Towards quantifying politicization in foreign aid project reports

Wang, S.; Eggers, G.; de Roode Torres Georgiadis, A.; Đo, T.A.; Gontard, L.; Carlitz, R.; Bloem, J.

Publication date
2024

Document Version
Final published version

Published in
The Second Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2024)

Towards quantifying politicization in foreign aid project reports

Sidi Wang[♣♠], Gustav Eggers[‡♠], Alexia de Roode Torres Georgiadis[‡♠], Tuan Anh Đo[♢], Léa Gontard[♢], Ruth Carlitz[‡♠] and Jelke Bloem[♢♠]

♠ Data Science Centre, University of Amsterdam
♣ Computational Linguistics and Text Mining Lab, Vrije Universiteit Amsterdam
♢ Institute for Logic, Language and Computation, University of Amsterdam
‡ Department of Political Science, University of Amsterdam
s.wang4@student.vu.nl, gustav.eggers2@student.uva.nl,
alexia.de.roode.torres.georgiadis@student.uva.nl, tuan.do@student.uva.nl,
lea.gontard@student.uva.nl, r.d.carlitz@uva.nl, j.bloem@uva.nl

Abstract

We aim to develop a metric of politicization by investigating whether this concept can be operationalized computationally using document embeddings. We are interested in measuring the extent to which foreign aid is politicized. Textual reports of foreign aid projects are often made available by donor governments, but these are large and unstructured. By embedding them in vector space, we can compute similarities between sets of known politicized keywords and the foreign aid reports. We present a pilot study where we apply this metric to USAID reports.

Keywords: Politicization, foreign aid, document embeddings, Development Experience Clearinghouse

1. Introduction

When foreign aid is provided for political vs. altruistic interests, aid effectiveness is expected to suffer. However, evidence for this relationship – and the mechanisms through which it operates – is limited. This is due in large part to the fact that politicization tends to be operationalized quite bluntly. In addition, most studies of aid project effectiveness exclude the world's largest donor (the United States Agency for International Development, USAID), since USAID does not rate project effectiveness on a common numerical scale. However, the agency does make project evaluations publicly available through the agency's Development Experience Clearinghouse (DEC).[1]

The DEC provides access to over 10,000 evaluations spanning a range of activities and time periods. Unlike many of its peer foreign aid agencies, USAID does not have an independent evaluation agency but rather contracts evaluation out to various private firms. The evaluations thus comprise a range of formats and styles.

As a survey by Németh (2023) shows, NLP methods have been applied extensively and fruitfully to study the related notion of political polarization, showing that this concept can be successfully modeled on the basis of models trained on natural language data such as word embeddings. Unstructured natural language data is available in the DEC, annotated with categorical metadata representing variables of interest such as sectors. As the reports are fairly substantial (about 16k tokens per report on average) there should be enough in-domain training material for statistical NLP methods in these reports.

In this work, we aim to develop a metric of politicization by investigating whether this concept can be operationalized computationally. We also present a pilot study using a Doc2Vec-based method to quantify politicization of foreign aid reports in a sample of the DEC corpus.

[1] `https://dec.usaid.gov/dec/home/Default.aspx`

2. Related work

In the context of foreign aid, politicization has occurred when "disagreements over the means to achieve a given goal are drawn along ideological lines that correspond to distinct political constituencies" (Carlitz, 2023, p. 9). This may affect the effectiveness of aid projects. The politicization of foreign aid has been studied primarily in terms of donor characteristics, as well as the dyadic relationships between particular donors and recipients. The most prominent operationalization of politicization considers whether donors and recipients are in some way aligned – where allegiances are measured using voting patterns in the UN general assembly (Bobba and Powell, 2007) or looking at joint membership in the UN Security Council (Dreher et al., 2018). Scholars have also examined the influence of political misalignment and ideological distance between donor and recipient governments (Dreher et al., 2015). Scholars have further inferred donor motives (and thus politicization) by examining the effect of aid given for developmental vs. 'strategic' purposes (Kilby and Dreher, 2010). Such blunt operationalizations make it diffi-

Proceedings of the Second Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2024), pages 85–90
21 May, 2024.

cult to distinguish relative politicization of different activities funded by the same donor, or otherwise provide for nuanced analysis.

2.1. Political NLP

The use of natural language processing to extract information from political texts and discourses has been explored from various angles, often driven by practical research questions. For example, one line of work is applying dimensionality reduction techniques, such as Latent Semantic Indexing (LSI), to identify political preferences in US elections (Bonica, 2013, 2014). Rheault and Cochrane (2019) investigated the potential of applying n-gram language modelling and Principal Component Analysis (PCA) for capturing ideological placements of parties in the US House.

Parallel efforts at the document level have employed NLP to analyze polarization in parliamentary systems (Peterson and Spirling, 2018), party affiliation (Yu et al., 2008) and news coverage (Chinn et al., 2020). Work on uncovering linguistic indicators of polarization often employs unsupervised learning methodologies. Moreover, the task of classifying political affiliations based on speech (Binder, 1999) and tweet texts (Demszky et al., 2019) has been explored with various machine learning algorithms, such as random forest classifiers. In the context of legal texts, Nay (2016) extended the Word2Vec model to embed institution-specific representations into a shared vector space, taking temporal relationships between them into account. This allows for the comparison of policy differences across US Congresses and sitting Presidents. However, in the landscape of international development projects, as Moore et al. (2023) note, there is a lack of work that specifically employs embedding techniques to extract, label and rate text from foreign aid evaluation reports.

Document embeddings have gained significant attention in the field of computational social science due to their ability of capturing abstract semantic information from textual data. Introduced by Le and Mikolov (2014), Doc2Vec extends the Word2Vec model (Mikolov et al., 2013) to generate a fixed-length representation of a given variable-length piece of text, allowing the model to be easily adapted to infer dense vector representations of sentences, paragraphs or entire documents in an unsupervised manner. There are two main approaches in Doc2Vec, so-called Distributed Bag-of-Words (DBOW) and Distributed Memory Paragraph Vectors (DMPV). DBOW treats each document as a single representation for context word prediction, ignoring the order of words within the document. DMPV preserves the order by using both document representation vector and the word vectors in the context to make predictions. Recent applications of Doc2Vec include sentiment analysis (Chen and Sokolova, 2021; Shuai et al., 2018; Liang et al., 2020), text classification (Dogru et al., 2021; Aubaid and Mishra, 2020; Lee and Yoon, 2018), topic modelling (Budiarto et al., 2021), polarized news detection (Srivastava et al., 2019) and political polarization on Wikipedia (Gode et al., 2023). The model's success on these related tasks suggests that the rich semantic representations of documents that Doc2Vec provides also have the potential to operationalize a metric of politicization.

3. Data and method

USAID's Development Experience Clearinghouse (DEC) represents a rich and largely untapped resource capturing information on aid projects funded by the US government. USAID's evaluation policy (USAID, 2020) stipulates that external evaluations must be carried out for (1) all activities with a total cost exceeding $20 million and (2) each "intermediate result"[2] within a country strategy. The policy further stipulates that plans for the dissemination and use of evaluations must be developed and that evaluation final reports and their summaries must be submitted within three months of completion to the DEC.

Scholars have just begun to leverage the rich information contained in the DEC. For instance, Moore et al. (2023) have developed a standardized taxonomy for benchmarking projects in the agriculture sector. This work lays the foundation for a machine learning algorithm that extracts information on the effectiveness of different interventions and developed standard metrics.

Our study focuses on health projects, for which the DEC contains 4,000 evaluations spanning 70 years. We expect politicization to vary across sectors and activities, arguing that reproductive and maternal healthcare is more politicized than, e.g., malaria control. Following the approach of Moore et al. (2023), we used a balanced sample of 99 reports written from 2003 to 2021 on projects in the health sector.

In selecting the sample, we addressed the limitations inherent in the keyword tagging system of the DEC. Recognizing the frequent inaccuracies

[2]According to USAID's Program Cycle Operational Policy, an intermediate result [IR] is defined as, "A component of a Results Framework in a Mission's CDCS [Country Development Cooperation Strategy]. Intermediate Results are seen as an essential contribution to advancing a DO [Development Objective]. IRs are measurable results that may capture a number of discrete and more specific lower-level results and often define the purpose of projects" (USAID, 2022, p. 127).

in the DEC's keyword-based search functionality, our methodology employed the Development Evidence Large Learning Model (DELLM)[3], a proprietary Large Language Model fine-tuned in collaboration with USAID technical experts. This model demonstrates enhanced capability in accurately categorizing project reports by sector.

The DEC database API was used to operationalize this approach. This integration facilitated an exhaustive analysis wherein DELLM processed the entirety of the DEC's repository to accurately label documents as either 'final evaluations' or 'final grantee reports' within the health sector. Subsequent to this categorization process, a balanced random sampling technique was applied to select a representative subset of 99 labeled reports for further analysis. The sample was balanced to have an even representation of years and countries in which the projects took place.[4]

All but one report are in English and the resulting corpus is 1.6M tokens in size. A vast majority of the reported projects in the sample took place on the African continent but that is representative of the data. Some relevant metadata for the reports is available on the DEC website, most importantly including standardized USAID thesaurus keywords (Donnelly, 2021) for the topics covered in the report. We use these document keywords as labeled data for evaluation.

3.1. Keyword coding

We derived keywords that describe health-related topics from the USAID thesaurus (Donnelly, 2021). The USAID thesaurus keywords are based on 165.000 USAID documents, from across the world, spanning more than 50 years of USAID activities. The USAID thesaurus keywords are commonly used to classify the contents of documents, including USAID project reports (USAID-KSC, 2012). Keywords can be understood as representing the subjects, targets, and interventions of USAID activities. Examples of keywords are 'health', 'HIV/AIDS', and 'bednets'. We derived our keywords from the thesaurus categories relevant to the health sector. Specifically, our keywords are taken from the section 'health and safety' and the 'family planning' sub-section within the 'population and demography' section.[5]

We classify our keywords as politicized (scored 3), non-politicized (1), or potentially politicized (2).

[3] https://www.developmetrics.com/our-capabilities/

[4] Sample selection and text extraction from the DEC was performed in collaboration with DevelopMetrics, https://www.developmetrics.com/. None of the authors of this study are affiliated with DevelopMetrics.

[5] Sections K and S14 in the USAID thesaurus.

Following Carlitz's (2023) notion that the reproductive health sector is more politicized than other sectors, we classified such keywords as politicized. We classify keywords that are not related to reproductive health as non-politicized. Lastly, we classify keywords that capture interventions/targets that can be related to either reproductive or non-reproductive health as potentially politicized. Examples of keywords within the three categories are 'condoms', 'eye diseases', and 'health education'. The classification was done by co-authors with expertise in political science.

3.2. Model

We use Doc2Vec (Le and Mikolov, 2014) in its Gensim (Řehůřek and Sojka, 2010) implementation, trained on the aforementioned DEC corpus, to obtain a potential politicization metric. As our dataset is small for training Doc2Vec, we follow Lau and Baldwin's (2016) approach in initializing Doc2Vec with pretrained word embeddings.[6] The pre-trained word-embedding used is the Common Crawl 300-d vectors with 840b tokens. We chose the DMPV training algorithm which can retain order and thus usually generate better results[7].

We use the model to generate 300-dimensional vectors for each report in the DEC corpus. Based on a list of keywords, it can retrieve the most or least similar documents to the keyword's vectors. We create query vectors by averaging the vectors of query words, with the word vectors coming from the trained model. If the keyword contains more than one word, we split it into single words and take the average vector; we also skip words that are not in the vocabulary of the pre-trained word-embedding model. Using these document embeddings and the hand-coded politicized keywords, we can obtain a potential metric of politicization for a target document by calculating the cosine similarity between the average vector of keywords coded as politicized and the target document.

3.3. Evaluation method

Ideally we would evaluate this approach directly by manually assigning each report a gold standard politicization score and computing the correlation with our metric, but the political scientists in our team consider this an infeasible annotation task due to the abstract nature of the concept.

Instead, we use an indirect 'silver standard' approach based on the report metadata available in the DEC. We score the reports based on whether

[6] https://github.com/maohbao/gensim

[7] Model hyperparameters: vector_size: 300; min_count: 1; epochs: 50; dm: 1; seed: 240123. Punctuation and stopwords were removed.

the reports are labeled with politicized keywords in the DEC metadata, and call this the silver score. We then test whether our metric correlates with this silver score, hypothesizing that reports with a higher silver score also get a higher similarity score from our Doc2Vec model. On average, every report has 8 keywords in the DEC metadata, which may be coded differently (scored between 1 and 3 where 3 is politicized, cf. section 3.1). We turn this into a silver standard score by computing the average score of all keywords. If a keyword was not scored by our annotators (e.g. it is not related to the health sector) it gets a score of 1. Documents with a larger proportion of keywords that we coded as politicized thus have a higher silver score.

We consider this a valid evaluation because the Doc2Vec model does not have access to this keyword metadata. The USAID thesaurus keywords are not explicitly listed in the report, although if the keyword is a common word like 'disease', it will be mentioned in the running text. Some more abstract keywords such as 'mass media' do not occur in the report text at all. By receiving an average vector of politicized keywords, the model only has access to our politicization coding at the keyword level, not at the document level. Thus the connection to documents is not given and should be inferred.

4. Results

We compute the cosine similarity between the average politicized keyword vector and the document vectors, using this similarity as our metric. We use the Spearman correlation coefficient to estimate the correlation between our metric and the silver score for all documents. The coefficient obtained is $\rho = 0.280$ with a p-value of 0.005, a weak but statistically significant correlation.

Figure 1 shows all documents ranked by their similarity score plotted against their silver score. This figure shows that top ranked documents on average cover topics that are more politicized according to our annotators, but with some clear deviations from the linear trend around the middle ranks. This suggests that there may be a clustering of documents in the center of the vector space that are not clearly differentiated by politicization.

Among 99 reports, a report on the Mozambique Malaria Program (PA00MGHW) has the highest cosine similarity with politicized keywords. PA00MGHW also has a relatively high silver score of 2.0. While we were initially surprised at a report on a malaria project receiving such a high politicization score, we note that the project included as one of its three main objectives, "Expand access and quality of malaria in pregnancy activities in targeted districts." In the metadata, the report also has keywords related to this topic. This

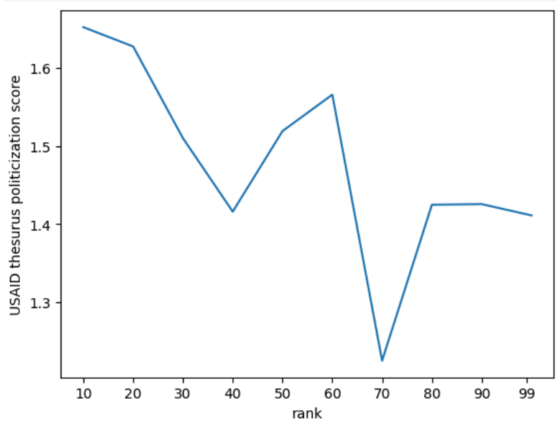

Figure 1: Politicization metric and silver politicization scores, binned in groups of 10 reports.

still lends scope for politicization as we understand it, and points to the importance of going beyond pre-determined keywords. Furthermore, the report also describes a predecessor program more focused on reproductive health, thus influencing the document embedding in a politicized direction, and the report is relatively short. This suggests that segmenting reports into their descriptions of distinct interventions may improve results.

A report on public health training in Ethiopia (PDACG247) has the lowest score on our metric. It also has a silver score of 1, the lowest possible. The main objectives, (1) Development of teaching materials in-country; 2) Strengthen staff through training in pedagogical, supervisory and writing skills; 3) Enhancement of the teaching-learning environment), were indeed not politicized according to our understanding of the concept.

An outlier with the third highest similarity but low silver score is report PA00MGHW. This report appears to be incomplete – that is, the actual evaluation is missing but rather this document is only a series of Annexes, presumably part of a comprehensive evaluation report. Thus, the low silver score indicates what we can miss by relying on externally applied keywords, as the information presented in the annexes does indeed appear to reflect politicized interventions as we understand them (e.g., comprehensive sex education).

A low similarity outlier (rank 97) discusses an Ethiopian reproductive health project, correctly tagged and thus receiving a high silver score of 2.25. The low similarity score was surprising, given the report mentions politicized topics like unsafe abortion. However, the low politicization score may reflect some form of self-censorship and thus may still be capturing a 'real' phenomenon of interest to scholars of politicization.

5. Discussion

While we have shown that our approach yields a metric that correlates with politicized content in foreign aid reports to some extent, there are some clear limitations. First, there is a dependence on manual annotation of politicized keywords. Inducing such keywords from political data sources external to the foreign aid reports would enable easier generalization beyond the health sector. Second, having one vector representation for an entire document proved to be too coarse-grained. Segmenting each report into descriptions of interventions, as also done by Moore et al. (2023), would reduce noise and better represent projects that address a variety of themes. However, as the reports are not consistently structured, this would require manual work. A further limitation is that we were not able to intrinsically evaluate the Doc2Vec model for this domain or perform hyperparameter tuning, due to limited availability of domain-specific resources.

A challenge we encountered throughout our work was coming up with a straightforward conceptualization of politicization that can be grounded in textual data, and identifying documents other than the corpus of reports that we could use to capture politicization. For instance, we searched for policy documents corresponding to Republican vs. Democrat health priorities but failed to find sufficient information. The method is likely more applicable to documents that are more clearly the output of political processes – e.g., comparing political party manifestos to policy documents produced by different parties. In future work we hope to integrate more explicitly political variables to engage more directly with Political Science questions.

The use of static embeddings precludes the possibility of observing different degrees of politicization for the same topics used in different contexts. In much political science work, operationalizations of politicization are conditional on the aid donor and therefore this contextual aspect should be represented in metrics of politicization. Therefore, we propose contextualized embedding-based methods as a future approach. By comparing keywords vector distance in different polarized contexts, we could attribute them a contextual politicization score and develop a politicization metric for keywords at the document level. This metric could be used to study the relation between politicization and project effectiveness. Through its grounding in contextual lexical semantics, this approach could yield deeper insight into the semantic nuances of language used in political discourse and reveal the extent to which political ideologies shape international aid strategies across different donor governments.

Acknowledgements

We are grateful to Lindsey Moore of DevelopMetrics for her help in extracting data from the DEC API. This project was supported by a University of Amsterdam Social and Behavioural Data Science Centre (SoBe DSC) Valorization Grant.

6. Bibliographical References

Asmaa M. Aubaid and Alok Mishra. 2020. A rule-based approach to embedding techniques for text document classification. *Applied Sciences*, 10(11).

Sarah A. Binder. 1999. The dynamics of legislative gridlock, 1947–96. *The American Political Science Review*, 93(3):519–533.

Matteo Bobba and Andrew Powell. 2007. Aid effectiveness: politics matters. Inter-American Development Bank Working Paper #601.

Adam Bonica. 2013. Ideology and interests in the political marketplace. *American Journal of Political Science*, 57(2):294–311.

Adam Bonica. 2014. Mapping the ideological marketplace. *American Journal of Political Science*, 58(2):367–386.

Arif Budiarto, Reza Rahutomo, Hendra Novyantara Putra, Tjeng Wawan Cenggoro, Muhamad Fitra Kacamarga, and Bens Pardamean. 2021. Unsupervised news topic modelling with Doc2Vec and spherical clustering. *Procedia Computer Science*, 179:40–46. 5th International Conference on Computer Science and Computational Intelligence 2020.

Ruth Carlitz. 2023. Aid politics at home and abroad. Downstream consequences for project effectiveness. In *13th Annual Conference of the European Political Science Association*.

Qufei Chen and Marina Sokolova. 2021. Specialists, scientists, and sentiments: Word2Vec and Doc2Vec in analysis of scientific and medical texts. *SN Comput. Sci.*, 2(5).

Sedona Chinn, P. Sol Hart, and Stuart Soroka. 2020. Politicization and polarization in climate change news content, 1985-2017.

Dorottya Demszky, Nikhil Garg, Rob Voigt, J. Zou, M. Gentzkow, James Shapiro, and booktitle = Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) Jurafsky, Dan. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings.

Hasibe Dogru, Sahra Tilki, Akhtar Jamil, and Alaa Hameed. 2021. Deep learning-based classification of news texts using doc2vec model. pages 91–96.

Joseph Donnelly. 2021. USAID Thesaurus 2021 dataset. [dataset].

Axel Dreher, Vera Z Eichenauer, and Kai Gehring. 2018. Geopolitics, aid, and growth: The impact of UN Security Council membership on the effectiveness of aid. *The World Bank Economic Review*, 32(2):268–286.

Axel Dreher, Peter Nunnenkamp, and Maya Schmaljohann. 2015. The allocation of German aid: Self-interest and government ideology. *Economics & Politics*, 27(1):160–184.

Samiran Gode, Supreeth Bare, Bhiksha Raj, and Hyungon Yoo. 2023. Understanding political polarization using language models: A dataset and method. *AI Magazine*, 44(3):248–254.

Christopher Kilby and Axel Dreher. 2010. The impact of aid on growth revisited: Do donor motives matter? *Economics Letters*, 107(3):338–340.

Jey Han Lau and Timothy Baldwin. 2016. An empirical evaluation of Doc2Vec with practical insights into document embedding generation. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 78–86. Association for Computational Linguistics.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196.

Hana Lee and Young Yoon. 2018. Engineering Doc2Vec for automatic classification of product descriptions on O2O applications. *Electronic Commerce Research*, 18:1–24.

Yinghong Liang, Haitao Liu, and Su Zhang. 2020. Micro-blog sentiment classification using Doc2vec + SVM model with data purification. *The Journal of Engineering*, 2020(13):407–410.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space.

Lindsey Moore, Mindel van de Laar, Pui Hang Wong, and Cathal O'Donoghue. 2023. Making impact with agricultural development projects: The use of innovative machine learning methodology to understand the development aid field. *UNU-MERIT Working Paper*, (2023-011).

John J. Nay. 2016. Gov2Vec: Learning distributed representations of institutions and their legal text. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 49–54, Austin, Texas. Association for Computational Linguistics.

Renáta Németh. 2023. A scoping review on the use of natural language processing in research on political polarization: trends and research prospects. *Journal of computational social science*, 6(1):289–313.

Andrew Peterson and Arthur Spirling. 2018. Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems. *Political Analysis*, 26(1):120–128.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. `http://is.muni.cz/publication/884893/en`.

Ludovic Rheault and Christopher Cochrane. 2019. Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis*, 28:1–22.

Qianjun Shuai, Yamei Huang, Libiao Jin, and Long Pang. 2018. Sentiment analysis on Chinese hotel reviews with Doc2Vec and classifiers. In *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 1171–1174.

Vertika Srivastava, Ankita Gupta, Divya Prakash, Sudeep Kumar Sahoo, RR Rohit, and Yeon Hyang Kim. 2019. Vernon-fenwick at SemEval-2019 Task 4: Hyperpartisan news detection using lexical and semantic features. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1078–1082.

USAID. 2020. USAID evaluation policy.

USAID. 2022. Program cycle operational policy. ADS Chapter 201.

USAID-KSC. 2012. The USAID thesaurus 6th edition. ID: PN-AEA-100.

Bei Yu, Stefan Kaufmann, and Daniel Diermeier. 2008. Classifying party affiliation from political speech. *Journal of Information Technology & Politics*, 5(1):33–48.