

Predicting Political Frames Across Policy Issues and Contexts

Shima Khanehzar Andrew Turpin Gosia Mikolajczak

School of Computing and Information Systems

The University of Melbourne

skhanehzar@student.unimelb.edu.au

{aturpin, malgorzata.mikolajczak}@unimelb.edu.au

Abstract

Politically-contested issues are often discussed with different emphases by different people. This emphasis is called a *frame*. In this paper, we examine the performance of classifiers trained using the media frames Corpus (MFC) (Card et al., 2015); a collection of US news labelled with fifteen different frame categories. Specifically, we compare pre-trained language models (XLNet, Bert, and Roberta), fine-tuned using MFC, against results from the literature and simpler models in their ability to predict frames from text. We also test these models on a new corpus that we have derived from Australian parliamentary speeches. Our experimental results first show that the fine-tuned models significantly outperform the current best methods on MFC. We also show that the model fine-tuned on US news articles can be convincingly applied to predict policy frames in Australian parliamentary speeches, though the accuracy is significantly reduced, suggesting potential discrepancy in framing strategies and/or text usage between US News and Australian Parliamentary Speeches.

1 Introduction

Politicians and the media often portray political issues in a subjective way in an attempt to shape public attitudes (Chong and Druckman, 2007). For example, a politician opposing the same-sex marriage (SSM) might frame the issue using the lens of tradition and religious beliefs, whereas a politician supporting SSM might frame a speech using fairness and equality as the base. Due to its complexity and linguistic subtleties, issue *framing* (Entman, 1993) remains challenging for automated text methods. To address these challenges, recent work by Boydston et al. (2013) defines broad categories of common policy frames and annotates US News articles to build the media frames Corpus (MFC) (Card et al., 2015). Follow-

up studies have used the MFC to investigate the accuracy of models that attempt to classify the dominant frames of US news articles. In this paper, we aim to extend this work and answer the following question: can recent pre-trained neural classifiers learn to predict dominant frames across issues and communication contexts? To answer this question, we provide the following contributions.

- We investigate the effectiveness of the pre-trained language models XLNet, Bert and Roberta in predicting dominant frames within each issue on the MFC.
- We investigate whether our models can learn to predict frame categories across issues. Our results show that we can apply trained models on a new issue without training data for that particular issue.
- We annotate a small subset of Australian parliamentary speeches on Same-Sex Marriage (SSM).
- We evaluate whether our models can learn to predict frames across communication contexts, applying the models fine-tuned on the MFC dataset on the Australian parliamentary speeches.

2 Background and Related Work

Natural Language Processing techniques have been applied to identify several aspects of the political discourse including ideology (Iyyer et al., 2014), sentiment (Godbole et al., 2007; Balahur et al., 2010), and stance (Mohammad et al., 2016).

Earlier studies focusing specifically on frame detection usually¹ employ topic modeling (Boydston et al., 2013), (Nguyen, 2015), (Tsur et al.,

¹for an exception, see Baumer et al. (2015) who use classifiers to identify the language of framing in the news

2015). This approach allows for automated detection of frames within specific corpora, but does not easily allow results and methods to be used across issues or contexts that are not part of the corpus on which the model is built. To address this shortcoming, [Boydston et al. \(2013\)](#) proposed a list of 15 broad frames (e.g., Economic, Morality, or Legal; plus an “Other” category) commonly used when discussing different policy issues (such as abortion, immigration, foreign aid, etc.), and in different communication contexts (news stories, Twitter, party manifestos, legislative debates, etc.). The frames have been defined in the [Policy Frame Codebook \(PFC\)](#)

The **The Media Frames Corpus (MFC)** [Card et al. \(2015\)](#) includes news articles from 13 U.S. newspapers, covering five policy issues: *same-sex marriage*, *immigration*, *tobacco*, *gun control*, and the *death penalty*, published between 1980–2012. Approximately 12,000 articles have been annotated with the dominant frame from the list of categories proposed in PFC. The annotations also identify exact text spans associated with each of the 15 frames. Since the frame distribution is imbalanced and not reported in the original paper, here we show the the statistical distribution of the frameworks in table 1.

The MFC has been previously used for training and testing classification models. For example, [Card et al. \(2016\)](#) provide an unsupervised model that clusters articles with similar collections of “personas” (i.e., characterisations of entities) and demonstrate that these personas can help predict the coarse-grained framing annotations in the MFC.

The current best result for predicting the dominant frame of each article in the MFC comes from [Ji and Smith \(2017\)](#), who proposed a recursive neural discourse structure network with a new attention mechanism of the text for text categorization. They report the average accuracy across 10-fold cross-validation using the immigration issue which we report in Table 2 (column 4).

[Field et al. \(2018\)](#) used the MFC to investigate agenda-setting and framing in Russian News. They introduced embedding-based methods for projecting frames of one language into another (i.e., English to Russian). It is worth mentioning that their approach is applicable to languages suffering from lack of training data.

3 Method

In this paper, we explore three general approaches to classify text with the frames from the PFC. First, We create baseline models with Support Vector Machine (SVM) and Weighted Support Vector Machine (Weighted-SVM). SVMs are often used for text classification problems, as the algorithms perform classification by finding hyperplanes to differentiate the classes. Weighted-SVM is often used for dataset with skewed distribution to reduce bias, and it is more suitable for MFC, which has an imbalanced class distribution. We implement SVM and Weighted-SVM using the default parameters in the sklearn python library. Second, we use the MFC to form a lexicon (bag of words) for each frame and classify new texts using the Okapi text similarity metrics ([Robertson and Zaragoza, 2009](#)) from each lexicon. Last, we employ pre-trained language models, and fine-tune them with the MFC. Since our primary goal is to investigate if framing shares similar patterns across domains, we evaluate these models across issues and contexts. For across-issue evaluation, we fine-tune our models on four issues from the MFC (i.e., excluding immigration), and then evaluate them on the immigration subset. For across-context evaluation, we evaluate the models on a subset of the Australian Parliamentary Speeches (APS), which we describe in more detail below.

3.1 Framing Lexicons

Based on the approach by [Field et al. \(2018\)](#), a lexicon related to each frame f in the PFC is derived by taking the top 50 words with the highest pointwise mutual information $I(f, w) = \log p(w|f) - \log p(w)$, where w is a word. We compute $P(w|f)$ by taking the number of occurrences of w in all the text segments annotated with the secondary frame f in the MFC divided by the total number of words in those segments. Quantity $P(w)$ is computed similarly over the entire corpus. As in [Field et al. \(2018\)](#), we discard all words that occur in fewer than 0.5% of documents or in more than 98% of documents.

In order to classify a document into one of the 15 frames, we take the highest ranked lexicon using the document as a query against a collection of the 15 lexicons, measuring similarity using Okapi scoring ([Robertson and Zaragoza, 2009](#)). We use the default parameters in the Okapi formula as implemented in the Gensim Python Library.

Frame Category	MFC SSM	MFC no SSM	MFC IM	MFC no IM	APS
Economic	136	1400	414	1122	0
Capacity and Resources	4	245	210	39	0
Morality	405	406	76	735	6
Fairness and Equality	196	653	155	694	24
Legality Constitutionality Jurisdiction	1173	3747	957	3963	9
Policy Prescription and Evaluation	178	1938	473	1643	2
Crime and Punishment	20	2167	803	1384	0
Security and Defence	1	609	286	324	0
Health and Safety	50	1330	239	1141	0
Quality of Life	294	790	410	674	6
Cultural Identity	298	1335	556	1077	5
Public Sentiment	364	758	243	879	11
Political	1215	3547	969	3793	35
External Regulation and Reputation	22	290	132	180	2
Other	0	11	10	1	0
All	4356	19226	5933	17649	100

Table 1: Frame statistics in MFC and APS used in our experiments.

3.2 Neural models

Bert (Devlin et al., 2019) is a bi-directional language model based on now ubiquitous Transformers (Vaswani et al., 2017) with a Cloze Test objective, and trained on a large text corpus. The pre-trained Bert model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. In this work, we add an extra task-specific neural layer followed by a non-linear layer and softmax for text classification on top of Bert. Then, the extra layers are jointly fine-tuned with the pre-trained Bert. A prominent limitation of Bert is that it takes at most 512 word tokens, which is often too small for document level tasks.

XLNet (Yang et al., 2019) is an unsupervised language representation learning method based on a novel generalized permutation language modeling objective. XLNet does not suffer from the pre-train-fine-tune discrepancy that Bert is subject to due to the Cloze Test objective during training. Additionally, XLNet employs Transformer-XL (Dai et al., 2019) as the backbone model, exhibiting excellent performance for language tasks involving long context. Overall, XLNet achieves high accuracy on various downstream language tasks including question answering, natural lan-

guage inference, sentiment analysis, and document ranking.

Roberta (Liu et al., 2019) is an improved version of Bert trained on a larger dataset with longer sequences. It also modifies the original design of Bert by removing the next sentence prediction objective and dynamically changing the masking pattern during pre-training. The author of Roberta claims that Roberta is comparable with XLNet on all GLUE (Wang et al., 2019) tasks and SQUAD (Rajpurkar et al., 2016), and achieves the state-of-the-art performance on 4/9 of the GLUE tasks.

3.3 The APS Dataset

The Australian Parliamentary Speeches (APS) dataset includes transcripts of second reading speeches related to same-sex marriage (SSM) bills presented in the the House of Representatives of the Australian Parliament between 2004-2017. The data has been obtained from the [Federal Parliament website](#). A random sample of 100 speeches was given to an honour student in political science, who was asked to identify 15 frame categories from the PFC, and to indicate the relevant passages representing each frame. The rater was also asked to indicate the dominant frame of each speech. We report the APS frame statistics in table 1.

Training data	MFC SSM	MFC no SSM	MFC IM	MFC no IM
Testing data	MFC SSM (SISC)	MFC SSM (AISC)	MFC IM (SISC)	MFC IM (AISC)
Roberta-Base	72.5	69.0	65.8	55.5
Xlnet-Base-Case	72.1	67.9	64.1	54.7
Bert-Base-Case	70.6	67.2	62.5	53.4
SVM	64.5	60.59	57.2	47.24
Weighted-SVM	65.5	61.45	58.4	49.26
Framing Lexicons	66.2	62.34	58.3	49.44
Ji and Smith (2017)	–	–	58.4	–
Card et al. (2016)	–	–	56.8	–
Field et al. (2018)	–	–	57.3	–

Table 2: Mean accuracy of Same-Issue and Same-Context (SISC); Across-Issue and Same-Context (AISC) evaluated on both the Same-Sex Marriage (SSM) and Immigration (IM). The training and testing data are indicated in the heading of each column.

4 Experiments and Discussion

We divide our experiments into four parts: Same-Issue and Same-Context (SISC); Across-Issue and Same-Context (AISC); Same-Issue and Across-Context (SIAC); Across-Issue and Across-Context. We follow the same setup as in Card et al. (2016) and report average accuracy across 10-fold cross validation. We use the Bert-Base-Cased, Roberta-Base, Xlnet-Base-Cased models. We use the pre-trained model from Huggingface package. We set the maximum sequence length to 256 since the average number of tokens for SSM and IM are 253 and 254 respectively. For more details about the pre-trained models’ parameters, we refer to the Huggingface package.

Same-Issue and Same-Context (SISC) We fine-tune and evaluate our models on the Same-Sex Marriage (SSM) and Immigration (IM) issues from the MFC dataset, and compare the results for IM with the previously proposed models, since to the best of our knowledge, IM is the only issue with results reported in previous work. Table 2 columns 2 and 4 show that the neural models outperform the basic classifier and lexicon-based methods. A paired t-test between Roberta-Base and Framing Lexicons method confirms the difference is statistically significant ($p < 0.001$). The difference between Roberta-Base and Xlnet-Base-Case is not statistically significant ($p = 0.061$), while the difference between Roberta-Base and Bert-Base-Case is ($p = 0.008$).

Across-Issue and Same-Context (AISC) To examine if our models can learn to predict frames across issues, we first exclude the SSM and IM data, respectively, from the MFC dataset and fine-tune our models on the data for the remaining issues. Then, we evaluate the models on the SSM and IM data and compare our results with the previously proposed models. Columns 3 and 5 of Table 2 show that there is a decrease from SISC in mean accuracy of about 4% for SSM, and 9% for IM. However, the classifiers are still well above chance, which is about 27.9% for SSM and 16.3% for IM if we default to the most common frame in the the respective issues.

Same-Issue and Across-Context (SIAC) To examine if our models can learn to predict frames across communication context, we fine-tune our models on the SSM data from the MFC, and then evaluate our models on the APS dataset. Table 3 (column 3) shows that there is a further drop in mean accuracy here for all models, but again still above chance, which is about 35.0% for APS if we default to the most common frame in the the respective issues.

Across-Issue and Across-Context (AIAC) To examine if our models can still learn to predict frames across both issue and communication context, we fine-tune our models on all other MFC data excluding SSM data, and then evaluate our models on APS dataset. Table 3 (column 2) shows that there is a further drop in mean accuracy here, about 9.3% on average for all models, compared to

SIAC, but again still above chance, which is about 35.0% for APS if we default to the most common frame in the the respective issues.

Training data	MFC no SSM	MFC SSM
Testing data	APS (AIAC)	APS (SIAC)
Roberta-Base	41.0	43.0
Xlnet-Base-Case	43.0	46.0
Bert-Base-Case	40.0	47.0
SVM	32.0	35.0
Weighted-SVM	33.0	37.0
Framing Lexicons	34.0	38.0

Table 3: Mean accuracy of Same-Issue and Across-Context (SIAC); Across-Issue and Across-Context (AIAC) evaluated on both the APS dataset. The training and testing data are indicated in the heading of each column.

5 Discussion

The previous best mean accuracy for predicting the dominant frame on the Immigration subset of the MFC is 58.4%. Our best model (Roberta-Base) fine-tuned with data on the same issue improves the performance by 12.7%, and our best model (Roberta-large—not shown in Table ??) fine-tuned on data not including the Immigration subset has 56.26% accuracy; still comparable performance against previous methods.

Our best model outperforms the previous best models on the MFC by a large margin. Notably, the performance of pre-trained language models is comparable to the previous best models, even with only fine-tuning on data not specific to the issue being classified, proving that pre-trained neural classifiers can learn to predict dominant frames across domains. However, fine-tuning on small amount of domain-specific data still outperforms the same models fine-tuned on out-of-domain datasets.

6 Conclusion

Using a pre-trained Roberta (Liu et al., 2019) model with added issue- and context-specific data to predict the dominant frame of a text improves upon the current state-of-the-art. Such a model that is trained on U.S. media articles can be convincingly applied to predict frames in Australian

political speeches, though the accuracy is significantly reduced, suggesting potential discrepancy in framing strategy between US News and Australian Parliamentary Speeches, and/or different uses of language in the two contexts. Over the coming months, we will work on improving the size and quality of the APS data and examine ways to improve the prediction of dominant frames in Australian political text.

References

- Alexandra Balahur, Ralf Steinberger, Mijail A. Kabadjov, Vanni Zavarella, Erik Van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2010. [Sentiment analysis in the news](#). In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*.
- Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. [Testing and comparing computational approaches for identifying the language of framing in political news](#). In *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015*, pages 1472–1482.
- Amber E Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues.
- Dallas Card, Amber E Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. 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 (Volume 2: Short Papers)*, pages 438–444.
- Dallas Card, Justin H. Gross, Amber E. Boydston, and Noah A. Smith. 2016. [Analyzing framing through the casts of characters in the news](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 1410–1420.
- Dennis Chong and James N Druckman. 2007. Framing public opinion in competitive democracies. *American Political Science Review*, 101(4):637–655.
- Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G. Carbonell, Quoc Viet Le, and Ruslan Salakhutdinov. 2019. [Transformer-xl: Attentive language models beyond a fixed-length context](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 2978–2988.

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.
- Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. [Framing and agenda-setting in russian news: a computational analysis of intricate political strategies](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3570–3580.
- Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. 2007. [Large-scale sentiment analysis for news and blogs](#). In *Proceedings of the First International Conference on Weblogs and Social Media, ICWSM 2007, Boulder, Colorado, USA, March 26-28, 2007*.
- Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. [Political ideology detection using recursive neural networks](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122, Baltimore, Maryland. Association for Computational Linguistics.
- Yangfeng Ji and Noah A. Smith. 2017. [Neural discourse structure for text categorization](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 996–1005.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. [SemEval-2016 task 6: Detecting stance in tweets](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.
- Viet An Nguyen. 2015. *Guided Probabilistic Topic Models for Agenda-Setting and Framing*. Ph.D. thesis.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. [Squad: 100, 000+ questions for machine comprehension of text](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 2383–2392.
- Stephen E. Robertson and Hugo Zaragoza. 2009. [The probabilistic relevance framework: BM25 and beyond](#). *Foundations and Trends in Information Retrieval*, 3(4):333–389.
- Oren Tsur, Dan Calacci, and David Lazer. 2015. [A frame of mind: Using statistical models for detection of framing and agenda setting campaigns](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 1629–1638.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5998–6008.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. [GLUE: A multi-task benchmark and analysis platform for natural language understanding](#). In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. [Xlnet: Generalized autoregressive pretraining for language understanding](#). *CoRR*, abs/1906.08237.