Tübingen at PoliticIT: Exploring SVMs, Pretrained Language Models, and Linguistic Transfer for Ideology Detection in Social Media

Çağrı Çöltekin[1], Matteo Brivio[1] and Fidan Can[1]

[1]University of Tübingen, Department of Linguistics

Abstract

This paper describes our approach to the EVALITA 2023 PoliticIT task on predicting political ideology and gender from Italian tweets. Furthermore, we investigate the effects of out-of-domain data (transcripts of parliamentary speeches) and cross-lingual transfer from Spanish. Overall, our simple traditional SVM classifier performed best according to the shared task evaluation, ranking first in ideology detection and second in gender identification. We also demonstrate promising results for out-of-domain data and cross-lingual transfer learning.

Keywords

Political Ideology Detection, classification, SVM, BERT, XLM-Roberta, Transfer learning, Multitask learning

1. Introduction

Although difficult to define, a relatively uncontroversial definition of political ideology is 'a set of beliefs about the proper order of society, and how it can be achieved' [1, 2]. As the definition suggests, people's political ideologies, and especially those of politicians, have a significant impact on society. Much like other personal characteristics, such as gender, age, or native language [3, 4, 5], political ideology can help to understand individual and social behavior [6, 7, 8].

Besides analyzing social and political discourse, detecting ideology is a crucial step for correctly understanding political texts. Certain words or phrases have different intended meanings depending on the ideological position of speakers or authors [9, 10], and a message can also be understood differently based on the audience. In its extreme case, so-called 'dog whistles' [11] may allow politicians to send 'coded messages' to only a part of the public who share their ideological position. As a result, identifying political ideology is also important for natural language understanding tasks.

This paper describes our contribution to the EVALITA 2023 [12] shared task [13] on predicting political orientation and gender of politicians from social media posts in Italian. The task is defined as three related classification sub-tasks: (1) binary political orientation (`left`, `right`), (2) fine-grained political orientation (`left`, `moderate_left`, `moderate_right`, `right`) and (3) binary gender (`female`, `male`). We evaluated both simple linear classifiers and models based on deep pretrained networks. We also experimented with additional resources from different domains (parliamentary speeches) and similar datasets in another language (Spanish).

The rest of this paper reports on our approach and results. The code, the models hyper-parameters, and any additional resources used in our contribution can be found at https://github.com/fidan-c/PoliticIT23-ideology-detection.

EVALITA 2023: 8th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, Sep 7 – 8, Parma, IT
ccoltekin@sfs.uni-tuebingen.de (Ç. Çöltekin); matteo.brivio@student.uni-tuebingen.de (M. Brivio); fidan.can@student.uni-tuebingen.de (F. Can)
https://coltekin.net/cagri/ (Ç. Çöltekin)
0000-0003-1031-6327 (Ç. Çöltekin); 0000-0001-6273-6900 (M. Brivio); 0009-0006-8900-2038 (F. Can)

2. Method and Experiments

2.1. Data

The PoliticIT dataset comprises between 80 and 100 tweets from 1751 politicians (1298 for training and development, 453 for testing). As usual, the test set was released only at the end of the competition. All tweet instances are anonymized by masking references to politicians, political parties and other Twitter account mentions. Further information on data collection and labeling can be found in Russo et al. [13].

Besides the PoliticIT dataset, we experiment with a few additional resources: transcripts of the parliamentary speeches from the Italian section of the parliamentary corpora collection ParlaMint [14] and the PoliticES 2023 shared task dataset [15, 16].

Parliamentary speeches We use the Italian section of the ParlaMint 3.0 pre-release, which contains speeches both from the Senate and the Chamber of Deputies, spanning from March 2013 to September 2022. We filter out samples that belong to the chairperson and those that are less than 50 characters long. We also exclude speakers without a known party affiliation, as well as those who are affiliated with parties whose political orientation is either not specified or reported to be ‘center’ or ‘big tent’.[1] Compared to the dataset released for the present shared task, the political orientation in the ParlaMint data is more fine-grained. For binary class labels, we map ParlaMint orientation labels to `left` if the class includes `L` (left) and to `right` if it includes `R` (right). We follow the same approach for the multi-class classification task, but mark orientation labels as `moderate` if they include `C` (center). For example, samples with original labels `CL` (center-left) or `CCL` (center to center-left) are mapped to `moderate_left`, while samples marked as `R` or `FR` (far-right) are mapped to `right`. The resulting corpus includes 167 `moderate_left`, 153 `moderate_right`, 89 `right` instances, and no instances with (non-center) `left` labels.

Compared to social media posts, parliamentary speeches tend to be much longer, and a few speakers have many more speeches than others in the corpus. To get a balanced dataset, we randomly select at most 10 speeches from each speaker and concatenate them with a separator token. The resulting corpus contains samples with an average of 2575.95 space-separated tokens per instance.

PoliticES 2023 shared task data The dataset consists of 2797 instances, each containing 80 tweets written by different users who share the same gender and political views. All tweets are anonymized, mirroring the anonymization procedure for the PoliticIT dataset. We use the PoliticES 2023 shared task data to conduct cross-lingual transfer learning experiments between Italian and Spanish. Further information about the data collection and labeling can be obtained from García-Díaz et al. [16].

2.2. Linear models

As a first, simple approach, we use linear support vector machines (SVMs) with bag-of-n-grams features. For all the results reported here, we train a separate linear SVM classifier for each task, with both word and character n-grams as features and weigh each feature using tf-idf. For each task, we tune the SVM margin parameter, `C`, (range 0.001-5.0), the maximum word n-grams (range 0-4) and the character n-grams (range 0-6). Lastly, we test whether n-grams should be case normalized or not (word, character, both or none). We draw 10 000 configurations uniformly from this hyper-parameter space and pick the setting with the highest average F1-score using 10-fold cross validation. The input to the linear models are all tweets belonging to each instance combined with an arbitrary separator symbol inserted between each one of them. To optimize the models trained on the ParlaMint data, we use the complete PoliticIT training set as the development set. Besides the predictions from the best model after the hyper-parameter search, we also experiment with the majority vote of the top-n best performing hyper-parameters. As the voting ensembles performed worse than the single-best model, we do not report their results here.

Since the binary ideology classifier was more accurate than the multi-class one in our initial experiments, we also include a post-processing step, where we adjust multi-class labels when they conflict with binary labels. If the multi-class classifier predicts (`moderate_`)`left` while the binary classifier predicts `right`, we set the multi-class label to `moderate_right`. Similarly, we resolve conflicts where the binary classifier predicts `left` by setting the multi-class label to `moderate_left`. All linear models are implemented using scikit-learn [17].

2.3. Transformer-based models

We also experiment with multi-task and cross-lingual transfer learning by fine-tuning pretrained language models. Specifically, we work both with a multilingual model, XLM-RoBERTa [18], and a monolingual version of BERT [19].[2] We customize both models, adding a ‘common block’ followed by three linear classification heads. Both solutions are implemented with PyTorch [20] and the Transformers library [21].

The ‘common block’ is a 2-layer, bi-directional Recurrent Neural Network which takes as input the language model’s representation of the `CLS` token for each tweet in a given group. We include a ReLU activation function and a dropout layer after each RNN layer, except the last one. The model leverages the hard-shared parameters in the RNN and tries to learn the three tasks simultaneously (i.e. in a multi-task fashion) by minimizing the sum of their training losses: binary cross entropy loss for the two binary tasks and categorical cross entropy loss for the multi-class one. As observed by Liang and Zhang, it is reasonable to assume that in such a setting, the tasks have different levels of difficulty, so we rely on their *balanced multi-task learning* (BMTL) framework to mitigate this issue [22]. This requires wrapping each loss ℓ into a function h, such that $h(\ell) = \exp(\frac{\ell}{T})$, where T is a temperature parameter that needs to be tuned.

We train the XLM-R model in three different settings: first on the combination of the full PoliticES 2023 dataset and the PoliticIT 2023 training set, then on the Italian dataset alone and lastly only on the Spanish data. Fi-

[1]Political parties encouraging a broad spectrum of views among their members.

[2]https://huggingface.co/dbmdz/bert-base-italian-cased.

nally, the BERT-based model is also trained on the Italian dataset alone.

In an attempt to improve the multi-class ideology task score, we include a post-processing step similar to the one described in 2.2. If the binary score is `left` and the multi-class prediction is `moderate_right`, we set the multi-class label to `moderate_left` and the other way around if the two scores are `right` and `moderate_left`. Similarly, if the binary score is `left` and the multi-class prediction is `right`, we set the latter to `left` and the other way around if the two scores are `right` and `left`.

All models use the AdamW optimizer and a scheduler with a learning rate that decreases following the values of the cosine function. As a pre-processing step, we remove all punctuation marks and masking tokens originally used to anonymize the tweets (e.g., `[POLITICIAN]`, `[POLITICAL_PARTY]`), expand all hashtags and convert emojis into their text descriptions.

All models are trained for 15 epochs and their hyperparameters are optimized with Ray Tune [23], relying on a Bayesian optimization strategy.

3. Results

Our best performing system (based on linear SVMs) ranked first among 7 participating teams on both the binary and the multi-class ideology prediction tasks, achieving a macro-averaged F1-score of 92.82 and 75.15, respectively. With respect to the gender prediction task, our system ranked second with a score of 79.25.

Besides these simple models, we experimented with parliamentary speeches as an out-of-domain dataset and cross-lingual transfer, using pretrained language models. We summarize our results in Table 1.

Out-of-domain data To investigate the effectiveness of training on out-of-domain data, we train an SVM model on parliamentary speeches from the ParlaMint dataset and test it on the PoliticIT test-set. We present the results in Table 1 (indicated as SVM-ParlaMint).

ParlaMint includes both political orientation and gender data for the speakers in the Italian parliament. However, as noted in Section 2.1, it does not include speakers belonging to any (non-center) left parties for the multi-class classification. As a result, the scores for multi-class ideology detection are rather poor. Nonetheless, both the binary ideology and gender scores are not too far behind the models trained on the in-domain data.

Cross-lingual transfer The results obtained with monolingual and cross-lingual pretrained models, Italian BERT and XLM-RoBERTa, are also presented in Table 1. Contrary to our expectations, the BERT-based model performs rather poorly compared to SVM-PoliticIT, with the sole exception of the gender classification task, where the monolingual model achieves a macro-averaged F1-score of 81.17, outperforming the SVM model.

Even though it does not completely exceed the performance of the SVM-PoliticIT classifier, the multilingual model trained on both the Italian and the Spanish data improves the scores substantially. Specifically, the model achieves a macro-averaged F1-score of 80.56 in the gender task, outperforming SVMs. The effectiveness of cross-lingual transfer learning can be observed by examining the results of the monolingual BERT model which, with the only exception of gender, are significantly lower. However, it should be noted that XLM-RoBERTa and BERT are pretrained using different training regimes. To get a clearer understanding of the potential benefits of cross-lingual transfer, we trained XLM-RoBERTa on the Italian dataset alone. The scores we report for the latter model are marginally lower than those obtained with the model trained on both the Italian and the Spanish data, suggesting a beneficial effect of cross-lingual transfer learning. This is further corroborated by the results in a zero-shot setting, where the XLM-RoBERTa model was trained on Spanish data only. Although the scores are well below those of the other models, they are clearly better than random, indicating signal in the cross-lingual data.

4. Discussion and Concluding Remarks

We presented our experiments on how to identify political orientation and gender from social media posts as part of the EVALITA 2023 PoliticIT shared task. Our systems based on linear SVMs achieved the best overall scores. We also experimented with classifiers based on deep pretrained language models, focusing particularly on cross-lingual transfer.

Our first finding is that ‘traditional’ linear models show a better overall performance than the deep learning ones. Given their simplicity and lack of information other than immediate surface features in the training data, their success may come as a surprise. However, this is not the first time that such models have been found to perform similarly or better than deep networks [24, 25, 26, 27, just to name a few]. Similarly, the logistic regression classifiers of Mosquera [28] obtained the third place, only slightly behind the first two systems, in the PoliticES 2022 shared task. This, of course, may be due to a lack of proper tuning of the pretrained models. However, the fact that the linear SVMs outperformed all other participating teams (presumably the majority of them using pretrained models) shows that these simpler models have some strong merits in this task.

Our preliminary experiments on out-of-domain data

Model	Task	Precision	Recall	F1-score
SVM-PoliticIT	ideology binary	93.25	92.58	92.82
	ideology multi	80.12	73.41	75.15
	gender	81.72	77.71	79.25
SVM-ParlaMint	ideology binary	79.49	78.55	78.79
	ideology multi	44.89	56.55	49.49
	gender	73.39	76.73	74.17
BERT-based model	ideology binary	84.62	78.06	78.34
	ideology multi	73.72	63.09	64.12
	gender	80.92	81.43	81.17
XLM-R-based model (it)	ideology binary	89.75	86.55	87.14
	ideology multi	74.73	68.84	69.76
	gender	75.43	79.26	76.26
XLM-R-based model (es)	ideology binary	60.16	55.90	52.30
	ideology multi	37.42	23.53	17.40
	gender	58.12	53.12	33.04
XLM-R-based model (es-it)	ideology binary	90.16	90.28	90.21
	ideology multi	75.28	68.44	70.08
	gender	79.47	82.62	80.56

Table 1
Precision, recall and F1-scores of all models. All scores are macro-averaged scores on the official PoliticIT shared task test set.

show that training an SVM model on parliamentary speeches results in lower, but comparable performance to training on in-domain-data. The results clearly indicate that there is a considerable cross-domain signal for both ideology detection and gender prediction. Although we did not experiment with the use of both in- and out-of-domain data together in this study, the results highlight the potential for using out-of-domain data to improve ideology prediction in social media.

Our experiments also show promising results on cross-lingual transfer. The best results we achieve with pre-trained language models come from fine-tuning a cross-lingual model using both Spanish and Italian data. Part of this success is probably due to the similar methodologies followed by the organizers of both shared tasks. Further investigation of cross-lingual and cross-country transfer on ideology detection may shed light into universal and culture-specific aspects of 'ideology'.

Our best-performing model is a bag-of-words SVM classifier, relying only on token unigrams as features. Compared to the transformer-based ones, this model allows for better interpretability[3] with its unigram features revealing some linguistic indications of political orientation. Words such as *immigrazione* 'immigration', *clandestini* 'illegal immigrants' and *confini* 'borders' appear constantly in right-wing discourse, while terms such as *diritti* 'rights', *democrazia* 'democracy' and *solidarietà* 'solidarity' seem to point to more leftist views. The model also picked up unmasked tokens left in the data that could suggest a party or politician's name. This is the case with words like *Giorgia*, *FI* and *Democratico*, which allude to the current right-wing Prime Minister Giorgia Meloni, a well-known Italian right-wing party *Forza Italia*, and a prominent Italian left-wing party *Partito Democratico*, respectively. Further and better approaches to explain model behavior is another direction for future research that may shed light in narratives of different ideological groups.

[3] Although one should be cautious because tokens are not independent features, the weights assigned to individual tokens are, nevertheless, easier to interpret than weights of a neural network or weights in a model with overlapping features.

References

[1] J. T. Jost, C. M. Federico, J. L. Napier, Political ideology: Its structure, functions, and elective affinities, Annual review of psychology 60 (2009) 307–337.

[2] R. S. Erikson, K. L. Tedin, American public opinion: Its origins, content and impact, Routledge, 2015.

[3] C. Stachl, F. Pargent, S. Hilbert, G. M. Harari, R. Schoedel, S. Vaid, S. D. Gosling, M. Bühner, Personality research and assessment in the era of machine learning, European Journal of Personality 34 (2020) 613–631.

[4] H. Christian, D. Suhartono, A. Chowanda, K. Z. Zamli, Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging, Journal of Big Data 8 (2021) 1–20.

[5] F. Rangel, P. Rosso, M. Montes-y Gómez, M. Potthast, B. Stein, Overview of the 6th author profiling task at PAN 2018: multimodal gender identification

in Twitter, Working notes papers of the CLEF (2018) 1–38.

[6] M. Fatke, Personality traits and political ideology: A first global assessment, Political Psychology 38 (2017) 881–899. URL: http://www.jstor.org/stable/45095184.

[7] B. Verhulst, L. J. Eaves, P. K. Hatemi, Correlation not causation: The relationship between personality traits and political ideologies, American journal of political science 56 (2012) 34–51.

[8] J. A. García-Díaz, R. Colomo-Palacios, R. Valencia-García, Psychographic traits identification based on political ideology: An author analysis study on Spanish politicians' tweets posted in 2020, Future Generation Computer Systems 130 (2022) 59–74. URL: https://www.sciencedirect.com/science/article/pii/S0167739X21004921. doi:https://doi.org/10.1016/j.future.2021.12.011.

[9] G. Lakoff, Moral politics: How liberals and conservatives think, University of Chicago Press, 2010.

[10] M. Iyyer, P. Enns, J. Boyd-Graber, P. Resnik, Political ideology detection using recursive neural networks, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Baltimore, Maryland, 2014, pp. 1113–1122. URL: https://aclanthology.org/P14-1105. doi:10.3115/v1/P14-1105.

[11] W. Safire, Safire's political dictionary, Oxford University Press, 2008.

[12] M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi, Evalita 2023: Overview of the 8th evaluation campaign of natural language processing and speech tools for Italian, in: Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR.org <http://CEUR.org>, Parma, Italy, 2023.

[13] M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi, Evalita 2023: Overview of the 8th evaluation campaign of natural language processing and speech tools for italian (2023).

[14] T. Erjavec, M. Ogrodniczuk, P. Osenova, N. Ljubešić, K. Simov, A. Pančur, M. Rudolf, M. Kopp, S. Barkarson, S. Steingrímsson, Ç. Çöltekin, J. de Does, K. Depuydt, T. Agnoloni, G. Venturi, M. Calzada Pérez, L. D. de Macedo, C. Navarretta, G. Luxardo, M. Coole, P. Rayson, V. Morkevičius, T. Krilavičius, R. Darģis, O. Ring, R. van Heusden, M. Marx, D. Fišer, The ParlaMint corpora of parliamentary proceedings, Language resources and evaluation 57 (2022) 415–448. doi:10.1007/s10579-021-09574-0.

[15] J. A. García-Díaz, S. M. Jiménez Zafra, M. T. Martín Valdivia, F. García-Sánchez, L. A. Ureña López, R. Valencia García, Overview of PoliticEs 2022: Spanish author profiling for political ideology, Procesamiento del Lenguaje Natural 69 (2022) 265–272. doi:10.26342/2022-69-23.

[16] J. A. García-Díaz, S. M. Jiménez Zafra, M. T. Martín Valdivia, F. García-Sánchez, L. A. Ureña López, R. Valencia García, Overview of PoliticES2023@IberLEF: Political ideology detection in Spanish texts, Procesamiento del Lenguaje Natural 71 (2023).

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[18] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[19] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[21] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. M. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https:

//www.aclweb.org/anthology/2020.emnlp-demos.6.

[22] S. Liang, Y. Zhang, A simple general approach to balance task difficulty in multi-task learning, ArXiv abs/2002.04792 (2020).

[23] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, I. Stoica, Tune: A research platform for distributed model selection and training, arXiv preprint arXiv:1807.05118 (2018).

[24] M. Medvedeva, M. Kroon, B. Plank, When sparse traditional models outperform dense neural networks: the curious case of discriminating between similar languages, in: Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial), Association for Computational Linguistics, Valencia, Spain, 2017, pp. 156–163. URL: https://aclanthology.org/W17-1219. doi:10.18653/v1/W17-1219.

[25] Ç. Çöltekin, T. Rama, Tübingen-Oslo at SemEval-2018 task 2: SVMs perform better than RNNs in emoji prediction, in: Proceedings of the 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 34–38. URL: https://aclanthology.org/S18-1004. doi:10.18653/v1/S18-1004.

[26] A. Caines, P. Buttery, REPROLANG 2020: Automatic proficiency scoring of Czech, English, German, Italian, and Spanish learner essays, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 5614–5623. URL: https://aclanthology.org/2020.lrec-1.689.

[27] K. Amponsah-Kaakyire, D. Pylypenko, J. Genabith, C. España-Bonet, Explaining translationese: why are neural classifiers better and what do they learn?, in: Proceedings of the Fifth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 2022, pp. 281–296. URL: https://aclanthology.org/2022.blackboxnlp-1.23.

[28] A. Mosquera, Alejandro mosquera at PoliticEs 2022: Towards robust Spanish author profiling and lessons learned from adversarial attacks, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022). CEUR Workshop Proceedings, CEUR-WS, A Coruna, Spain. D. Moctezuma, and V. Muniz-Sánchez, 2022.