Computational Frame Analysis:

A Singapore Study

Kien Trung Khuong  
*Department of Computer Science*  
*University of Exeter*Exeter, UK  
tk402@exeter.ac.uk

Internal Supervision: Chico Camargo  
*Department of Computer Science*  
*University of Exeter*Exeter, UK  
f.camargo@exeter.ac.uk

*Abstract*—Framing is prioritizing some elements of a reality and strengthen their salience in such a way as to promote a particular interpretation [1]. Framing can be used by politicians in their speeches to shape public opinions or by journalists in their articles to influence readers. This study aims to apply the latest tools and techniques in Machine Learning and Natural Language Processing to detect frames in Singapore parliamentary debates and whether there are any semantic changes in the meanings of words over time. In addition, because the opposition is under-represented in Parliament, I will investigate how the alternative media frame parliamentary debates and other local political events to determine different framing strategies employed by the government and its dissents.

Keywords—Natural Language Processing, Machine Learning, Frame Analysis, Singapore, Political Communication

# Introduction

The Parliament of Singapore, primarily modelled after the Westminster system, is responsible for “making laws, taking up a critical/inquisitorial role to check on the actions and policies of the Government and scrutinizing the State's finances” [2]. Therefore, it is utterly important to understand the wide-ranging topics discussed in Parliament if one wishes to follow Singapore’s development, study its public policy and its principles of governance.

In many cases, to advance their agendas, Members of Parliament (MP) must be selective when presenting their arguments: what to say, what not to say, and what to explain and highlight. This leads us to framing, a central concept in political communication. The most popular definition is “to frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described” by Robert Entman [3]. Consequently, it is widely accepted that “how issues are framed has a large impact on the way that people hear, interpret, and act on the different policy proposals” [4]. However, the detection of frames remains challenging as it is hard to identify a frame which is an abstract concept, and it has traditionally required a tedious and labor-intensive manual annotation process that is subject to human bias [5].

Fortunately, the availability of online data in recent years presents an unprecedented opportunity. Many existing tools and techniques in Natural Language Processing (NLP) and Machine Learning (ML) can now be employed to discover framing efficiently in a larger scale.

Since independence in 1965, Singapore has been ruled by one political party, the People’s Action Party (PAP). It was the only political party present in parliament from 1965 to 1981. Even now, the PAP still controls the Parliament, occupying 83 out of 93 seats. Once frames have been identified, it will be interesting to study their development across multiple years to understand their evolutionary process especially before and after the presence of opposition in parliament.

I will also investigate frames used by the Singapore press when covering topics discussed in Parliament. Nevertheless, because Singapore has a tightly controlled media environment, it is expected that Singapore mainstream media use similar frames when covering Parliamentary debates. As a result, frames used by alternative media will be identified and examined to form a better picture of framing methods employed by different sides of the political spectrum.

# Research Context

## Computational Frame Analysis

Thanks to the development of many advanced ML and NLP techniques, many studies on computational frame analysis have been conducted to apply these methods to identify frames, most of which focused on news media. Some attempted to use supervised models while few others proposed unsupervised approaches. One recent study introduced a semi-supervised model [6].

For supervised approaches, unfortunately, there are few available labelled training datasets. Those datasets include the ComArg corpus [7]–which covers the topics “gay marriage” and “under god in pledge”–the Media Frame Corpus [8] – which covers the topics of “smoking”, “same-sex marriage” and “immigration”–and the Gun Violence Frame Corpus [1] – which covers the topic of “gun violence”. Supervised models enjoyed some promising results. For example, [1] and [9] modified and fine-tuned a pre-train Bidirectional Encoder Representations from Transformers (BERT) model on MC and a set of New York Times articles respectively. However, there are some concerns when applying supervised learning in this project. First, while issues debated in parliaments, not just Singapore Parliament, can be very diverse, these datasets only cover few topics. Second, the idea of applying a model trained on a dataset like MFC to parliamentary debates is unlikely to succeed due to possible differences in framing strategies and/or words used as shown in [10]. Third, creating labelled training data is not feasible in this project due to labour and time constraints. As a result, a supervised approach to this project will not be considered. On the other hand, experiments with these algorithms are still necessary and will be carried out.

As for the semi-supervised approach, the study was conducted on the “immigration” portion of MFC. The model, called FRISS, outperformed existing models such as RoBERTa. However, this approach suffers the same issue as the supervised approach: it is limited to a pre-defined topic and hence, very unlikely to work on the parliamentary debates. However, it is worth investigating this proposal to determine if it is feasible to extend this model.

I will be focused on applying unsupervised ML techniques. A common idea is topic modelling, an unsupervised machine learning technique that aims to cluster word groups and similar expressions that best characterize a set of documents. These words represent the topic of those documents. For topic modelling, Latent Dirichlet Allocation (LDA) [11] is to go-to tool while others prefer the Structural Topic Model (STM) [12], which is similar to LDA but it allows us to incorporate meta data (author, date, etc) into our model to improve the assignment of words to latent topics in a corpus. Another approach called Analysis of Topic Model Networks (ANTMN) was proposed by [13] in which the authors demonstrated that “frame elements could be identified using topic modelling, and that frame elements can then be automatically grouped into frame packages using community detection techniques applied to the topic network”.

An interesting approach that I will explore is an extension of LDA called supervised hierarchical LDA (SHLDA) [5], which discovers “the context-independent effect of individual words” and “context-specific” effect of second-level agendas.

Another algorithm that I would like to explore is Anchored Correlation Explanation (ACE) [14]. ACE avoids many assumptions made by LDA and hence, “searches for topics that are *maximally informative* about a set of documents.” Moreover, this model can be “guided” to look for certain under-represented topics that are of interest.

## Computational Frame Analysis of Singapore Parliament debates, Singpore newspapers and alternative media

As far as I am aware, there have been no studies on frame analysis of Singapore parliamentary debates and alternative media, let alone a computational one. Although there are some studies on Singapore news frames such as [15] [16], they focused on a particular event such as a general election. Additionally, these studies involved political researcher who applied some statistical models on a very small, labelled dataset (429 news articles for [15] and 158 news releases for [16]). Should this project be successful, it will be the very first study about computational frame analysis of parliamentary debates, mainstream media, and alternative media in Singapore.

## Novel work

As discussed, there have been no previous works done on frame analysis of Singapore parliamentary debates and alternative media. However, according to [17], there are studies on other parliamentary text data such as US Congress and UK Parliament, some of which are related to frame detection such as [7]. There was a similar project on frame analysis on UK Parliament debates by Ben Dennes [18]. Therefore, it is possible to learn from these experiences when working on this project. However, unlike other parliamentary democracies, Singapore Parliament has its own unique features, most notably the Non-constituency Members of Parliament (NCMP) and Nominated Members of Parliament (NMP) and Singapore as an Asian country is different from the UK, a European country, making some novel works possible in this project.

# Aims and objectives

The research questions I will ask and the ways in which I

will attempt to answer them are outlined below:

*1. Can any of existing computational frame analysis methods be applied successfully to identify frames in Singapore Parliament debate transcripts, newspapers, and alternative media?* Experiments with many advanced ML and NLP techniques will be carried out and the results of these techniques will serve as the foundation for Research Aim 2.

*2. What modifications can be done to improve the performance of existing algorithms?* By studying results of Research Aim 1, I can investigate the feasibility of combining some successful components of multiple models or I can attempt to add some augmentations to existing models.

*3. How do frames evolve overtime? If there are changes, does it mean MPs or parties changed their attitudes or strategies?* Once a successful computational framing analysis method is developed, a method to consolidate individual MP’s contributions based on topics and frames will be implemented. After this, it is possible to observe and measure any changes in frames used by MPs and parties. It would be nice if all frames are visualized in one chart, possibly a timeline.

*4. Can we use the same model for newspapers and alternative media, or does it need to be retrained?* The model will be applied on news articles and its performance will be examined. If the answer is negative, it is likely that the model needs to be retrained because vocabularies used in newspaper are very different that those used in parliamentary debates.

*5. How different are frames used by ruling party MPs and opposition MPs? How about mainstream media and alternative media?* At this point, we can compare frames used by individuals who have different political perspectives. It would be fascinating to compile all information and perform political analysis to uncover some hidden truths.

# Data & Resources

## Data

The primary source of data for this project is *Official Reports - Parliamentary Debates (Hansard)* consisting of “speeches and debates made in the Parliament Chamber and provides a record of parliamentary business and proceedings in a Sitting” [19]. The platform allows users to query reports date as far back as 1955. Unfortunately, it does not have a “download” feature to produce documents in a structural format. Hence, a challenging task is to implement scraping scripts that save these reports in a better format. The platform is publicly accessible and can be used for free.

For mainstream media, I will gather news articles from *The Straits Times* (ST) and *Channel News Asia* (CNA), two dominant English newspapers in Singapore. The LexisNexis database contains articles from ST since 1992 and from CNA since 2000. A quick search “Singapore” reveals more than 867,000 articles from ST and 151,000 articles from CNA. These articles can be downloaded as RTF files. However, the database only allows downloading up to 500 articles each time. Scripts to automate this task needs to be implemented. The LexisNexis database requires subscription to access.

For alternative media, I am going to use *New Naratif*, “a movement for democracy, freedom of information and freedom of expression in Southeast Asia” [20], and *The Independent* that claims to “bring in-depth perspective and analysis on current affairs, economics and politics in Singapore” [21]. Both are typical online news media which require scraping scripts to save their content. For New Naratif, many of their materials are in the form of podcast (mp3 files). A speech-to-text technique will be applied to convert them to text. The Independent articles are free while New Naratif has some membership-only content. New Naratif’s membership costs US$ 5/month.

## Software Tools

Python together with ML and NLP packages such as ntlk, spacy, huggingface as well as other common data processing libraries like pandas and matplotlib will be used extensively in this project.

In addition, Javascript tools such as TimelineJS will be utilized to create interactive visualization in a more user-friendly browser environment.

# Methods & Experiment Design

## Perform exploratory data analysis

Before diving into any machine learning algorithms, it is a must to explore and understand the dataset. Debates and articles will be collected and saved in JSON format. I will start by analyzing text statistics. Furthermore, as part of exploratory data analysis, I will apply “Term frequency – inverse document frequency” (TF-IDF) to gauge the “favorite” terms of every MP.

## Experiments with existing NLP techniques

I will experiment with as many as possible existing NLP techniques: LDA, STM, SHLDA, ANTMN, ACE, BERT, FRISS, etc to understand how they works as well as their advantages and disadvantages. This task will be accompanied by continued extensive research of literature on framing.

## Develop a novel algorithm

Results in experiments with existing NLP techniques will provide me with clues and directions for introducing modifications to them or looking for ways to combine them to arrive at the “best-of-many-worlds” solution.

## Study the evolutionary process of frame development

Once frames have been identified successfully, it is necessary that their connections should be examined. This can be done by employing network theory.

## Adapt the model to newspapers articles

It might be necessary to retrain the model on newspapers articles as news articles are most likely written for layman while parliament debates are between highly educated individuals. This task might not be challenging but will be very time-consuming as the number of articles is much greater than parliament speeches.

## Perform political analysis

After all frames have been identified, it would be very interesting to analyze them: how the ruling party as well as the opposition has evolved, and generally, how Singapore society has evolved. It might be possible to visualize the whole development of Singapore public policies since independence.

# Data Governance and Ethics

Transcripts of Parliament debates used in this project will come from Official Reports - Parliamentary Debates (Hansard), which have been made available on the Internet since 2000, meaning these speeches are available for inspection to the entire population. ST and CNA articles can be obtained from LexisNexis under the institutional license held by the University of Exeter. The Independent articles are free while New Naratif free materials are collected. Any paid content, if needed, will be paid for, but I expect free ones should suffice.

All data collected will be used for research purposes only. Therefore, there should not be any ethical or legal concerns about using them in this project.

# Project Plan

The milestones for each of the deliverables stated in section V are as follows:

## Perform exploratory data analysis: 15 Apr

## Experiments with existing NLP models: 15 May

## Develop a novel algorithm: 30 Jun

## Study the evolutionary process of frame development: 7 July

## Adapt the model to newspapers articles: 14 July

## Perform political analysis: 30 July

# Risk Assessment

There are 2 key risks that might adversely affect the project:

1. Lack of computing resources: training a ML model on a vast dataset requires a significant amount of time, even on a decent laptop. The project may seriously fall behind schedules if it takes much more time to experiment with a model than previously estimated. One way to mitigate this issue is to use better and faster computing resources in terms of GPU and RAM. This can be done by using a virtual machine on Google Cloud.

2. Unsuccessful development of a novel framing analysis algorithm: in the case of unsuccessful attempts to develop a novel computational frame analysis, I will resort to using topic modeling to detect frames like [22]. It might not identify many possible frames but the fact that this is the first study on computational frame analysis of Singapore politics should be useful for future reference.

# Conclusion

In this project I aim to propose a novel technique of frame analysis by drawing on the experiences of previous studies on ML and NLP techniques. It will be applied to detect frames used by MPs in Singapore Parliament. By identifying these frames across multiple years, it will be possible to observe the evolution of their development. Furthermore, the algorithm will be applied to alternative media to paint a better picture of frames employed by the many sides of the political spectrum.

# References

|  |  |
| --- | --- |
| [1] | L. G. K. M. M. B. D. T. W. Siyi Liu, "Detecting Frames in News Headlines and Its Application to Analyzing News Framing Trends Surrounding U.S. Gun Violence," in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, Hong Kong, China, Association for Computational Linguistics, 2019, pp. 504-514. |
| [2] | "Functions," [Online]. Available: https://www.parliament.gov.sg/about-us/parliament-information/functions. [Accessed 22 3 2022]. |
| [3] | R. Entman, "Framing: Toward Clarification of a Fractured Paradigm," *Journal of Communication,* pp. 51-58, 1993. |
| [4] | E. a. M. National Academies of Sciences, Framing the Dialogue on Race and Ethnicity to Advance Health Equity: Proceedings of a Workshop, Washington, DC: The National Academies Press, 2016. |
| [5] | J. H. G. P. R. N. A. S. Amber E. Boydstun, "Identifying Media Frames and Frame Dynamics Within and Across Policy Issues," 2013. |
| [6] | S. a. C. T. a. M. G. a. T. A. a. F. L. Khanehzar, "Framing Unpacked: A Semi-Supervised Interpretable Multi-View Model of Media Frames," 2021. |
| [7] | N. NADERI, "Argumentation Mining in Parliamentary Discourse," in *OSSA Conference Archive*, Toronto, Ontario, Canada, 2016. |
| [8] | A. E. B. J. H. G. P. R. N. A. S. D. Card, "The Media Frames Corpus: Annotations of Frames Across Issues," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China, 2015. |
| [9] | J. A. Y.-Y. A. Haewoon Kwak, "A Systematic Media Frame Analysis of 1.5 Million New York Times Articles from 2000 to 2017," arXiv, 2020. |
| [10] | A. T. G. M. S. Khanehzar, "Modeling Political Framing Across Policy Issues and Contexts," in *Proceedings of The 17th Annual Workshop of the Australasian Language Technology Association*, Sydney, 2019. |
| [11] | A. Y. N. M. I. J. D. M. Blei, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.,* vol. 3, pp. 993-1022, 2003. |
| [12] | B. S. D. T. E. A. M. Roberts, "The Structural Topic Model and Applied Social Science," 2013. |
| [13] | Y. O. D. Walter, "News Frame Analysis: An Inductive Mixedmethod," *Communication Methods and Measures,,* vol. 13, no. 4, pp. 248 - 266, 2019. |
| [14] | K. R. D. K. G. V. S. Ryan J. Gallagher, "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge," *Transactions of the Association for Computational Linguistics,* vol. 5, 2017. |
| [15] | C. S. Y. ,. H. T. J. ,. L. L. Q. &. T. X. L. Mark Cenite, "Perpetual development journalism? Balance and framing in the 2006 Singapore election coverage," *Asian Journal of Communication,* vol. 18, no. 3, pp. 280-295, 2008. |
| [16] | S. T. Lee, "Predictors of H1N1 Influenza Pandemic News Coverage: Explicating the Relationships between Framing and News Release Selection," *International Journal of Strategic Communication,* 2014. |
| [17] | R. B. Gavin Abercrombie, "Sentiment and position‑taking analysis of parliamentary debates: a systematic literature review," *Journal of Computational Social Science,* 2019. |
| [18] | B. Dennes, "Finding Frames in Parliament: Computational Framing Analysis of UK House of Commons Speech," University of Exeter, Exeter, UK, 2021. |
| [19] | "Official Reports - Parliamentary Debates (Hansard)," [Online]. Available: https://sprs.parl.gov.sg/search/home. [Accessed 24 3 2022]. |
| [20] | "New Naratif," [Online]. Available: https://newnaratif.com/about-us/. [Accessed 24 3 2022]. |
| [21] | "The Independent Singapore News," [Online]. Available: https://theindependent.sg/about-us/. [Accessed 24 3 2022]. |
| [22] | T. &. E. V. &. K. A. Ylä-Anttila, "Topic Modeling as a Method for Frame Analysis: Data Mining the Climate Change Debate in India and the USA," 2018. |
| [23] | L. G. K. M. M. B. D. T. W. S. Liu, "Detecting Frames in News Headlines and Its Application to Analyzing News Framing Trends Surrounding U.S. Gun Violence," *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL),* pp. 504-514, 2019. |