# Politics, Morals and Tweets

## Identifying Political Affiliation Utilizing Moral Foundations Theory and Contextual Embeddings*

Raphael Mitsch

University of Vienna, Universitätsring 1, 1010 Vienna, Austria
a1006529@unet.univie.ac.at
https://github.com/rmitsch/righteous-mind

**Abstract.** The *Moral Foundations Theory* (MFT), popularized by Jonathan Haidt in [8], is an influential theoretical concept in social psychology. It proposes to explain behavioral differences in humans by coupling these differences with a set of fundamental moral values which do not hold the same significance for everybody. MFT emphasizes the cultural and political context, i. e. that the moral values prioritized depend on a person's political ideology and cultural context and/or vice versa. This implies that people with opposing political affiliations judge the value of these moral foundations differently, potentially leading to stark conceptional divergences in opinions on the importance of moral values between people attracted to disparate political ideologies.

Building on the MFT, we are applying state-of-the-art contextual embeddings to extract the importance of moral values for a set of professional politicians in the US. We empirically evaluate whether political preferences can be deduced from these personal moral matrices by training a machine learning model on these moral matrices to distinguish between political ideologies.

**Keywords:** Moral Foundations Theory · Natural Language Processing · Social Media · Word Embeddings.

## 1 Introduction

We motivate this work of research primarily by attempting to answer the question of whether Twitter users' inclination towards political ideologies can be inferred from their prioritization of moral values (MVs). This is heavily inspired by and building on research undertaken in [7,6,8], for which tens of thousands of questionnaires were gathered for an empirical analysis. We approach this problem from a different perspective, with a diverging tool set and no need for user interaction: We want to deduce which political ideology Twitter users are partial to without requiring them having to explicitly provide information as was

* Project and paper were created in partial fulfillment of the requirements for the course *Social Media Data: Quantitative Text Analysis of Big Data*, University of Vienna, Winter Semester 2018/2019.

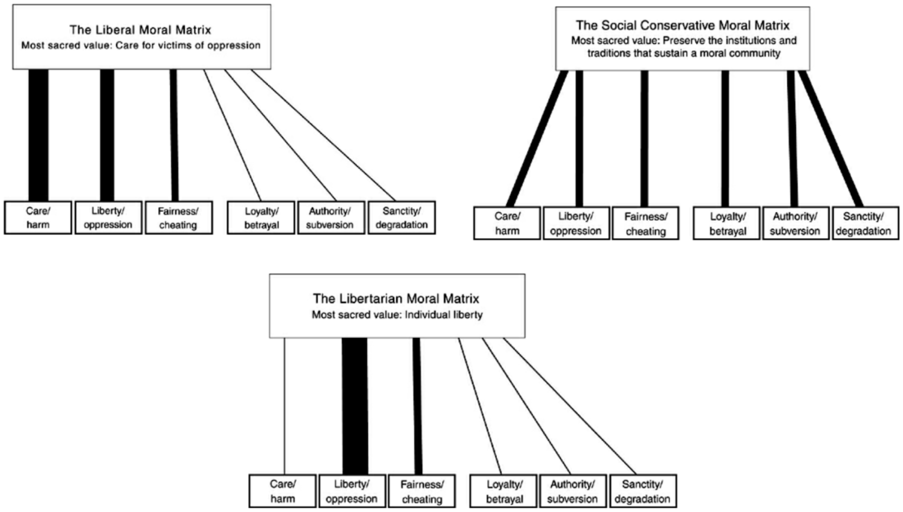


Fig. 1: Moral matrices for participants in surveys on https://moralfoundations.org/ for participants with a declared Liberal, Conservative or Libertarian ideology. Line width represents the priority of this moral foundation for participants with this ideology. Moral values from left to right: Care/harm, liberty/oppression, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation.

required by Haidt's et al. surveys. Thus our goal is twofold: (1) Predict users' political ideology based on their moral priorities and (2) analyze feature importance, i. e. the importance of moral values (MVs), in the classification model applied for prediction in regard of its similarities and differences from the moral priorities as specified in [8]. This requires extracting users' moral priorities as a first step.

The three political ideologies that were examined in the light of MFT were *Conservative*, *Liberal* and *Libertarian*[1]. Through the mentioned surveys conducted during the work on [7,6,8], data was gathered that enabled the authors to create a *moral matrix* for each of these three political ideologies - see figure 1[2]. A moral matrix in this context is the set of priorities assigned to the set of moral values defined by the MFT.

In order to simplify the tasks illustrated above, we...

[1] We capitalize these political ideologies throughout this document to emphasize that there are multiple definitions of these terms and we follow a specific one.

[2] These images were taken from http://clsbluesky.law.columbia.edu/2017/06/07/corporate-governance-as-moral-psychology/ and were created by Jonathan Haidt.

- ...investigate only US-American politics, for which Twitter data is comparatively abundant (thereby conforming with [8]). In US politics, *Conservative* roughly translates into an affiliation with the Republican Party, whereas *Liberals* are found in the Democratic Party (or the Green Party, which we do not take into account). *Libertarians* are predominantly represented by a school of thought in the Republican Party. Therefore we use members of the Democratic Party as a proxy for Liberals and members of the Republican Party as a proxy for Conservatives[3].
- ...take exclusively tweets from professional politicians into consideration, since we can assume a reliable party affiliation for those. Independent candidates are either dropped from the dataset or affiliated with the party they bear the strongest ideological similarity with - we pool together e. g. Bernie Sanders and members of the Democratic Party as Liberals.
- ...limit the number of possible outcomes to either Republican or Democrat. This is due to two reasons: (1) The number of libertarians is comparatively low, therefore we expect higher variance in our data. (2) We didn't find a version of the MFD providing a set of keywords for the moral foundation Liberty, which is essential in the self-identification of Libertarians [8]. Gathering and validating a set of keywords by ourselves was not in the scope of this paper.

It is common in works dealing with MFT to determine a textual entity's similarity to a MV by comparing it with terms in the Moral Foundations Dictionary (MFD) [7,9], which lists representative terms by their corresponding MV. A downside of this approach is that it requires an exact match between the word or token in the dictionary and the one appearing in the examined document.

Text embeddings, which have been highly popular in the NLP community and beyond that since [13], are able to mitigate or even fully solve this problem by inferring one fixed-length vector for a single word or a series of them. While the length of the sequence to infer is arbitrary, semantic meaningfulness is more likely to decrease with longer texts. For techniques supporting character n-grams or single characters - as opposed to entire words - as fundamental tokens, even misspelled or otherwise similar words have highly similar word embeddings. We therefore propose an embedding-based approach for computing a document's similarity to any of the MVs. This approach is further detailed in 4.

Once the similarities between tweets and Moral Values are obtained, the similarities to all MVs for each user in the dataset are aggregated. These aggregated similarities in conjunction with information on which party this politician is affiliated with are fed into a classifier. The concluding evaluation discusses

1. whether a meaningful separation on the grounds of MVs is possible and
2. if so, which features are how important and whether they resemble those presented in [8].

[3] Therefore please consider the terms "Liberal" and "Democrats" respectively "Conservative" and "Republicans" as exchangeable for the remainder of this paper.

## 2 Related Work

MFT builds on the framework introduced by Graham and Haidt in [7] and [6] and popularized by Haidt in [8], in which one additional moral foundation, liberty, is considered and examined with regard to how strongly the moral matrix of Libertarians depend on it. We use version 2.0 of the MFD, provided by Frimer. Using this version over the original one is recommended in [9].

There is a considerable corpus of work on MFT-related topics which we can't exhaustively treat here. Amongst the newer entries in this corpus are papers aiming at identifying moral sentiment that take word embeddings into consideration [5] and use knowledge graphs to inject background knowledge into the inference process [11].

One of the first widely disseminated word embeddings is [13], which is still exceedingly popular as a baseline embedding in many fields. More recent developments in this field are ELMo [14] and BERT [4], both of which are deep neural network-based contextualized systems enabling and encouraging transfer learning[4]. Here, ”contextualized” means that a word's context is taken into account when encoding it in an embedding vector[5].

## 3 Dataset

A dataset of roughly $1.2 \cdot 10^6$ tweets from 543 professional politicians in the US and up to 3200 tweets per user [16] builds the foundation for our analysis. We ignore all features in this dataset except user names and tweet's texts.

We first considered pulling data from the Twitter API directly, but quickly diverted our attention to finding a readily available dataset due to the limitations on the number of tweets that might be downloaded. The set of tweets we would have been able to gather from Twitter's API might not have been large enough for our purposes.

Although even professional politicians are not either simply Liberal or Conservatives and this dichotomy in itself is a strong simplification of the existing multi-dimensional political continuum, having tweets by declared members of one distinctly Conservative and one distinctly Liberal group at least somewhat reduces the risk of mislabeled users.

## 4 Methodology

Our approach is briefly summarized in section 1. In the following we detail the methodology and techniques used as well as various challenges we encountered in this context. The workflow can be divided into the steps listed below, assuming

[4] See e. g. `https://jalammar.github.io/illustrated-bert/` and `http://ruder.io/nlp-imagenet/` for introductions into how it works why and how this is important.

[5] Exemplary: The embeddings ”bank” in ”the river bank” and ”money in the bank” would differ.

all requested data - see section 3 - is available. The steps' order of enumeration is consistent with their chronological sequence.
For a description of the evaluation process see section 5.

### 4.1 Estimate Emotional Intensity for Tweets

Emotional intensity might be relevant in classifying personal relevance of a MV, if one assumes that topics prioritized highly by the author are not only written about more often, but also more emotionally. It is with this intuition that we estimate each tweet's emotional intensity. We utilize VaderSentiment [10] to extract the positive, negative and neutral sentiment. VADER normalizes their sum so that $pos_{text} + neu_{text} + neg_{text} = 1$. We therefore define the emotional intensity of a tweet as $int_{tweet} = 1 - neu_{tweet}$. This is motivated by the demand to extract a scalar representing how emotional a piece of text is. It does not matter if the text has positive or negative connotations, since all we care about is the author's emotional involvement in the tweet and, consequently, the tweet's topics.

### 4.2 Preprocessing and Cleaning

User handles and links are removed from tweets, since they most likely don't encode information semantically relevant for our purposes. The same applies for special characters and the "RT" marking a retweet. The "#" in hashtags are removed, the text following after it is split up in several words, if possible[6]. Finally, stop words are removed.

We do not lemmatize tokens, since words with slightly different spelling are still semantically similar with character-based embedding techniques like BERT.

We also fetched politicians' affiliation with political parties as proxy for their political ideology. This was done by parsing the Google results for the search phrase "[NAME] political party" and manually adding this information for the few names for which this procedure didn't yield automatically processable results. Polling Wikipedia articles on the politicians in question and consulting the Google Knowledge Graph API was also attempted, but either didn't yield the relevant information directly or was less robust than parsing the Google's search result.

### 4.3 Compute Embedding for Moral Phrases

We infer one embedding vector for all phrases in the MFD. While almost all phrases are 1-grams, some are 2- or 3-grams. Hence we use BERT's n-gram embedding capability to infer *one* single embedding over all tokens in a phrase, thereby simplifying a comparison and enabling uniform treatment of all moral phrase vectors.

[6] The python module `wordsegment` is used for splitting words. See https://github.com/grantjenks/python-wordsegment.

### 4.4 Train Model for MV Prediction

An essential idea in this project was not to compare each token with each moral phrase, but instead to train a classification model predicting the relevance of moral foundations for a specified word from its 768-dimensional BERT embedding. Specifically, a tree-based XGBoost [3] model was trained on a stratified sample with a size of 70% of all moral phrases. The usage of this model is outlined in 4.6. Running precision, recall and F1 computations on the held-out test set yields the results shown in the subsequent table.

| Moral Value | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Care | 0.69 | 0.70 | 0.70 | 108 |
| Fairness | 0.82 | 0.82 | 0.82 | 117 |
| Loyalty | 0.80 | 0.69 | 0.74 | 88 |
| Authority | 0.77 | 0.56 | 0.65 | 48 |
| Sanctity | 0.79 | 0.90 | 0.84 | 165 |
| Micro | 0.78 | 0.78 | 0.78 | 526 |
| Macro avg. | 0.78 | 0.74 | 0.75 | 526 |
| Weighted avg. | 0.78 | 0.78 | 0.77 | 526 |

While the results certainly surpass a randomized baseline, they could be improved[7] for a better result on downstream tasks - the quality of this classifier is crucial for the analysis pipeline. Compare with sections 6 and 7 for a deliberation on this.

### 4.5 Compute Tweet Embeddings

As described above, we aim to feed the embedding vector for a single token into our classifier and fetch the resulting MV probabilities - reflecting the semantic similarity between any MV and this token. Therefore we first infer the embedding vector for each word remaining in every tweet after the cleaning process. Note that this is by the far the most computationally intense step in the workflow; GPU acceleration is strongly recommended[8]. A token-wise embedding is produced by not pooling the output of the last layers of the network.

### 4.6 Predict MV Probabilities for Tweets

The outcome of step 4.5, i. e. the 768-dimensional embedding vectors for all words in a tweet, is forwarded to the classifier described in 4.4. The MV probabilities

[7] Note that the recall score for Authority is an outlier and particularly and rather poor.

[8] Computing the embeddings on a i7 dual-core machine without a GPU took several days. We didn't store the results due to space limitations, instead we directly computed the MV probabilities as described in step 4.6. Batch-wise inference would speed up the computation even more on a GPU.

for all tokens in the current tweet are predicted and summed up so that the resulting vector $\mu$ is of size five (one scalar per MV):

$$\mu_t = P(MVs \mid t) = \sum_{i=1}^{n_{w,t}} \theta(w_{t,i})$$

with $w_{t,i}$ being the $i$-th word in a tweet $t$ and $n_{w,t}$ the number of words in tweet $t$. $n_{w,t}$ represents the number of words $t$. Let $\mu_t$ be a vector of fixed length 5, holding the probabilities that $t$ is semantically most similar to the corresponding MV. $\theta$ is the inference function for the classifier trained to predict probabilities of word $w$ being relevant for any MV.

The numbers of words and tweets w.r.t. the authors are kept track of.

### 4.7 Aggregate MV Probabilities by Politician

Having collected the MV probabilities for all tweets, the emergent MV preferences for a politician are determined by simply averaging the sum of MV probabilities over the number of total words in this politician's corpus of tweets:

$$\xi_u = n_{u,w}^{-1} \cdot \sum_{i=1}^{n_{u,t}} \mu_{u,i}$$

where $u$ is any user, $n_{u,w}$ is the number of words in tweets authored by this user, $n_{u,t}$ the number of tweets authored by user $u$ and $\mu_{u,i}$ the sum of MV probabilities for user $u$'s $i$-th tweet as described in 4.6. $\xi_u$ is the resulting normalized MV priority vector of length 5 for user $u$.

### 4.8 Train Classifier For Prediction of Political Ideology

Finally, we train another classifier using politicians' aggregated MV probabilities as input to predict which political ideology they are inclined towards. As for the MV probability classifier we use a tree-based XGBoost model. A 20-fold cross-validation on the trained classifier yields a macro-F1 score of 0.73 with a standard deviation of $\sigma^2 = 0.2$ on the held-out test containing 50% of all users.

### 4.9 Technical Challenges

Some of the assumptions made during this workflow's inception are outlined in the following:

- BERT is able to accurately infer embeddings that capture the semantic information of any given word and phrase. We rely on the evaluation undertaken in [4] that reports SOTA performance in several tasks requiring semantically useful embeddings as sufficiently strong support for this assumption. Note that we used the pretrained, smaller $BERT_{Base}$ model[9] producing 768-dimensional embeddings due to memory constraints.

[9] The excellent https://github.com/hanxiao/bert-as-service was utilized for inferring embeddings with the pretrained BERT graph.

– The phrases contained by the MFD accurately reflect the MVs they represent. While this is a mostly subjective stance, we presume this to be approximately correct, since the MFD (deducing from our limited literature research in this context) seems to be the consensus amongst research in this field.

### 4.10 Alternatives

Other embedding approaches might be used as surrogates for BERT - e. g. ELMo [14].

SIF [1] is a strong baseline and viable alternative for constructing n-gram embeddings out of several tokens.

BERT can be fine-tuned, allowing the optimization of the network for an arbitrary downstream task. We consider this preferrable to our approach (training an XGBoost model), but lacked the resources to conduct the training in time.

Google's Universal Sentence Encoder [2] is another network-based method that could be used as a drop-in replacement for BERT, but seemingly supports neither contextual word embeddings - only sentence respectively document embeddings - and does not handle out-of-vocabulary items well.

A different direction for cross-referencing texts with MFD phrases is to use a technique like Guided Topic Modeling [15], which accepts a set of seed words per topic and composes topics by bundling words appearing in the same context as the specified seed words in topics.

## 5 Evaluation

### 5.1 Results

The metrics yielded by the model that was used to predict the affiliation with a political party are presented in the following - for raw precision, recall and aggregate measures see table 5.1. We also provide a precision-recall curve, a ROC curve and a confusion matrix in figure 2.

| **Moral Value** | **Precision** | **Recall** | **F1-score** | **Support** |
|---|---|---|---|---|
| Liberal/Democrat | 0.78 | 0.73 | 0.75 | 152 |
| Conservative/Republican | 0.68 | 0.74 | 0.71 | 121 |
| Micro avg. | 0.73 | 0.73 | 0.73 | 273 |
| Macro avg. | 0.73 | 0.73 | 0.73 | 273 |
| Weighted avg. | 0.74 | 0.73 | 0.73 | 273 |

This is a decent preliminary result considering that this was the first run and there are a couple of ways in which the results can be further improved - see section 6 for more information. Having said that, these results be seen as a starting point, but a deeper exploration and further performance tuning is definitely in order.

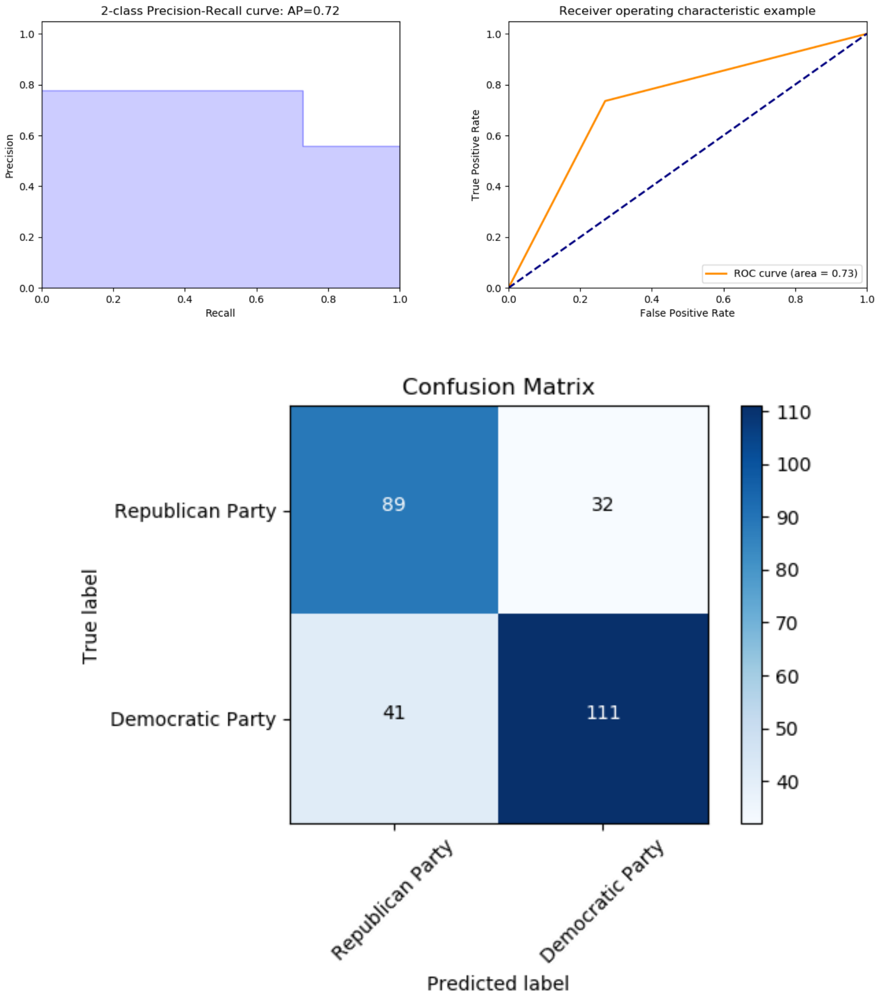


Fig. 2: ROC curve, precision/recall curve and confusion matrix for model predicting the political party based on MV relevance scores.

### 5.2 Examination of Feature Importance

Figure 3 shows the aggregated relative relevance of MVs for members of the Democratic and the Republican Party, respectively. Noticeably they are close to identical, and yet our model is able to separate Liberals and Conservatives decently well using very features. This leaves two options: Either those small differences are significant enough for the model to distinguish between Conservatives and Liberals linearly or those five features form clusters in higher-dimensional subspaces.

Either way, this result was surprising - we didn't expect an exact match with the MV weighting as presented in figure 1, but a coarse approximation. The divergence could be explained by various factors - errors in predicting MV priorities from embeddings, not factoring in emotional intensity, uneven weighting of seeds words in the MVD, perhaps the signal-to-noise ratio in tweets is simply too low for this kind of predictions - but exploring the reasons is considered to be future work.

We continued investigating the issue of feature importance utilizing SHAP [12], an approach to explaining the output of machine learning models. It offers a *force plot* showing how much individual features contributed to the classification of all datapoints. As can be seen in figure 4, SHAP displays some stark differences in effects of MVs between Liberals and Conservatives. Especially *care* can be seen as strongly pushing a politician's classification towards Liberalism. In general SHAP, in absolute values at least, indicates considerably more positive contributions towards Liberalism than towards Conservatism, which impedes a direct comparison.

## 6 Future Work

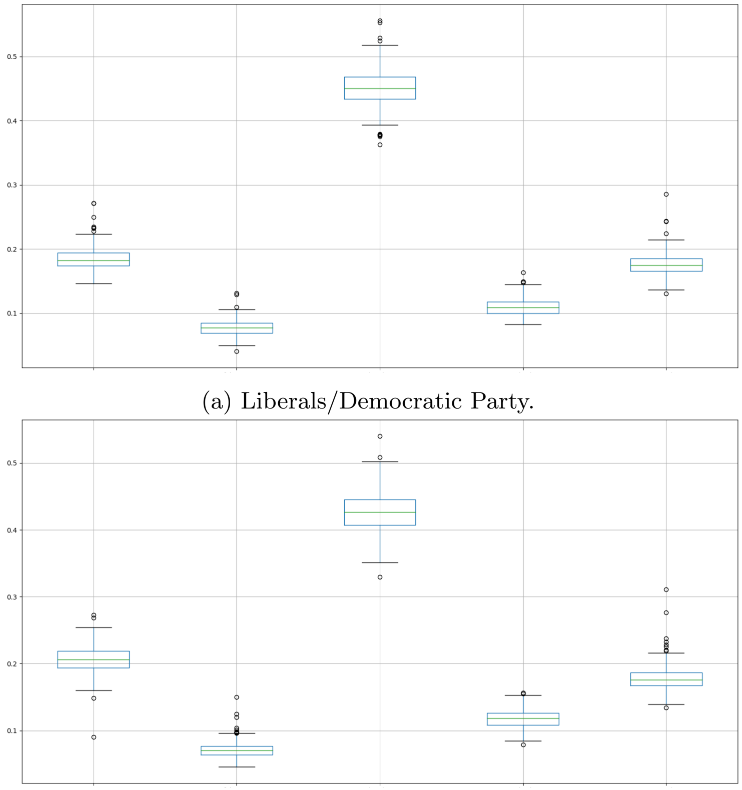
First and foremost, further exploring why a direct, simple statistical comparison shows moral preferences amongst Conservatives and Liberals to be almost identical could be immensely fruitful in advancing this line of research.

The techniques in the presented workflow most certainly can be further improved, chief amongst them the classification of context embeddings w.r.t. MVs - here, fine-tuning BERT towards this goal seems a more natural and direct way to train an accurate and efficient model. We used the base BERT model due to resource constraints - employing the larger variant might be beneficial.

The MFD contains a number of variations of the same words in different inflections, e. g. ”compassion” and ”compassionately”. While this might be useful for a one-to-one comparison, it's rather useless and perhaps actually has adverse effects for our approach. Inflected forms of the same word are expected to have similar embeddings. Even if the inflected form is an out-of-vocabulary word, i. e. it was not in the training set, its lexical similarity still implies a semantic similarity in character-based embedding models. Having multiple inflected forms of the same word in the MFD therefore only artificially increases the weight of this word and biases the embedding classifier towards it[10].

Another point of interest is that we weren't able to retrieve a version of the MFD containing seed words for the foundation *Liberty* - it stands to reason that introducing it would augment the model's predictions.

[10] Note that this paragraph solely reflects our intuition. We did not conduct explicit test to empirically confirm this.



(a) Liberals/Democratic Party.

(b) Conservatives/Republican Party.

Fig. 3: Boxplots showing the distribution of relevance scores. From left to right: Care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation.

Finally, we generated the tweet intensity scores discussed in subsection 4.1, but haven't exploited them in the party prediction model yet. Although we don't expect a strong change compared to our baseline performance, using emotional intensity as tweet weights might boost performance.

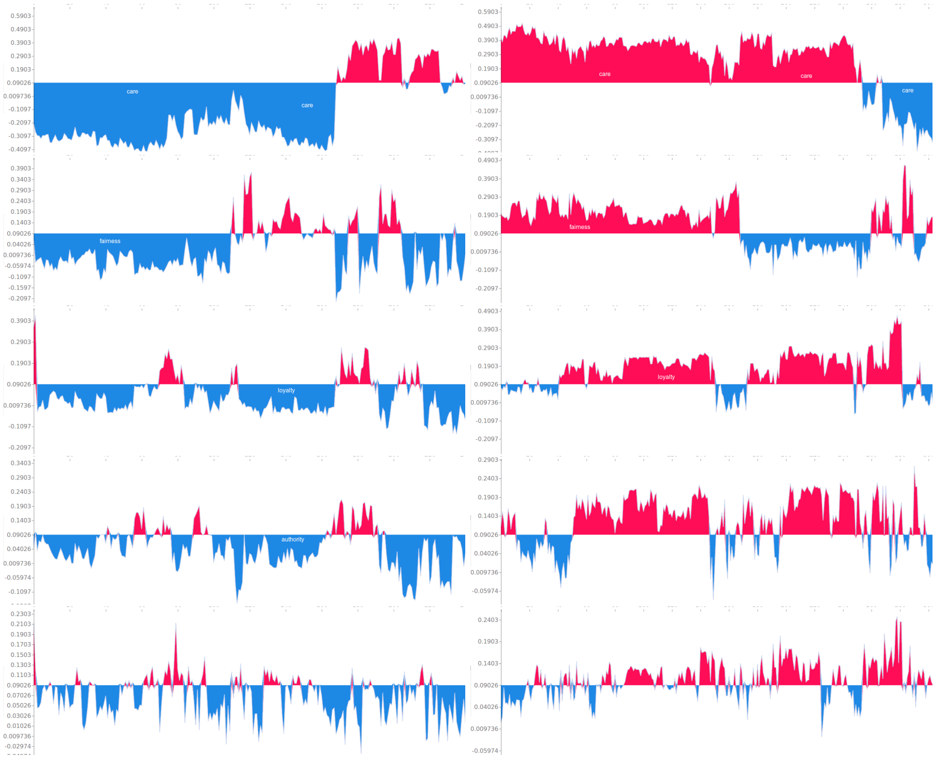


Fig. 4: Contribution of MVs as features to the classification of individual politicians with SHAP. Blue represents detrimental, red incremental effects. On the left: Republicans, on the right: Democrats. Each row represents one moral value, from top to bottom: Care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation. Sequences are sorted by similarity on the corresponding feature's importance.

## 7 Conclusion

We show that it is possible[11] to to classify politicians' political ideologies as Liberal or Conservative reasonably well with (1) state-of-the-art NLP embedding techniques, (2) a modified approach[12] for handling comparisons to moral seed phrases and (3) letting the classifier learn only from tweet-based estimations of the users' moral values.

[11] Under some simplifying conditions: We assume Republicans/Conservatives and Democrats/Liberals to be semantically interchangeable definitions in the context of US politics and politicians' tweets to be representative of their personal moral values.

[12] It might actually be a novel approach to estimating the priority of moral values in texts. Our lack of familiarity with this field prohibits a definitive claim though.

This problem setting involves at its core two principal sub-components: Extracting moral preferences and predicting political ideology based on these preferences. Of these two, we consider the extraction of moral values as a considerably more difficult issue, particularly so when analyzing tweets - which are short, have to be regarded as unrelated to each other and contain a lot of noise irrelevant in this context. A lack of available ground truth data - i. e. texts labeled with information on how relevant which moral values are in each text - only aggravates the issue.

The endeavour of extracting moral values from short, independent texts is a challenging topic and has some interesting sociopolitical applications in times of increasing partisanship in political matters. An augmented model could hence be fruitful for other lines of research as well. To this end we discuss some feasible improvements in section 6. We hypothesize that our model's performance to be lifted significantly given a more accurate representation of moral value preferences in the underlying set of tweets though.

## References

1. Arora, S., Liang, Y., Ma, T.: A Simple but Tough-to-Beat Baseline for Sentence Embeddings. Tech. rep., Princeton University (2017), https://openreview.net/forum?id=SyK00v5xx
2. Cer, D., Yang, Y., Kong, S.y., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.H., Strope, B., Kurzweil, R.: Universal Sentence Encoder (mar 2018), http://arxiv.org/abs/1803.11175
3. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System (mar 2016). https://doi.org/10.1145/2939672.2939785, http://arxiv.org/abs/1603.02754http://dx.doi.org/10.1145/2939672.2939785
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (oct 2018), http://arxiv.org/abs/1810.04805
5. Garten, J., Boghrati, R., Hoover, J., Johnson, K.M., Dehghani, M.: Morality Between the Lines : Detecting Moral Sentiment In Text (2016), https://www.semanticscholar.org/paper/Morality-Between-the-Lines-{%}3A-Detecting-Moral-In-Garten-Boghrati/6f640c4f10e8f1d39ed38a563ab3fe6301bf0735
6. Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S.P., Ditto, P.H.: Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. Advances in Experimental Social Psychology **47**, 55–130 (jan 2013). https://doi.org/10.1016/B978-0-12-407236-7.00002-4, https://www-sciencedirect-com.uaccess.univie.ac.at/science/article/pii/B9780124072367000024
7. Graham, J., Nosek, B.A., Haidt, J., Iyer, R., Koleva, S., Ditto, P.H.: Mapping the moral domain. Journal of personality and social psychology **101**(2), 366–85 (aug 2011). https://doi.org/10.1037/a0021847, http://www.ncbi.nlm.nih.gov/pubmed/21244182http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3116962
8. Haidt, J.: The righteous mind : why good people are divided by politics and religion. https://books.google.at/books?hl=en{&}lr={&}id=

ItuzJhbcpMIC{&}oi=fnd{&}pg=PR11{&}dq=the+righteous+mind{&}ots=H39aAFiv2y{&}sig=DiIKSdWLbL2lKZTzInlzP38hGh0{&}redir{_}esc=y{#}v=onepage{&}q=therighteousmind{&}f=false

9. Haidt, J., Graham, J., Dehgani, M., Boghrati, R., Frimer, J.: Moral Foundations Dictionaries for Linguistic Analyses, 2.0. Tech. rep. (2017), http://www.jeremyfrimer.com/uploads/2/1/2/7/21278832/summary.pdf
10. Hutto, C.J., Gilbert, E.: VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. ICWSM (2014), https://www.semanticscholar.org/paper/VADER{%}3A-A-Parsimonious-Rule-Based-Model-for-Analysis-Hutto-Gilbert/a6e4a2532510369b8f55c68f049ff11a892fefeb
11. Lin, Y., Hoover, J., Dehghani, M., Mooijman, M., Ji, H.: Acquiring Background Knowledge to Improve Moral Value Prediction (sep 2017), http://arxiv.org/abs/1709.05467
12. Lundberg, S.M., Erion, G.G., Lee, S.I.: Consistent Individualized Feature Attribution for Tree Ensembles (feb 2018), http://arxiv.org/abs/1802.03888
13. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space (jan 2013), https://arxiv.org/abs/1301.3781
14. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations (feb 2018), http://arxiv.org/abs/1802.05365
15. Singh, V.: How we Changed Unsupervised LDA to Semi-Supervised GuidedLDA (2017), https://medium.freecodecamp.org/how-we-changed-unsupervised-lda-to-semi-supervised-guidedlda-e36a95f3a164
16. U/Stuck_In_the_Matrix: Over one million tweets collected from US Politicians (President, Congress and Governors) : /r/datasets (2017), https://www.reddit.com/r/datasets/comments/6fniik/over{_}one{_}million{_}tweets{_}collected{_}from{_}us/