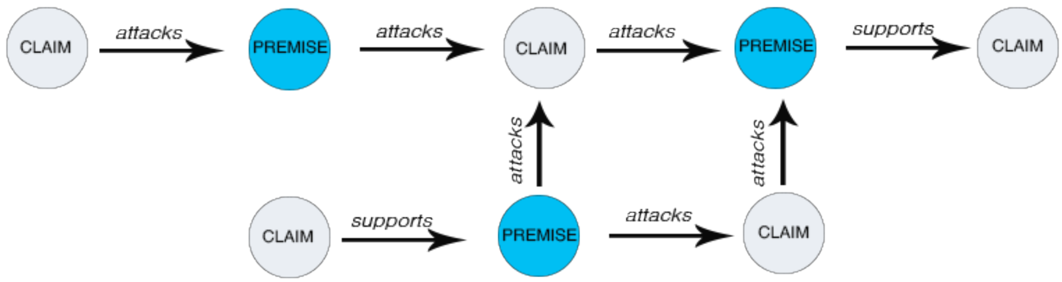

Fig. 2. Example connected component of the graph of a debate.

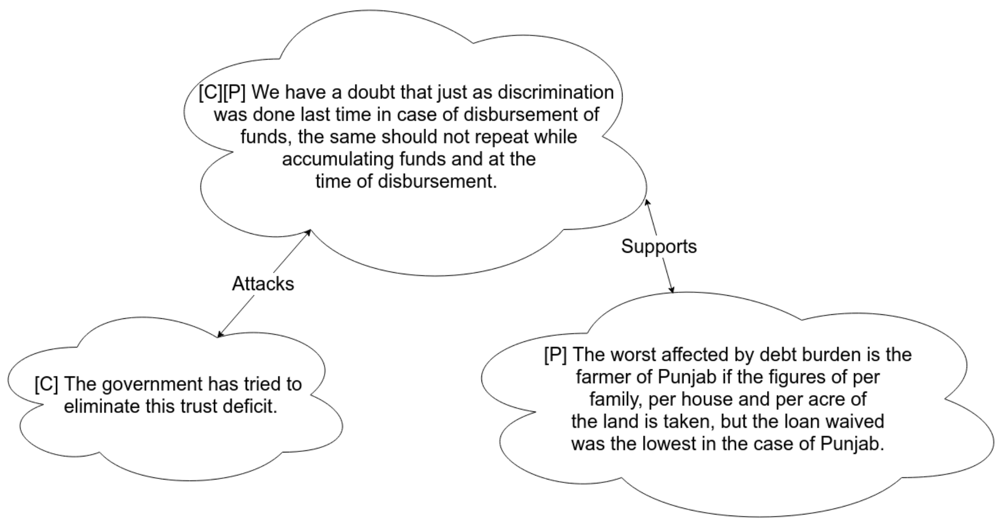

Fig. 3. Example connected component of the graph of a debate.

E. Features

Additionally we use sentiment and contextual features which are proven to be effective in summarization techniques [7].

1) Sentiment Score: We use VADER [30] to calculate the sentiment intensity of a sentence. If the sentence has high intensity score, we include it the list of sentences for the next stage. Sentiment rich sentences have proven be an effective feature in the extractive summarization techniques [12].

2) Sentence Length: In this stage, we remove sentences which have less than 5 words. This feature was used as sentences with longer length have more information. [7].

3) Duplicate Sentences: In this sub stage, we remove the duplicate sentences using word2vec [31] features of a sentence. Cosine similarity score is calculated for each pair of sentences and if the score is greater than a threshold they are considered to be duplicate. In the snippet, sentences (2) and (3) have high similarity score and so the longest sentence i.e (2) is considered.

F. Summary

This is the last stage where in the generated summary of desirable length is presented to the user. The length variable can take two values 1000 and 1500. Usually, a summary is 15 to 30 percent of the original text. As all our debate texts are approximately of the length 5000 words, we came up with those values.

Finally, in the snippet, the sentences (2) and (4) are included.

> (2) *But then because of the provisions of Article 370, the Constitution Amendment was applicable to the whole country except to the State of Jammu & Kashmir.* (4) **Jammu & Kashmir went through that legislative process and finally passed a Resolution in their State Assembly**.

IV. Results & Analysis

We have used 5 debates for the purpose of evaluation. Two students of philosophy were asked to write summaries of 2 different lengths. We calculated the inter editor-agreement as specified in [7] which came to 72.3 % . The agreement was based on the number of semantically similar sentences across the total dataset. Semantic similarity was calculated using word2vec [31] with a threshold cosine similarity value of 0.7. The idea of semantic similarity was inspired from [32]. We used ROUGE [33] as an evaluation metric which calculates a score, given the system generated summary and the gold summary as inputs.

We used the following three popular high performing summarization algorithms to compare with our system:

1) **TextRank** In this system, the sentences are ranked based on graph techniques and then the summaries are generated composed of highly ranked sentences. [28].
2) **LexRank** LexRank is an unsupervised approach to text summarization based on graph-based centrality scoring of sentences. [34].
3) **Nenkova** In this system, the summaries are generated using using three factors related to frequency: content word frequency, composition functions for estimating sentence importance from word frequency, and adjustment of frequency weights based on context. [35].
4) **W/O Argumentation Feature** We have evaluated the system without the argumentative feature to check the relevance of the argumentative feature in the pipeline. We have selected sentences in a random manner instead of the argumentative feature.

The following tables show the F-scores of the ROUGE-1,ROUGE-2 and ROUGE-l for 1000 and 1500 word sumarries.

System	ROUGE-1	ROUGE-2	ROUGE-L
TextRank	0.34	0.13	0.30
LexRank	0.30	0.15	0.30
Nenkova	0.36	0.12	0.33
Our System W/O Arg	0.35	0.13	0.31
Our System	**0.36**	**0.15**	**0.33**

TABLE I
ROUGE SCORES OF AVERAGE F-MEASURE OF SYSTEM SUMMARIES - 1000 WORDS

System	ROUGE-1	ROUGE-2	ROUGE-L
TextRank	0.39	0.17	0.34
LexRank	0.38	0.18	0.34
Nenkova	0.40	0.16	0.37
Our System W/O Arg	0.41	0.19	0.39
Our System	**0.44**	**0.21**	**0.41**

TABLE II
ROUGE SCORES OF AVERAGE F-MEASURE OF SYSTEM SUMMARIES - 1500 WORDS

As you can see, our system reports higher values for all ROUGE metrics. This verifies that argument components act as an important feature in summarization.

We also analyze which topic model performs better in this particular pipeline. According to the table below, NMF performs better than other topic models as it captured different viewpoints. [11] also reports NMF performing better than LDA.

Topic Model/ROUGE-L	1500
TextRank	0.39
LSA	0.39
LDA	0.40
NMF	**0.41**

TABLE III
ROUGE SCORES OF AVERAGE F-MEASURE OF SYSTEM SUMMARIES FOR 1500 WORD SUMMARY FOR DIFFERENT TOPIC MODELS.

We also experiment different heuristics as in which nodes to consider in the graph of a debate. DASO is the method we mentioned in the approach section. In Low degree nodes, we consider all nodes in an ascending manner until the required length is achieved. In All support relations method, we consider all nodes which have only support relation.

System	ROUGE-1	ROUGE-2	ROUGE-L
All Support Relations	0.33	0.14	0.30
Low Degree Nodes	0.35	0.14	0.32
DASO	**0.36**	**0.15**	**0.33**

TABLE IV
F-SCORE VALUES OF ROUGE METRICS OF DIFFERENT GRAPH EXPERIMENTS FOR A 1000 WORD SUMMARY.

V. CONCLUSION & FUTURE WORK

In this research, we found out that a semantic feature such as an argumentative theory plays an important role for creating summaries in the case of parliamentary proceedings. We achieved significantly higher ROUGE scores when compared to popular high performing extractive summarization techniques. We used the dataset of Indian Parliament for conducting this research.

We plan to add more features to the pipeline to further increase the results. As specified above, this approach requires annotated dataset. So, this system becomes semi-automatic. In the future, we will investigate if we can get similar results using computational models with the help of features used in [6]. We also plan to give a suitable layout for the summary of the parliamentary debates as proposed in [8]. In future, we would like to model this problem as an optimization technique for getting the best results keeping various stages of pipelines as variables. We plan to explore more graph based algorithms which can further improve the scores. We also plan to test this approach on multiple datasets to check the efficiency of this approach.

REFERENCES

[1] O. Onyimadu, K. Nakata, T. Wilson, D. Macken, and K. Liu, “Towards sentiment analysis on parliamentary debates in hansard,” in *Joint International Semantic Technology Conference*. Springer, 2013, pp. 48–50.

[2] C. Stab and I. Gurevych, "Parsing argumentation structures in persuasive essays," *CoRR*, vol. abs/1604.07370, 2016. [Online]. Available: http://arxiv.org/abs/1604.07370

[3] I. Persing and V. Ng, "End-to-end argumentation mining in student essays," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1384–1394.

[4] R. Mochales and M.-F. Moens, "Argumentation mining," *Artificial Intelligence and Law*, vol. 19, no. 1, pp. 1–22, 2011.

[5] G. Petasis and V. Karkaletsis, "Identifying argument components through textrank," in *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, 2016, pp. 94–102.

[6] C. Stab and I. Gurevych, "Identifying argumentative discourse structures in persuasive essays," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 46–56.

[7] S. Ranade, J. Gupta, V. Varma, and R. Mamidi, "Online debate summarization using topic directed sentiment analysis," in *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*. ACM, 2013, p. 7.

[8] C. Egan, A. Siddharthan, and A. Wyner, "Summarising the points made in online political debates," in *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, 2016, pp. 134–143.

[9] A. D. Chowanda, A. R. Sanyoto, D. Suhartono, and C. J. Setiadi, "Automatic debate text summarization in online debate forum," *Procedia Computer Science*, vol. 116, pp. 11–19, 2017.

[10] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2013.

[11] P. Sobhani, D. Inkpen, and S. Matwin, "From argumentation mining to stance classification," in *Proceedings of the 2nd Workshop on Argumentation Mining*, 2015, pp. 67–77.

[12] D. Wang and Y. Liu, "A pilot study of opinion summarization in conversations," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011, pp. 331–339.

[13] K. Ganesan, C. Zhai, and E. Viegas, "Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 869–878.

[14] N. Sanchan, A. Aker, and K. Bontcheva, "Automatic summarization of online debates," *arXiv preprint arXiv:1708.04587*, 2017.

[15] S. Tarnpradab, F. Liu, and K. A. Hua, "Toward extractive summarization of online forum discussions via hierarchical attention networks," *arXiv preprint arXiv:1805.10390*, 2018.

[16] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480–1489.

[17] D. Fišer and J. Lenardic, "Parliamentary corpora in the clarin infrastructure," in *Selected papers from the CLARIN Annual Conference 2017, Budapest, 18–20 September 2017*, no. 147. Linköping University Electronic Press, 2018, pp. 75–85.

[18] N. Hajlaoui, D. Kolovratnik, J. Vyrynen, R. Steinberger, and D. Varga, "Dcep -digital corpus of the european parliament," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), may 2014.

[19] M. A. Porter, P. J. Mucha, M. E. Newman, and C. M. Warmbrand, "A network analysis of committees in the us house of representatives," *Proceedings of the National Academy of Sciences*, vol. 102, no. 20, pp. 7057–7062, 2005.

[20] M. Thomas, B. Pang, and L. Lee, "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*. Association for Computational Linguistics, 2006, pp. 327–335.

[21] Z. Salah, F. Coenen, and D. Grossi, "Extracting debate graphs from parliamentary transcripts: A study directed at uk house of commons debates," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*. ACM, 2013, pp. 121–130.

[22] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining." in *Lrec*, vol. 10, no. 2010, 2010, pp. 2200–2204.

[23] M. Lippi and P. Torroni, "Argument mining from speech: Detecting claims in political debates." in *AAAI*, 2016, pp. 2979–2985.

[24] D. Vilares and Y. He, "Detecting perspectives in political debates," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 1573–1582.

[25] S. V. K. Rohit and N. Singh, "Analysis of speeches in indian parliamentary debates," *arXiv preprint arXiv:1808.06834*, 2018.

[26] D. Kuang, J. Choo, and H. Park, "Nonnegative matrix factorization for interactive topic modeling and document clustering," in *Partitional Clustering Algorithms*. Springer, 2015, pp. 215–243.

[27] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[28] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004.

[29] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, *Handbook of latent semantic analysis*. Psychology Press, 2013.

[30] C. H. E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*, 2014.

[31] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[32] N. Sanchan, A. Aker, and K. Bontcheva, "Gold standard online debates summaries and first experiments towards automatic summarization of online debate data," *arXiv preprint arXiv:1708.04592*, 2017.

[33] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, 2004.

[34] G. Erkan and D. R. Radev, "Lexrank: Graph-based lexical centrality as salience in text summarization," *Journal of artificial intelligence research*, vol. 22, pp. 457–479, 2004.

[35] A. Nenkova, L. Vanderwende, and K. McKeown, "A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2006, pp. 573–580.