Which Reveals Ideology Better? Comparing Self-Presentation and Public Rhetoric in the Facebook Climate Debate via Embeddings Analysis

Luigi Arminio[1] and Luca Rossi[2]

[1] IT University of Copenhagen, Copenhagen 2300, Denmark
luia@itu.dk
[2] IT University of Copenhagen, Copenhagen 2300, Denmark
lucr@itu.dk

Abstract. Information about the ideological orientation of social media users is crucial to analyse a large number of social issues. However, the lack of structured data about users' beliefs is a common issue in social science. To address this gap, after a review of the state-of-the-art approaches for automated ideology detection, this research focuses on the use of a text-based methodology for this goal, using word embeddings. Specifically, the study contributes to the existing literature by focusing on ideology detection in the specific case of the polarized climate debate, and by testing the less-utilized OpenAI "ada" model for this task. Moreover, this research compares the accuracy of posts, representing users' public rhetoric, and page descriptions, thought to reveal their self-presentation, to predict users' ideological orientation. Our findings suggest that post-based methods hold the highest accuracy, but clustering-based approaches using page descriptions also yield respectable results while allowing a reduced use of computational resources. Overall, embedding-based methodology is shown to be a valuable tool for analyzing the ideological leanings of users in polarized debates.

Keywords: ideology · detection · embedding · polarization · climate

1 Introduction

When conducting social research, data on individuals' ideological leanings are invaluable for studying a number of social issues, including how political orientation drives macro-level social processes such as collective health behaviour (Geana, Rabb, and Sloman 2021; Hornsey et al. 2021), attitudes towards climate action (Arbuckle 2017), consumption practices (Buechner et al. 2022), political polarization processes (Rogowski and Sutherland 2016) or homophilic behaviour based on political inclinations among individuals in social networks (Del Valle 2022). However, social media studies often lack structured individual data (Chen et al. 2015): this also applies to the political orientation of the users (Conover et al. 2011). This data scarcity issue has led researchers to employ automated methods for estimating users' political orientation (Barberá 2015; Bond and Messing 2015). Some of these techniques rely on the network patterns that characterize social media interaction and start from homophily assumptions, but their accuracy may vary due to differing levels of homophily across platforms and interactions (Hanteer et al. 2018). Furthermore, the availability of data for network construction is often limited. Alternative ideology detection approaches rely on users' social media posts' textual content, but the most traditional text-based approaches exhibited other significant limitations, such as the weakness in pinpointing words' contextual meanings and the necessity for prelabeled data to train the algorithms (Chinn, Hart, and Soroka 2020). In this sense, in recent years, advanced deep-learning models like word embeddings seemed to offer a promising alternative, although their employment remains relatively limited, although their employment remains relatively limited, primarily confined to a restricted number of political domains and types of textual content, relying on a narrow range of embedding models, neglecting potential advancements. In this context, our research seeks to extend the current understanding of ideology identification from textual data in the social media context through the pursuit of three delineated research goals: firstly, focusing on ideological polarity detection in the context of a specific social cleavage (i.e., the climate discourse) that has been underexplored in the context of embedding-based ideology identification. Second, to test the use of a newer language model that is more recent than the traditionally employed ones in embedding-based research on ideology. Lastly, to compare the efficiency of predicting users' ideology based on their self-presentation (how they describe themselves on Facebook) versus an approach that employs their public rhetoric (their published posts) as textual material for the embedding algorithms.

2 State of the art

Existing computational social science literature documents various approaches for automated detection of political orientation in social media actors. These methods can be categorized through various lenses. We can classify methods based on their focus on analysis, distinguishing between behaviour-based (focused on how the users act) and identity-based (focused on how the users identify themselves) approaches, or based on the nature of the analysed data, distinguishing between network-based and text-based methods.

Network-based methods include Bayesian ideal point estimation, initially employed with roll call data (Bafumi et al. 2005). In its application in social media studies, it uses MCMC methods to map users on a continuous political spectrum based on a user-to-user adjacency matrix built on follower-followee connections (Barberá 2015). An analogue network-based approach applies singular value decomposition (SVD) on Facebook data, extracting a continuous and unidimensional political spectrum from a user-to-user matrix and labeling users based on their connections to Facebook pages (Bond and Messing 2015).

Within text-based methods, Wordfish stands out as an unsupervised text scaling model that treats ideology as a latent variable (Chinn, Hart, and Soroka 2020; Hjorth et al. 2015) inferred from the semantic choices of the different political actors (Proksch and Slapin 2009). While this method has been repeatedly employed in social media studies (Ceron 2017), it resulted to struggle with shorter texts and stylistic nuances (Ruedin and Morales 2019), which may pose a significant limitation when working with social media data. An alternative approach is Wordscores, a supervised scaling model that identifies political ideology based on a predefined number of dimensions and labelled textual political data (Laver, Benoit, and Garry 2003). However, Wordscores requires extensive labeled datasets (Ruedin and Morales 2019), which limits its flexibility, and it exhibited poorer performance when working with extreme ideologies (Ruedin and Morales 2019).

The constraints of these methods substantiate the importance of exploring alternative advanced approaches, such as the embedding-based methodology. Embedding-based methods, that employ deep learning to convert text (such as the textual content of politically relevant Facebook posts, or the Facebook description of the official page of a political party) into vector representations to effectively capture semantic meaning and contextual relationships within textual documents (Jurafsky and Martin 2008). These methods have been tested both with parliamentary corpora (Rheault and Cochrane 2020) and social media text data (Rao and Spasojevic 2016), but their utilization remains somewhat limited: the study of their performance in the context of automated ideology detection is relegated to a reduced number of studies referring to specific dimensions of the political spectrum to measure political ideology, such as the US *Democratic-Republic* dichotomy (Rao and Spasojevic 2016).

This limited range of application highlights concerns tied to theoretical sociological considerations. By referring to the theory of cleavages, that identifies a range of societal divisions shaping political discourse (Lipset and Rokkan 1967), as the reference theoretical model for political ideology analysis, social polarization can be elucidated as an expression of existing social cleavages (Casañ et al. 2022). In this context, it is noteworthy to acknowledge the presence of several different cleavages shaping the public discourse, which also increased in number in recent years (Alter and Madsen 2021; Ramaciotti Morales 2022; Jahn 1993): the continuous emergence of new cleavages, each exhibiting diverse socio-demographic profiles and behavioral patterns (Lonsdale 2023), which may even be associated with variations in language production (Pelinka 2007; Dassonneville, Fréchet, and Liang 2022), suggests that methodologies that work for text-based ideology detection within one social divide, such as the Republican-Democratic divide, may not necessarily valid when applied to other divides, such as the cleavage on environmental values. More specifically, success in text-based ideological detection achieved within one ideological divide, like the Republican-Democrat dichotomy, cannot be assumed to translate equally across other polarized debates (such as the climate debate) due to the unique lexical items, semantic nuances, and potentially varying communicational tones that characterize each specific debate.

To address the challenge of applying text-based ideological detection within specific social cleavages, this research will focus on a specific politically polarized area of the public debate and conduct an in-depth exploration of how word embedding performs in detecting ideological positions of the actors within this divide. In this study, we focused on the polarized climate change debate, that emerged as a contentious social cleavage in the latest decades (Alter and Madsen 2021). The choice to focus on the social divide around climate change stems from the significant impact of this cleavage in reshaping political landscapes, with shifts in voting behaviour and increased engagement in environmental activism by the younger generations (Alter and Madsen 2021; Ramaciotti Morales 2022; Jahn 1993; Wallis and Loy 2021).

Furthermore, the current literature on embedding-based ideology detection mainly relies on traditional embedding models such as BERT (Baly et al. 2020), Word2Vec (Gurciullo and Mikhaylov 2017), and to some extent, ELMo (Jiang et al. 2019), and GloVe (Kamboj et al. 2020). In this context, we suggest that the inclusion of novel embedding models, such as those recently developed by OpenAI, like Ada (Neelakantan et al. 2022), may provide a relevant contribution to this field.

Concerning other limitations in the current literature on ideological detection, it is worth mentioning the lack of a comprehensive analysis focusing on the choice of the type of textual content to predict users' political orientation. In effect, in social media contexts, users release multiple types of data, that are structurally different. While previous studies employed embedding-based ideological detection in different ways, there remains a gap in understanding which specific kinds of social media text may be more predictive for this task. We believe that in contexts like Facebook, where we can access both textual data concerning the self-presentation of the users, given by their account descriptions, and their public rhetoric, given by the textual content that they produce through the posts, it is useful to determine which performs better in predicting ideology through text embeddings. For these reasons, we decided to compare the embedding-based ideological detection on these two kinds of available textual data.

3 Data and methods

3.1 Data collection

To pursue our research goal, we manually compiled a list of politically relevant Facebook pages within the climate debate from multiple countries, distinguishing between those advocating for climate action (pro-actors) and those promoting climate denial or skepticism (counter-actors). Data was collected through the Crowdtangle API from 113 actors and 58 counter-actors, encompassing both their Facebook page descriptions (*i*) and the textual content of their posts (*ii*) over a four-month period (from April 29 to August 29 2023).

3.2 Ideological scores generation and validation

Both textual datasets went through translation into English using the *deep_ translator* Python library (Kareem 2023) for linguistic uniformity. Then, the two corpora went through word embedding utilising the *ada* language model (Neelakantan et al. 2022), that allowed us to represent each textual document (post or page description) in a space of 1531 continuous dimensions. After that, we employed t-SNE (Van der Maaten and Hinton 2008) to project the embedded documents into one single dimension in order to transform the textual content of each post in a single scalar value, aimed at politically positioning the users along a continuous scale based on their textual content. In the case of the corpus of posts, it has been necessary to aggregate all the scores from posts of each same page by average, in order to have a single score for each stakeholder. At this point, each stakeholder was labeled through two continuous values: the Content score, representing the average projected embeddings of the published posts by a given stakeholder, and the Self-description score, given by the projected embedding value of the page description of each stakeholder.

After labeling each stakeholder with the two embedding-generated scores, we assessed their accuracy in reflecting their latent ideological position by analyzing the association between each of the two continuous scores and the manually annotated actor position (pro-actor or counter-actor). In light of the continuous nature of the embedding-based scores and the dummy nature of the actor status, we resorted to the Point-Biserial correlation (PB) (Kornbrot 2014). We specifically focused on the absolute value of the PB coefficient, in light of the absence of directional assumptions, to solely assess how effectively the embedding-based scores differentiated between users from distinct classes (pro and counter actors) in a way that mirrors the ground truth's classification, without assuming any predefined relationship between high or low embedding values and a specific class (pro or counter actor). After that, we assessed our data-driven scores' performance in determining actors' ideological polarity within the "pro-actors vs. counter actors" framework by categorizing our word embedding-generated scores.

3.3 Categorization methods and performance evaluation

Initially, we employed and compared two methods to categorize the scores. First, we resorted to mean-based dichotomization, assigning 0 to scores below the mean and 1 to scores above it. Second, we utilized threshold optimization, dividing the continuous score range into 100 equidistant points and then identifying a threshold holding the highest Normalized Mutual Information (NMI) (Forbes 1995) with the dichotomous ground truth (the pro-counter actor status).

The mean-based approach was used as a simple heuristic method for actor classification based on observed score distributions, which tend to be unimodal. The second one, named threshold optimization, maximizes the predictive power of our scores by dichotomizing data based on the ground truth.

After that, a third approach was involved, bypassing dimensionality reduction and instead employing. k-means clustering across all 1531 dimensions of word embeddings. The value of k was predetermined as 2 based on the premise of a polarized ideological environment (pro vs counter actors of climate action) rather than relying on data-driven optimization techniques, aiming to assess the embeddings' performance in accurately capturing the inherently dichotomous and polarized ground truth.

Table 1. Summary of Tested Methods for Ideology Detection

Data	**Analysis Focus**	**Categorization Method**
Posts	Public Rhetoric	t-SNE & Threshold Optimization
		t-SNE & Mean-Based Dichotomization
Descriptions	Self-Presentation	t-SNE & Threshold Optimization
		t-SNE & Mean-Based Dichotomization
Posts	Public Rhetoric	k-means Clustering (k=2)
Descriptions	Self-Presentation	k-means Clustering (k=2)

Subsequently, we assessed the alignment between the categorizations generated from the three methods and verified actor status using NMI and Matthews correlation coefficient (MCC) (Matthews 1975) . We chose to employ both MCC and NMI to provide comprehensive evaluation of our text-driven categorizations: while NMI measures overall agreement between embedding-based categorizations and ground-truth, MCC ensures balanced performance assessment, accommodating class imbalances (Boughorbel, Jarray, and El-Anbari 2017). This was applied to both the data types (descriptions and posts) and all three categorization approaches (optimization-based, mean-based, clustering-based). We only utilized the absolute value of MCC to focus on association magnitude, mirroring our approach with the point-biserial correlation. After that, we selected the optimal model (i.e., the one that showed the highest performance in terms of the above-mentioned metrics across different data types and categorization approaches).

Furthermore, we determined that if any of the content-based classifications (i.e., relying on the released posts) showed meaningful performance, we would have selected the most performative one to measure its sensitivity to time patterns and stylistic variations in content production. The decision to examine the performance sensitivity of the Content-based approach with respect to temporal variations stemmed from the premise that different time scales could influence political entities, leading to language shifts or evolving lexicon due to social events and political demands. To analyze this, we split our dataset, spanning months, into four subsets representing each month. We then computed key performance parameters, like NMI and MCC, for each month separately and for cumulative periods (e.g., first month, first and second month, and so on) to assess the hypothesized temporal variations of the classification quality. Average post length was calculated as the mean word count per post.

As mentioned before, in addition to the temporal sensitivity analysis, we evaluated factors impacting the optimal content-based classification performance. This involved analysing three parameters: posting frequency, average post length, and content diversity in published posts. Posting frequency was measured as the total number of posts per page within the specified time window, capped at 100 posts per page due to the data collection constraints. Content diversity was computed by constructing a corpus for each page, encompassing their published posts within the above-mentioned time frame. We treated each user's set of post embedding scores as a cluster, computing an average embedding vector as the page's centroid. Then, we determined content diversity as the average cosine distance between all posts embeddings on a page and its centroid.

After defining these three parameters, we estimated the relationship between classification accuracy of the actors and each parameter in order to assess their potential influence on the classification performance of the optimal Content-based approach.

4 Results

4.1 Overview of ideological categorization approaches

Both the textual content of the Facebook posts and page descriptions provided valuable insight into categorizing the stances of the main stakeholders involved in the polarized climate debate. In particular, the optimal approach emerged be the one that joins word embeddings and dimensionality reduction for continuous scoring of ideological polarity based on Facebook posts. However, the clustering-based approach's notable performance with page descriptions should be noted: in cases of limited computational resources: this method, by demanding fewer data points than post-based approaches, can facilitate efficient data processing while keeping respectable performance with respect to the optimal approach, while acknowledging a slight compromise in classification accuracy.

Concerning the third method, which employed k-means clustering, it showed more balanced performance with different data types (i.e., posts and descriptions), but none of the applications outperformed the optimal t-SNE-based approach on users' posts. As for the drawbacks of k-means approaches, he absence of a metric (like the Content score) to place actors on a continuous ideological spectrum, which may provide some research benefits. For example, in case of extension of the binary classification model to a tripartite model including a third category between the two extremes (i.e., referring to stances that are not explicitly identifiable as pro-climate or anti-climate action), continuous ideological scoring would facilitate the exploration and the detection of this middle ground. In effect, this approach would provide better control than the k-means-based one, where introducing a third cluster may yield uncertain outcomes in detecting an intermediate ideological class.

4.2 Ideological continuous scoring performance

As for the continuous scores generated through dimensionality reduction of the embeddings, Content score showed a significantly stronger PB correlation (0.66) with the true label than the Self-dscription score (0.22). Following score dichotomization, the performance gap between content produced and self-presentation further increases, with the NMI consistently higher for the Content score in both mean-based and optimization-based methods when compared to actor status.

However, when dividing users into two distinct ideological categories through k-means performed on the entire set of features (i.e., the 1531 embedding columns), the self-description of the users proved effective in estimating uers' ideological polarity, albeit with lower performance compared to the projection-based approach with threshold optimization on posts. In this sense, both the self-presentation of the climate debate stakeholders on social media and their public rethoric on Facebook showed some degree of effectiveness in predicting their ideological polarity.

Overall, the content-based embedding generated the best predictive performance and resilience to variations in the classification methodology, albeit with varying results. On the other hand, although clustering-based approach with Facebook pages descriptions yielded slightly lower classification performance compared to the optimal post-based classification, it still achieved satisfactory performance while effectively managing computational resources as a result of the reduction of the demanded data volume.

4.3 Influential factors on ideological classification performance

After comparing ideological detection with different data types and dichotomization processes, given the meaningful results obtained in ideological classification based on the public rhetoric of social media users, we examined how its predictive and classificational capabilities vary across different temporal scopes. This in-depth examination stems from the assumption that different time windows, due to their unique durations and political occurrences, may yield varying amounts of posts and linguistic nuances, potentially influencing the approach's performance.

First, we recomputed the score for individual months (e.g., the first month, the second month, etc.). Subsequently, we cumulatively aggregated posts and recalculated ideological scores for each aggregated data subset (the first month of data, the first two months, etc.). After that, we recomputed performance parameters for each time frame and assessed the differences in classification and predictive performance of the Content Score. We employed the PB Correlation

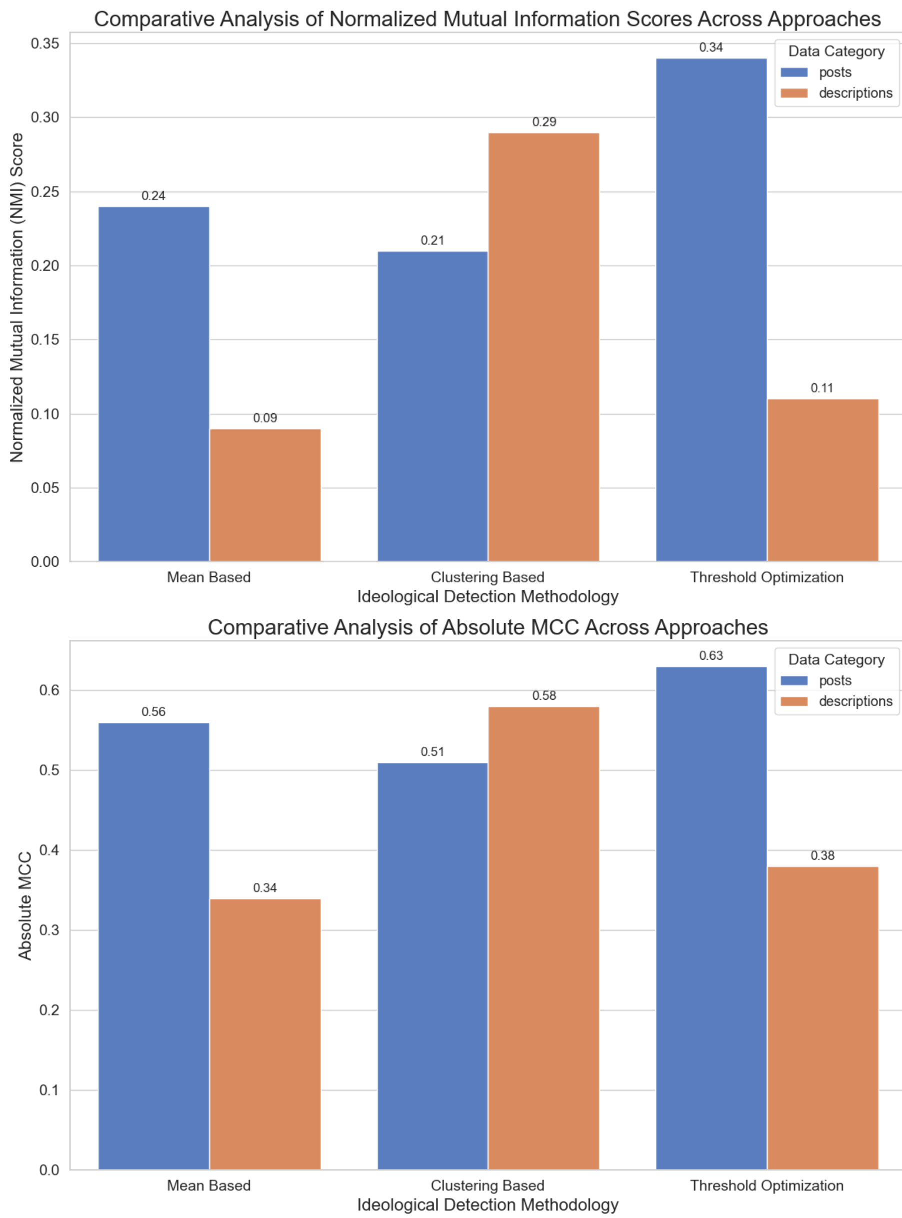

Fig. 1. Comparing approaches for categorizing ideological polarity via word embeddings. Clustering-based methodology provides the most balanced results across diverse data types, but threshold optimization of projected embedding scores on posts performs best.

between our score and the dummy ground truth as a metric for score performance (that we called *PB Performance*) and then, we calculated the Pearson Correlation between the number of posts within each time window and the PB Performance of each time-specific Content Score.

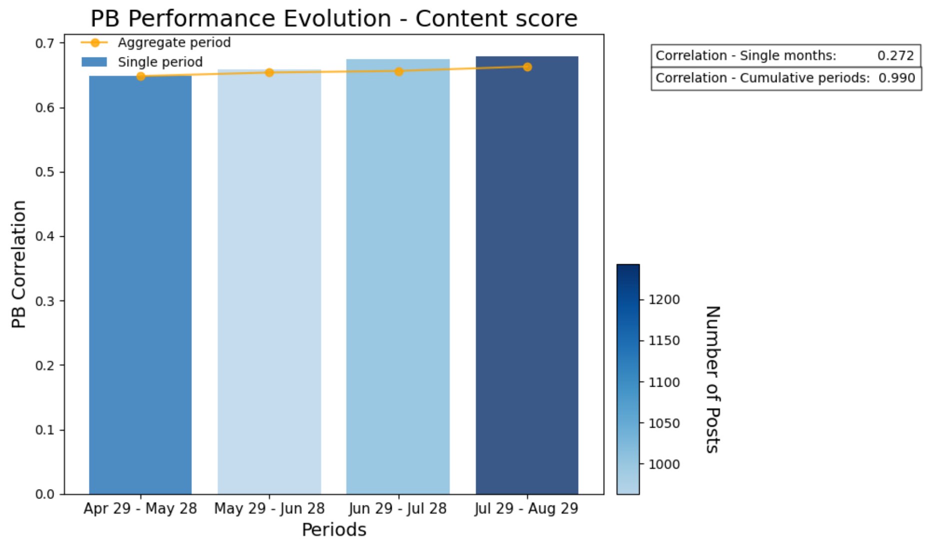

Fig. 2. PB Performance of Content Score over months. Bars represent PB Performance for each month, while dots indicate cumulative PB Performance (first month, first two months, etc.). Bar color tones reflect the number of posts per period.

Overall, different time periods have shown different levels of performance: the most recent months shown stronger association between the Content score and stakeholders' true labels. Furthermore, the number of released posts in a given month partially affected the performance of the Content score: when the dataset is segmented into four individual monthly periods, the correlation between the number of posts in a month and score performance is 0.27; when the correlation between the number of posts in each cumulative aggregation of months and the model performance, the correlation coefficient reaches 0.99. These findings suggest that the model performance is affected by features of the examination period. However, further investigations over prolonged time intervals are required to validate this observation.

Subsequently, our analysis focused on our dichotomized Content score, to measure the elements affecting its efficacy in aligning with the ground truth in estimating the ideological polarity of each Facebook page. We considered three potential factors affecting the probability of correct classification: the posting frequency of each page, the average length of the posts, and the diversity of the published content.

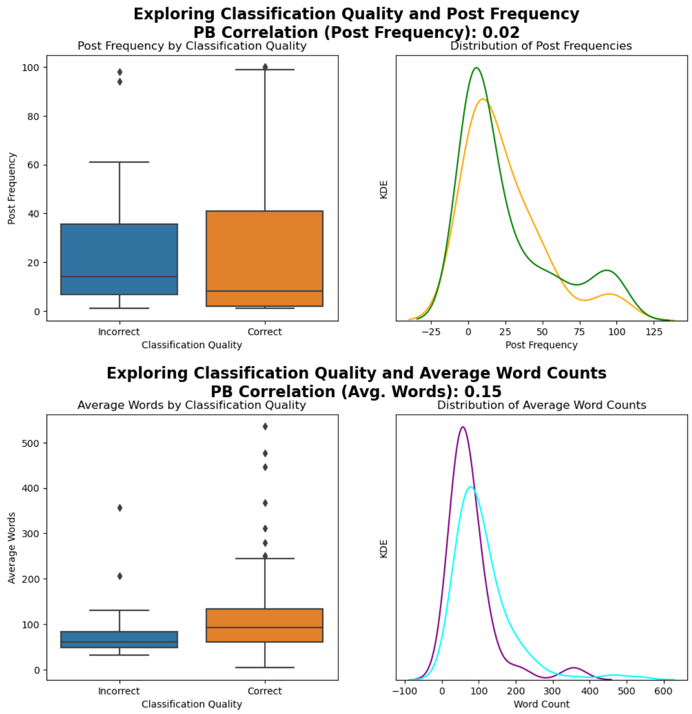

Fig. 3. Classification quality and user activity: The Content Score (post-based) shows that posting frequency does not significantly affect ideological polarity estimation on Facebook, while there is a weak positive correlation with average post length.

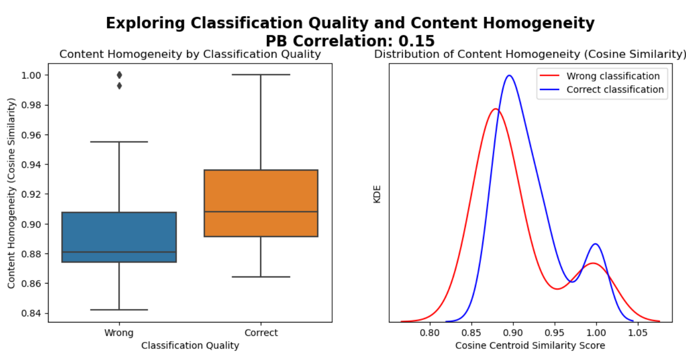

Fig. 4. Classification quality vs Content Homogeneity - The match between dichotomized Content Score and ideological ground truth is slightly affected by the diversity level of published content by each climate debate stakeholder

Our results indicate that, after Content Score dichotomization, rhetoric-based ideological polarity prediction among social media users was resilient to variations in terms of factors such as posting frequency of the Facebook pages: although misclassified pages exhibited a skewed distribution toward lower values, no statistically significant differences were observed in average tendencies. The approach maintained substantial resilience to variety of the produced content and average post length, though both higher posts homogeneity and average length exhibited a slight positive correlation with the effective categorization of the pages

5 Discussion and conclusion

In conclusion, our study provided compelling insights about the use of word embedding to determine the ideological polarity of social media actors in the context of the climate change debate. Our goal was to broaden the current knowledge of word embedding for ideological classification by examining a specific polarizing issue, the climate debate, which has been under-explored in the analysis of embedding-based methods for ideology detection. Our approach involved applying the OpenAI “ada" model, a promising yet underutilized method in this context, to assess the performance of embedding-based ideological classification with different textual data sources. In particular, we collected both users' page descriptions and their released posts over a four-month period. This allowed us to capture and compare two two distinct dimensions of social media users' textual production: their public rethoric (given by their published posts), and their self-presenation (given by their page descriptions) to determine which better predicts users' ideological polarity. Then, starting from the word embedding outcomes, we implemented three distinct approaches to bifurcate users into two polarized categories: two employing dimensionality reduction for continuous ideological scoring and one integrating embedding with k-means clustering for a two-cluster model representing the ideological polarization among the users. Our analysis revealed notable findings for both self-presentation and public-rhetoric in the context of ideological classification. In particular, the top-performing approach was the integration of word embeddings with dimensionality reduction applied to users' posts. The clustering-based approach applied to user descriptions was the second best method in terms of performance.
Overall, approaches relying on users' public rethoric (posts) showed higher resilience to methodological variations but were more computationally demanding. This textual embedding-based approach gains significance when the lack of interactional data prevents researchers from mapping the network dynamics among political actors or social media users as well as when the lack of homophilic patterns in social media interactions might cause network-based ideological estimation to fail. In light of the significance of both the produced content and the self-description of the pages in estimating ideology, the selection of the best approach should be guided by the available computational resources and the

interest in the use of a continuous score to place actors on an ideological scale.

Concerning the limitations, data constraints prevented including a neutral third group between pro-climate actors and counter-actors, which could have enriched our ideological classification. Moreover, the rhetoric-based ideological detection varied with the dataset's temporal span, as unique periods with specific discussion themes and posting frequencies impacted Content Score performance. Moreover, longer time spans boosted the quality of ideological predictions. At the same time, content-based ideology detection showed resilience to content variation.
In terms of future developments, this analysis can be replicated using other embedding models traditionally employed in ideology detection and social media studies, such as BERT and Word2Vec. This approach may allow us to compare the effectiveness of different embedding models in ideological classification of climate debate stakeholders based on either their self-presentation or their public rhetoric. Furthermore, given the variety of textual content that users produce on social media, including different tones, styles and register, it may be relevant to measure and explore how the subjectivity level of the produced content (Yaqub et al. 2018) can affect correct text-based ideological classification: understanding this relationship could increase the precision and reliability of ideology detection and may provide a better understanding of the interplay between discursive tone and the expression of specific ideological stances.

6 Supplementary material

Due to privacy and data protection reasons, the collected data through Crowdtangle during the current study are not publicly accessible.

However, the code employed in this study is available in our repository at *github.com/arminioluigi/ideology_estimation*.

References

Alter, K. J. and M. R. Madsen (2021). "The International Adjudication of Mega-Politics". In: *Law & Contemp. Probs.* 84. DOI: 10.2139/ssrn.3961417, p. 1.

Arbuckle, M. B. (2017). "The interaction of religion, political ideology, and concern about climate change in the United States". In: *Society & Natural Resources* 30.2. DOI: 10.1007/s11109-015-9323-7, pp. 177–194.

Bafumi, J. et al. (2005). "Practical issues in implementing and understanding Bayesian ideal point estimation". In: *Political Analysis* 13.2. DOI: 10.1093/pan/mpi010, pp. 171–187.

Baly, R. et al. (Nov. 2020). "We can detect your bias: Predicting the political ideology of news articles". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. DOI: 10.18653/v1/2020.emnlp-main.404, pp. 4982–4991.

Barberá, P. (2015). “Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data”. In: *Political analysis* 23.1. DOI: 10.1093/pan/mpu011, pp. 76–91.

Bond, R. and S. Messing (2015). “Quantifying social media’s political space: Estimating ideology from publicly revealed preferences on Facebook”. In: *American Political Science Review* 109.1. DOI: 10.1017/S0003055414000525, pp. 62–78.

Boughorbel, S., F. Jarray, and M. El-Anbari (2017). “Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric”. In: *PloS one* 12.6. DOI: 10.1371/journal.pone.0177678, e0177678.

Buechner, B. M. et al. (2022). “Political ideology and cultural consumption: The role of flexibility in shaping liberal and conservative preferences for global-local experiences”. In: *Journal of the Association for Consumer Research* 7.3. DOI: 10.1080/08941920.2016.1209267, pp. 266–275.

Casañ, R. R. et al. (2022). “Online polarization and cross-fertilization in multi-cleavage societies: the case of Spain”. In: *Social Network Analysis and Mining* 12.1. DOI: 10.1007/s13278-022-00909-5, p. 79.

Ceron, A. (2017). “Intra-party politics in 140 characters”. In: *Party politics* 23.1. DOI: 10.1177/1354068816654325, pp. 7–17.

Chen, X. et al. (2015). “A comparative study of demographic attribute inference in twitter”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 9. 1. DOI: 10.1609/icwsm.v9i1.14656, pp. 590–593.

Chinn, S., P. S. Hart, and S. Soroka (2020). “Politicization and polarization in climate change news content, 1985-2017”. In: *Science Communication* 42.1. DOI: 10.1177/1075547019900290, pp. 112–129.

Conover, M. D. et al. (Oct. 2011). “Predicting the political alignment of twitter users”. In: *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. DOI: 10.1109/PASSAT/SocialCom.2011.34. IEEE, pp. 192–199.

Dassonneville, R., N. Fréchet, and B. Liang (2022). “Linguistic cleavages in public opinion”. In: *Comparative public opinion*. DOI:10.4324/9781003121992-18. Routledge, pp. 284–309.

Del Valle, M. E. (2022). “Homophily and Polarization in Twitter Political Networks: A Cross-Country Analysis”. In: *Media and Communication* 10.2. DOI: 10.17645/mac.v10i2.4948, pp. 81–92.

Forbes, A. D. (1995). “Classification-algorithm evaluation: Five performance measures based on confusion matrices”. In: *Journal of Clinical Monitoring* 11. DOI: 10.1007/BF01617722, pp. 189–206.

Geana, M. V., N. Rabb, and S. Sloman (2021). “Walking the party line: The growing role of political ideology in shaping health behavior in the United States”. In: *SSM-population health* 16. DOI: 10.2139/ssrn.3961417, p. 100950.

Gurciullo, S. and S. J. Mikhaylov (Oct. 2017). “Detecting policy preferences and dynamics in the un general debate with neural word embeddings”. In:

2017 International Conference on the Frontiers and Advances in Data Science (FADS). DOI: 10.18653/v1/S19-2146, pp. 74–79.

Hanteer, O. et al. (Aug. 2018). “From interaction to participation: The role of the imagined audience in social media community detection and an application to political communication on twitter”. In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. DOI: 10.1109/ASONAM.2018.8508575. IEEE, pp. 531–534.

Hjorth, F. et al. (2015). “Computers, coders, and voters: Comparing automated methods for estimating party positions”. In: *Research & Politics* 2.2. DOI: 10.1177/2053168015580476, p. 2053168015580476.

Hornsey, M. J. et al. (2021). “Resolving the small-pockets problem helps clarify the role of education and political ideology in shaping vaccine scepticism”. In: *British Journal of Psychology* 112.4. DOI: 10.1111/bjop.12500, pp. 992–1011.

Jahn, D. (1993). “The rise and decline of new politics and the Greens in Sweden and Germany: Resource dependence and new social cleavages”. In: *European Journal of Political Research* 24.2. DOI: 10.1111/j.1475-6765.1993.tb00375.x, pp. 177–194.

Jiang, Y. et al. (June 2019). “Team bertha von suttner at semeval-2019 task 4: Hyperpartisan news detection using elmo sentence representation convolutional network”. In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. DOI: 10.18653/v1/S19-2146, pp. 840–844.

Jurafsky, D. and J. Martin (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Vol. 2. Prentice Hall.

Kamboj, M. et al. (2020). “Multimodal political deception detection”. In: *IEEE MultiMedia* 28.1. DOI: 10.1109/MMUL.2020.3048044, pp. 94–102.

Kareem, N. (2023). “Deep-translator (Version 1.11.4)”. In: *PyPI*. Retrieved January 10, 2024, from pypi.org/project/deep-translator/.

Kornbrot, D. (2014). “Point biserial correlation”. In: *Wiley StatsRef: Statistics Reference Online*. DOI:10.1002/0470013192.bsa485.

Laver, M., K. Benoit, and J. Garry (2003). “Extracting policy positions from political texts using words as data”. In: *American political science review* 97.2. DOI: 10.1017/S0003055403000698, pp. 311–331.

Lipset, S. M. and S. Rokkan (1967). “Cleavage structures, party systems, and voter alignments: An introduction”. In: *Party Systems and Voter Alignments: Cross-National Perspectives*. The Free Press, pp. 1–64.

Lonsdale, A. (2023). “Sociodemographic Inequalities and Political Cleavages: An Investigation of Recent Elections, Nativist Movements, and Individual-Level Voting Shifts”. CID: 20.500.12592/p2ngm6b. PhD thesis. ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES.

Matthews, B. W. (1975). “Comparison of the predicted and observed secondary structure of T4 phage lysozyme”. In: *Biochimica et Biophysica Acta*

(BBA)-Protein Structure 405.2. DOI: 10.1016/0005-2795(75)90109-9, pp. 442–451.

Neelakantan, A. et al. (2022). "Text and code embeddings by contrastive pre-training". In: *arXiv preprint arXiv:2201.10005.* 10.48550/arXiv.2201.10005.

Pelinka, A. (2007). "Language as a political category: The viewpoint of political science". In: *Journal of Language and Politics* 6.1. DOI: 10.1075/156921507781509626, pp. 129–143.

Proksch, S. O. and J. B. Slapin (2009). "How to avoid pitfalls in statistical analysis of political texts: The case of Germany". In: *German Politics* 18.3. DOI: 10.1080/09644000903055799, pp. 323–344.

Ramaciotti Morales, P. (Nov. 2022). "Multidimensional online American politics: Mining emergent social cleavages in social graphs". In: *International Conference on Complex Networks and Their Applications.* DOI: 10.1007/978-3-031-21127-$0_1$5. Cham: Springer International Publishing, pp. 176–189.

Rao, A. and N. Spasojevic (2016). "Actionable and political text classification using word embeddings and LSTM". In: *arXiv preprint arXiv:1607.02501.* DOI: 10.48550/arXiv.1607.02501.

Rheault, L. and C. Cochrane (2020). "Word embeddings for the analysis of ideological placement in parliamentary corpora". In: *Political Analysis* 28.1. DOI: 10.1017/pan.2019.26, pp. 112–133.

Rogowski, J. C. and J. L. Sutherland (2016). "How ideology fuels affective polarization". In: *Political behavior* 38. DOI: 10.1007/s11109-015-9323-7, pp. 485–508.

Ruedin, D. and L. Morales (2019). "Estimating party positions on immigration: Assessing the reliability and validity of different methods". In: *Party Politics* 25.3. DOI: 10.1177/1354068817713122, pp. 303–314.

Van der Maaten, L. and G. Hinton (2008). "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.11, pp. 2579–2605.

Wallis, H. and L. S. Loy (2021). "What drives pro-environmental activism of young people? A survey study on the Fridays For Future movement". In: *Journal of Environmental Psychology* 74. DOI: 10.1016/j.jenvp.2021.101581, p. 101581.

Yaqub, U. et al. (May 2018). "Analysis and visualization of subjectivity and polarity of Twitter location data". In: *Proceedings of the 19th annual international conference on digital government research: governance in the data age.* DOI: 10.1145/3209281.3209313, pp. 1–10.