Midterm Project Report for Newspaper Bias project

Tom Arend
t.arend@phd.hertie-school.org

Nicolai Berk
nicolai.berk@gmail.com

Abstract

The measurement of ideology is one of the major applications of text analysis in political science. However, researchers often face scarcity of available labelled data to train supervised models for their specific domain. Manual annotation is costly and often severely affected by subjective bias. We propose to fine-tune transformer models on available, labelled political texts issued by political parties to obtain a classifier of political ideology. Using a unique dataset of newspaper articles authored by politicians, we test such an application in the German context. Comparing transformer neural networks fine-tuned on a set of party press releases, a set of newspaper articles, or both, we present evidence on the feasibility of such an approach. This contributes to the broad literature on text analysis in the political domain, enabling researchers to train powerful deep learning models on political language with scarce training data. Additionally, we contribute a state-of-the-art deep learning model for the measurement of ideological bias in news articles. This is a crucial issue in the debates surrounding media effects on polarisation, turnout, and voting behaviour.[1]

1. Introduction

The measurement of ideology and political bias[2] are the subject of much research on political texts [1, 6, 9]. Despite significant advances in our understanding and detection of ideology, most researchers still face significant constraints when working with text. Notably, they face a lack of appropriately labelled training data. This is linked to the high costs incurred by manually annotating a significant number of speeches, texts, or sentences. Often, the detection of ideological bias might be highly dependent on the coders' subjective assessment. Instead, researchers could train models on other sources of text with clear and available labels and subsequently apply them to the desired texts using transfer learning.

We conduct experiments with differently fine-tuned deep learning models to understand if and how transfer learning can be used to measure bias in the absence of abundant training data. To test this, we fine-tune a deep neural network to predict the authoring political party of German press releases. Once the model is fine-tuned to predict the authoring party of press releases, it is applied to estimate the bias of newspaper. In addition, we compare this approach to models that are (additionally) fine-tuned in the domain of newspaper articles to assess the effectiveness of this alternative fine-tuning process. While the application of transformers to measure bias is not entirely new in political science [10, 11], we move beyond the current state-of-the-art by measuring the precise implications of different fine-tuning procedures.

Beyond testing the effectiveness of this two-step fine-tuning process, this project will develop a state-of-the-art deep learning model for the measurement of political bias in newspaper articles. Newspapers represent an important institution in the political world, affecting phenomena ranging from polarization to voter turnout. Much like the shadows in Plato's allegory of the cave, news provide elites and citizens with a representation of a reality they are not able to see themselves [8]. The media have the power to affect voting behaviour [2, 5], as well as polarise the electorate [7] or motivate them to turn out to vote [3].

Given this importance of the news media for the study of politics, it is surprising that few papers deploy state-of-the-art deep learning technologies to classify ideological bias in news articles. Gentzkow and Shapiro estimate slant in US newspapers by identifying bi- and trigrams' indicative of a congressional speakers' party, and apply the resulting dictionary to newspapers to scale them [3]. More recently, Widmer et al. have assessed polarisation in the US media environment using a supervised model. They train a classifier on bigrams, predicting whether content was produced by a left-leaning network (CNN) or a right-leaning network (Fox news) [12]. We believe both approaches are likely inferior to more complex deep learning models, as such novel approach would not incorporate idiosyncratic phrases used by the specific networks identified to train the data. Using

[1] GitHub repository: https://github.com/nicolaiberk/nlpdl_project

[2] 'Political bias', 'ideological slant' and variations of the two are used interchangeably in this report.

party labels to train classifiers is more straightforward, as it places newspapers within the existing context. If existing approaches to classify newspaper slant can be improved upon, or even just complemented, we could provide an additional tool for researchers to study drivers and effects of media bias. A working state-of-the-art model might even renew interest in the subject matter and encourage researchers to find new and exciting applications for it.

Outside of academia, a confident and robust classifier of newspaper bias might help readers to identify when they are reading an article that is overly partisan. This would perhaps encourage them to approach certain news sources with more scepticism and hold news outlets to higher editorial standards. In the long run, the highlighting of biases in articles might counteract the worrying polarization of entire electorates.

2. Proposed Method

We propose the use of transfer learning to identify newspaper slant: a pre-trained transformers model is fine-tuned to identify the authoring party of press releases. Then, this model is applied to a range of newspaper articles, indicating which parties' communication an article most resembles. This approach is useful, as political parties represent the major organisers of and reference points for political ideology. More importantly, party-issued political texts constitute data that already carry ideological labels and need no further labour for annotation. It therefor represents an efficient and broadly available source of data for transfer learning to identify ideological bias in newspaper articles. 'Bias' is here defined as the similarity of a set of text to a given party's communication. This is a reasonable definition, as political parties constitute the major points of reference and organizers of ideology in contemporary democracies.

In detail, we employ pre-trained BERT neural network models for German language[3] and compare how three different training processes affect model performance. We expect optimal performance from a transformer neural network that was fine-tuned in two steps. First we use the pre-trained model to classify party press releases by issuing party. Secondly, the model is applied in its actual domain to estimate ideological bias in news articles. We fine-tune and subsequently validate the model using op-eds by politicians. These unique data have the advantage of carrying party labels, allowing a direct transfer of the categories from the initial fine-tuning process (first step) to the outcome of interest. We compare this fine-tuned model to a BERT model that was not fine-tuned on party press releases and one that was not fine-tuned on newspaper data. This leaves us with three models:

1. A model only fine-tuned on newspaper articles authored by politicians.
2. A model only fine-tuned on party press releases.
3. A model both fine-tuned on press releases and politicians' newspaper contributions.

This approach assess by how much model performance is improved upon, when including information from both party press releases and newspaper data. This is especially useful as it allows us to comparatively assess the performance of a model that has never seen data from the domain of interest (model 2). If this model performs well, this constitutes strong evidence that party communication can be used to train classifiers which are applied in a diverse set of domains with scarce or no training data. If model 3 severely outperformed model 1, this would corroborate the idea that information from available sets of texts from political actors can be used to improve ideology measurements for other sets of text.

3. Experiments

Data: In this project, we draw on three distinct data sources:

- A dataset of over 40,000 German party press releases issued between 2010 and 2019, collected by the SCRIPTS project[4].
- A collection of over 2 million German newspaper articles from six major newspapers, published between 2013 and 2019, collected by one of the authors in a previous project ([4]).
- A set of German newspaper articles authored by politicians collected as part of this project (collection underway).

Evaluation method: So far, we have successfully fine-tuned a BERT-model to classify party-press releases by authorship (model 2, serving as the basis for model 3). The performance of this model is shown in 1. As the reader can see, the model performs at a very high, near-perfect level. We believe that the classification performance made hyper-parameter optimisation at this stage somewhat obsolete. It might however be useful in the second stage of fine-tuning. The impressive performance of this model on the press releases was rather surprising to the authors and shifted the initial focus of the project away from improving

[3]Link to the model page: `https://huggingface.co/transformers/model_doc/distilbert.html`

[4]`https://www.scripts-berlin.eu/index.html`. Special thanks go to Lukas Stötzer for the effortless (for us) provision of the data.

	label	class	f1	precision	recall	n
3	3	SPD	0.997927	0.997238	0.998617	1446
4	4	Linke	0.997237	0.997543	0.996931	1629
5	5	FDP	0.996416	0.996813	0.996019	1256
0	0	Greens	0.995416	0.996940	0.993898	1311
1	1	Union	0.990737	0.984224	0.997336	1126
2	2	AfD	0.989836	0.998423	0.981395	645

Figure 1. Performance of model 2 for the classification of party press releases.

the classification of party press releases to the comparison of improvements in the two-step fine-tuning process.

For the final assessment of classifier performance, we will use a unique, newly collected set of politicians' op-eds and interviews. This data enables us to compare the impact of different fine-tuning processes (using party press releases, newspaper articles, or both) on model performance, as the outcome categories (authorship party) are identical. We aim for a balance of precision and recall, and will therefore evaluate the models' F1-scores. Additionally, we will assess generalisability by assessing the placement of newspapers based on estimates for all news articles. Left-wing newspapers such as the TAZ should be placed closer to left-wing parties (Grüne, SPD, Linke) than right-wing newspapers (FAZ, WELT).

Experimental details: We use the 'distilbert-base-german-cased' model from the Huggingface transformer library[5] for all three models. This deep-learning model has been trained to efficiently solve different classification tasks on 12 GB of German language data, including Wikipedia, legal data, and news. Given the impressive performance on the party press releases, deviation from the default settings for most hyper-parameters was deemed unnecessary thus far. The same applies for the use of the larger, more powerful, but slower 'bert-base-german-cased' model. We decided on three training epochs, a training batch size of 16, and a weight decay of 0.01 for regularisation. This might change in the future, depending on the model's performance on the newspaper articles.

The configuration of our hyper-parameters hence largely follows the defaults and is as follows[6]:

- Training epochs: 3
- Training batch size: 16
- We apply a weight decay as a form of regularisation: 0.01
- Maximum sequence length (longer articles were truncated): 512
- Dropout for each layer: 10%

Results: We classified a training set of 4,000 newspaper articles using model 2 (fine-tuned on party press releases). We expected articles from conservative newspapers FAZ and Welt to be more similar to Union and FDP and possibly the AfD, especially compared to the progressive TAZ. Spiegel Online (SPON) is expected to be equally similar to right- and leftwing newspapers.

The results can be seen in table 1, which shows the average probability for an article to be classified as being authored by a given party. As expected, FAZ and Welt are very similar to Union (81%) and FDP (58%/55%), but also rather similar to the Greens (36%/35%). While they show the highest similarity to the AfD (8%/7%), similarity to the radical-right party is generally on a very low level among all newspapers. Spiegel Online (SPON) shows lower similarity to the FDP and closer to Union and Greens, but is generally rather similar to the right-wing newspapers. As expected, the TAZ shows a comparatively different profile, being very similar to the Greens (average likelihood 70%), and less similar to the Union parties (41%), the FDP (32%), and the AfD (5%). Surprisingly, it also shows the lowest similarity to the Linke (13%). Maybe most surprising is the general low similarity to SPD press releases (4%/5%). It seems the party has a rather distinctive style in its press releases.

Paper	Linke	Grüne	SPD	Union	FDP	AfD
FAZ	0.15	0.36	0.05	0.81	0.58	0.08
Welt	0.15	0.35	0.05	0.81	0.55	0.07
Spiegel	0.11	0.39	0.05	0.89	0.34	0.07
TAZ	0.13	0.69	0.04	0.41	0.32	0.05

Table 1. Mean similarity estimate to each party by newspaper.

Given that model 2 is likely the worst performing model, we consider these results to be promising at this early stage. However, we need to look into the odd performance of the SPD category, which we expected to be generally well-represented in media articles, especially in the left-wing TAZ.

[5] https://huggingface.co/bert-base-german-cased.

[6] The model's default settings can be found here: https://huggingface.co/distilbert-base-german-cased/blob/main/config.json.

4. Future work

Given the surprisingly good performance of the baseline model on the classification of party press releases, the project can move on to focus on the evaluation of the effectiveness of the two-step fine-tuning process as its major contribution. We identified several key tasks to meet this goal[7].

The collection of several hundred newspaper articles and interviews by politicians for fine-tuning and model evaluation is the most important and urgent next step. This is done mostly by scraping politicians' websites. Once this data is available, we will be able to fine-tune model 1 and model 3 and comparatively assess the performance of all three models on a held-out set. This will enable us to assess which model performs best, and - more importantly - how much the model is improved upon by each fine-tuning step.

Even though the assessment of the models' performance does constitute the major goal of the project, several additional tasks are conceivable in order for us to fully grasp the necessary model improvements. The extremely accurate performance of the model 2 in the classification of press releases might fit certain characteristics we do not want it to learn. We intend to assess this by using a test set that lies in the future, meaning beyond the end-date of the training set currently in use. This way, we can see whether the model's high accuracy owes to idiosyncrasies of party press releases at a given period in time (same topics, same politicians addressed). Should accuracy drop substantially, the model might suffer from overfitting and the classification accuracy might be due to the incorporation of time-specific noise. This is likely only necessary if model 2 performs rather badly on the test set of politicians op-eds.

In the same vein, the surprisingly low classification rate for news articles resembling the SPD will have to be investigated. This might entail an in-depth look at the input data. We will most likely take a qualitative approach to understand whether the SPD's press releases are radically different from others or whether the party is really underrepresented in the coverage of these newspapers.

There are some additional tasks that we would like to address should time permit it. Chief among these is extensive hyper-parameter optimisation to improve the performance of all models. More importantly, we might make additional efforts to understand *why* additional fine-tuning steps could be important (or not). One potential avenue to understand how different fine-tuning processes affect our models, we might build on recent work employing causal mediation analysis to assess the interaction of the structure and predictions in transformer models [11].

[7]The open issues can be found via `https://github.com/nicolaiberk/nlpdl_project/issues`.

5. Conclusion

As it stands, we are confident that the project is on the right track, despite the necessary shift in focus. While the surprising performance of the initial baseline model might have derailed our original efforts, we believe that we managed to successfully shift the project in a direction that promises to result in an original and ambitious contribution to the literature. We strongly believe that there is a value in pursuing an approach that is more focused on the transfer learning aspect than on the initial fine-tuning. Not only does it address one of the crucial shortcomings of deep learning (i.e. a need for large amounts of labelled data), but also opens up the benefits to wider applications. We are eager to hear your feedback regarding both the results thus far and our ideas for the experiments addressing the issue of fine-tuning transformers in the absence of appropriate training data.

References

[1] A. Bilbao-Jayo and A. Almeida. Automatic political discourse analysis with multi-scale convolutional neural networks and contextual data. *International Journal of Distributed Sensor Networks*, 14(11), 2018.

[2] C. F. Chiang and B. Knight. Media bias and influence: Evidence from newspaper endorsements. *Review of Economic Studies*, 78(3):795–820, 2011.

[3] M. Gentzkow and J. M. Shapiro. What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1):35–71, 2010.

[4] W. Krause and N. Berk. Right-Wing Terrorist Attacks, the Media's Reactions, and Radical Right Party Support. Working Paper. 2021.

[5] J. M. D. Ladd and G. S. Lenz. Exploiting a rare communication shift to document the persuasive power of the news media. *American Journal of Political Science*, 53(2):394–410, 2009.

[6] M. Laver, K. Benoit, and J. Garry. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331, 2003.

[7] Y. Lelkes, G. Sood, and S. Iyengar. The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect. *American Journal of Political Science*, 61(1):5–20, 2017.

[8] Plato. *Republic*.

[9] A. Simoes and M. Del Mar Castaños. Fine-Tuned BERT for the Detection of Political Ideology Stanford CS224N Custom Project. 6(2017):1–5, 2020.

[10] Z. Terechshenko, F. Linder, V. Padmakumar, M. Liu, J. Nagler, J. A. Tucker, and R. Bonneau. A Comparison of Methods in Political Science Text Classification: Transfer Learning Language Models for Politics. SSRN Scholarly Paper ID 3724644, Rochester, NY, Oct. 2020.

[11] J. Vig, S. Gehrmann, Y. Belinkov, S. Qian, D. Nevo, Y. Singer, and S. Shieber. Causal mediation analysis for interpreting neural NLP: The case of gender bias. *arXiv*, 2020.

[12] P. Widmer, E. Ash, and S. Galletta. Media Slant is Contagious. *SSRN Electronic Journal*, pages 1–42, 2020.