Geopolitically-Informed MultiModal BERT for Propaganda Detection in Political Tweets

Antoni Valls[1,2,*,†], Björn Komander[1,3,*,†] and Jesús Cerquides[1]

[1] IIIA-CSIC, Campus UAB, 08193 Cerdanyola, Spain
[2] University of Padova, Italy
[3] School of Computing Technologies, RMIT University

Abstract

People are becoming increasingly dependent on social media for their political information, with political accounts often being their primary source. The overwhelming flood of information makes it challenging for the average person to compare sources, critically evaluate the content, and form well-informed opinions. This context leave individuals susceptible to manipulation by political actors who exploit these vulnerabilities. Following this line, we presented a model for the DIPROMATS Task 1 of IberLEF 2024, which challenges participants to detect propaganda from diplomatic tweets in English and Spanish. Our approach proposed a MultiModal BERT-like model that combined a pre-trained TwHIN-BERT encoder with a selection of contextual features aimed at helping distinguishing the underlying geopolitical intentions behind propaganda use and at mitigating typical issues in text-stream scenarios, as spurious correlations or concept drift. Although our results do not reflect a perfect handling of this issues, we achieved competitive results that positioned us in the middle of the rankings, with better performance for the Spanish track. Adding contextual information to TwHIN-BERT improves the F1 score for the propaganda class, motivating further research in this domain.

Keywords

Automated Propaganda Detection, Text-Based Propaganda Detection, MultiModal BERT, Concept Drift

1. Introduction

Social media has become a crucial platform for delivering geopolitical speeches. The hyperconnectivity that social networks have brought to our lives translates into a direct political-audience communication line, either via tweets, Facebook posts, or YouTube videos. Politicians are now able to react almost instantly to events occurring in any part of the world, and their rhetoric may become an essential source of the political information their followers receive. In this sense, readers are vulnerable to any manipulative intentions of the political voices.

Furthermore, with the rise of social media, individuals now have the ability to connect with significantly broader audiences, a privilege once reserved for major news organizations [1]. The downside of this democratization of the speech is the overabundance of information that one receives, making it a very difficult and arduous task to critically compare and evaluate information between sources, and to build personal well-informed opinions, leaving us more exposed to misinformation.

In 1938, the Institute for Propaganda Analysis defined the term *propaganda* as the “expression of opinion or action by individuals or groups deliberately designed to influence opinions or actions of other individuals or groups with reference to predetermined ends”. Although the research on the detection of propaganda is not as popular as the detection of fake news or fact-checking, there has been some work done in recent years, focused primarily in articles [2, 3, 4], but also in tweets [5, 6, 7]. Recognizing the importance of this challenge, DIPROMATS (Automatic Detection and Characterization of Propaganda Techniques and Narratives from Diplomats of Major Powers) seeks to promote the research in the field of propaganda detection.

IberLEF 2024, September 2024, Valladolid, Spain

[*] Corresponding author.

[†] These authors contributed equally.

antoni.vallscifre@studenti.unipd.it (A. Valls); bkomander@iiia.csic.es (B. Komander)
0000-0002-8732-1258 (B. Komander); 0000-0002-3752-644X (J. Cerquides)

The DIPROMATS shared Task 1 at IberLEF 2024 [8, 9] encourages participants to classify propagandistic tweets from governmental and diplomatic sources according to three distinct subtasks:

- Binary classification problem: models should decide if tweets contain propagandistic techniques.
- Coarse-grained propaganda characterization: models should relate each tweet to one of the four available categories (*Not propagandistic, Appeal to commonality, Discrediting the opponent, Loaded language*).
- Fine-grained propaganda characterization: models should relate each tweet to one of the eight available categories (*Not propagandistic, Flag Waving, Ad Populum / Ad antiquitatem, Name Calling/Labelling, Undiplomatic Assertiveness / Whataboutism, Appeal to Fear, Doubt, and Loaded Language*).

Our work focuses on the first subtask in both English and Spanish. The detection of propaganda in essence is a text classification task with the aim to develop the best possible classifier over a set of text instances X and propaganda labels y. Specifically, the aim is to learn a function $f(X_{train}, y_{train})$ that minimizes a loss $\mathcal{L}(\hat{y}, y)$ and can predict well on unseen data. With recent advances in Natural Language processing, Large Language Models offer a promising tool for the detection of propaganda. However, we slightly deviate from the aim of developing the best possible text classifier, for the following reasons: Online propaganda is a complex concept, that does not only depend on the used language. Following Neriono [10], online propaganda can be characterized into the supply side, including the producers and production of propaganda; the demand side, including the potential receivers of propaganda and how sociocultural factors might impact both the demand-side and the supply-side. Furthermore, the use of online propaganda is driven by political interests and diplomatic strategies [11], which can influence the language used based on these factors.

Therefore it is important to consider contextual information for any text classification to better capture the concept of propaganda. Otherwise, the text classification model might overfit on spurious correlations, or suffer from concept drift [12], as propaganda is an evolving set of techniques and mechanisms [4, 3].

In this work we propose a MultiModal Bert-like model, that simultaneously learns on the input text and a rich array of contextual data, characterizing the supply side, demand side and the potential sociopolitical context. By incorporating relevant contextual features, our approach seeks to capture the underlying dynamics that drive propaganda campaigns, enabling the model to adapt to evolving propaganda strategies and political narratives.

While our approach does not outperform other models in this challenge, we find small improvements in the F1 for the propaganda class. However, future work is needed to determine a better set of contextual features, and if additional data allows a classification model to capture a more complete picture of online propaganda.

In the remainder of this work we briefly review related work, introduce the dataset and discuss our choice of contextual features. We then examine our model architecture and the performance of our model in the Dipromats challenge.

2. Related Work

An influential early work by Da San Martino et al. [3] classified text segments based on the specific propaganda technique employed, initially considering 18 different techniques. Subsequent studies by the same authors consolidated the taxonomy, reducing the number of distinct techniques to 14 and further grouping them into 6 higher-level clusters [1, 4]. Although, this line of research provided a framework for systematically analyzing and identifying propaganda in news articles [2], the same foundation has also influenced short-text propaganda characterization research, e.g. both DIPROMATS 2023 [7] and DIPROMATS 2024.

Various methodologies were applied in DIPROMATS 2023, most of them BERT-based [13]. The PropaLTL team [14] added contextual information to the tweet (sentiment, tweet type and country of the author) to the BERTweet [15] and RoBERTuito [16] models, obtaining the best overall performance in the binary task: top position for Spanish out of 18 runs, and the second position for English out of 30 runs. Alternatively to BERT-based models, the English winning team of task 1 was team Mario [17], who designed a system of cascades using two GPT-J models. Other methodologies can be found in the DIPROMATS 2023 report [7].

As most of the high-performance models are BERT-based models we decided to adopt the TwHIN-BERT [18] as our backbone model for all the experiments.

3. Data

The DIPROMATS labeled dataset consists of 12012 English tweets and 9591 Spanish tweets published between January 1st, 2020 and March 11th, 2021 by authorities and diplomats from China, Russia, US and the EU. The data is split with a temporal criterion, following a 70/30 train-test proportion. Table 1 shows the number of tweets and propaganda proportion for each actor in both languages.

Table 1
DIPROMATS dataset characteristics by actor

Actor	# English tweets	# Spanish tweets	Propaganda freq. in English	Propaganda freq. in Spanish
China	3022	2997	0.31	0.26
Russia	2960	1391	0.20	0.25
US	3114	2738	0.32	0.22
EU	2916	2465	0.09	0.05

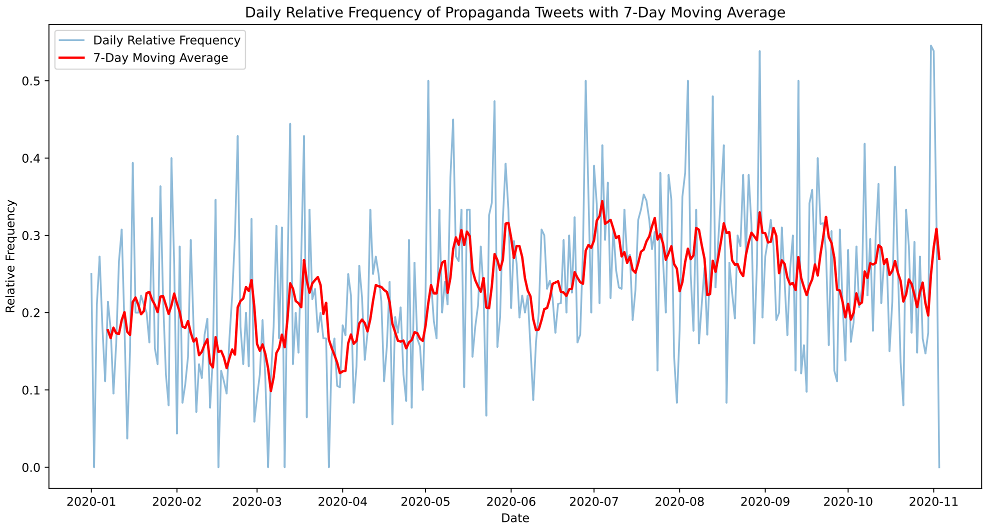
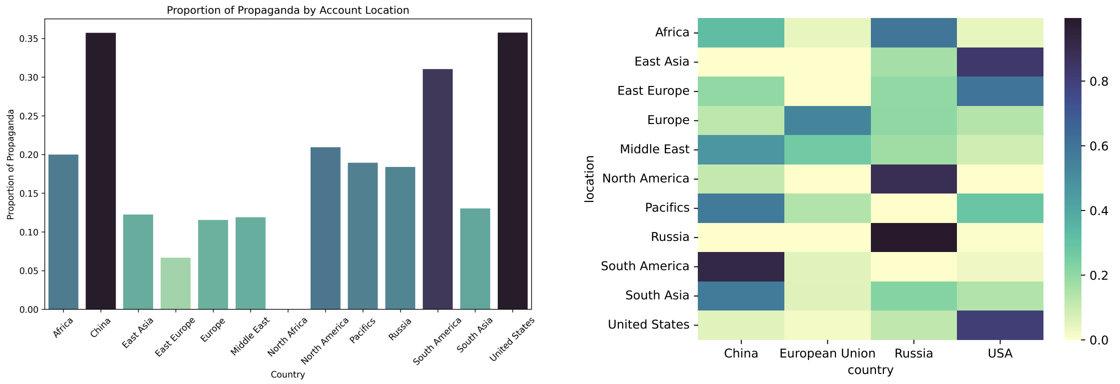


Figure 1: Evolution of propaganda frequency on the English training set.

Figure 1 shows the relative use of propaganda over time. Our primary hypothesis is that the presence of propaganda in political speeches reflects the geopolitical strategies of the author's country. The topic modeling analysis detailed in Appendix A supports this hypothesis. This analysis strengthens the argument that the BERT model should incorporate information about the state of the world. We implemented the following features to our experiments:

- **Country**: The country of origin of the diplomat who shared the tweet.
- **COVID-19 Metrics**: The 7-day moving average of new COVID-19 related deaths, segmented by China, Russia, the United States, Europe, and globally.

- **Armed Conflict Features**: The 7-day moving average of occurring armed conflicts, categorized by regions including South America, North America, the Middle East, Europe, and East Asia.



(a) Proportion of propaganda based on the account's registered locations.

(b) Proportion of Region's Propaganda by accounts country.

Figure 2: Characteristics of the location of the accounts for the English DIPROMATS dataset.

Furthermore, we extracted the geographic location for each account using the Twitter API, supplemented by manual mapping efforts. Figure 2a illustrates the proportion of propaganda based on the geographical regions or countries where the accounts are located. Notably, diplomatic discourse originating from accounts situated in China, the United States, and South America shows a higher tendency towards propagandistic content compared to those based in Europe. Figure 2b shows the main registered locations of the diplomats of each actor country, where the data reflects well the geopolitical interests of the countries. For instance, most propaganda in South America is produced by China, while China and Russia produce the most propaganda in Africa, and the US produces the most propaganda in East-Asia and Eastern Europe.

4. Methodology

Figure 3 shows the pipeline of our approach. This can be divided in three parts:

- **Text encoding.** In the pre-processing stage, we removed URLs from the tweets, retaining the text, hashtags, mentions and emojis. The remaining content was then encoded using the TwHIN-BERT-base model, a multilingual language model pre-trained on 7 billion tweets from over 100 distinct languages. This model is specifically designed to handle social media engagements on Twitter [18]. Figure 4 depicts the TwHIN-BERT diagram followed by the concatenation process.
- **Context data generation.** Based on the notion that contextual features reflecting the global context can help uncover the manipulative intentions of diplomats and politicians using propaganda, we performed topic modeling to discover the main concerns of the tweet authors. The topic modeling analysis, conducted using BERTopic [19], is detailed in Figure 6 from Appendix A. It depicts the temporal evolution of main topics of the English tweets and their associated propaganda frequency (Spanish tweets had similar topics). Unsurprisingly, COVID-19 emerged as one of the most prominent subjects. Consequently, we incorporated information about the COVID-19 pandemic and various global armed conflicts. Incorporating information about armed conflicts can be crucial because such events can be exploited for propaganda purposes, with diplomats and politicians attempting to shape narratives and influence public opinion. For COVID-19, we included five features: a 7-day smoothed count of new deaths attributed to COVID-19 in China, Russia, the US, Europe, and globally [20]. For armed conflicts, we used data from ACLED (Armed Conflict Location & Events Data) [21]. From their database, we counted the number of "Violence

against civilians" [1] and "Battles" [2] in South America, North America, Middle East, Europe, and East Asia for each day within the time range of the DIPROMATS dataset.

- **Concatenation and Classification Head.** The contextual embedding of the text was concatenated to a one-hot encoding of the country that the author represents and to the rest of the context data described in the previous step. The resulting vector was fed into a MLP that worked as a binary classifier. This MLP is composed of one hidden layer of 78 nodes with a ReLU activation function and dropout.

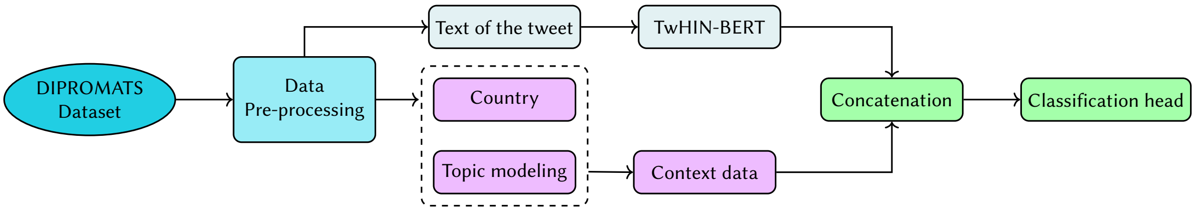


Figure 3: Representation of the pipeline of our model.

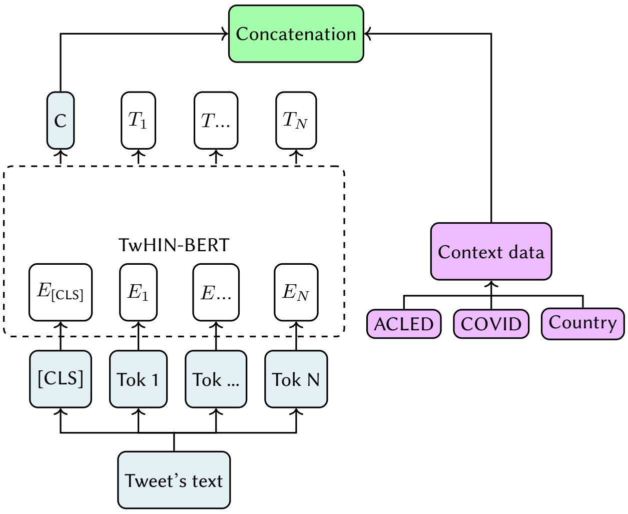


Figure 4: TwHIN-BERT diagram and concatenation step.

For participating in the shared task, we presented four different methodologies:

- **Method 1:** Train the model on the Spanish and English data together using context data.
- **Method 2:** Train one model per each language, separately and using context data.
- **Method 3:** Train the model on the Spanish and English data together without context data.
- **Method 4:** Train one model per each language, separately without using context data.

In all of them, the fine tuning was made freezing all layers of the pre-trained TwHIN-BERT model, except for the last two, and using cross-validation by splitting the training set into 5 equal length time periods. In order to offset to unbalanced propaganda presence, see Table 1, we used the Weighted Cross-Entropy loss function, with class weight equal to class frequencies in the training data. We used the following set of hyperparameters:

[1] ACLED defines this as *violent events where an organized armed group inflicts violence upon unarmed non-combatants.*
[2] ACLED defines this as *a violent interaction between two organized armed groups at a particular time and location.*
[3] With a reducing learning rate on plateau scheduler.

Hyperparameter	Value
Optimizer	AdamW
Learning rate	1e-5 [3]
Epochs	15-18
Batch Size	64
Max_len	128

Table 2
Hyperparameter configuration used.

5. Discussion

Table 3 presents the official results of our submissions in the competition. Our models achieved a moderate ranking overall, with notably better performance on the Spanish language task.

Method 3 emerged as our top-performing model across all language tracks, closely followed by Method 1. The ranking is sorted by the ICM metric [22]. Focusing on the F1-True score, which measures the ability to accurately identify true instances of propaganda, Method 1 outperformed the other methods for all languages. This could suggest that the external features incorporated in Method 1 provided additional discriminating power in detecting propaganda instances.

However, it's worth noting that the results are not very impressive, and the performance gap between Method 3 and Method 1 is relatively small. This implies that the classification performance is not heavily reliant on the external features compared to the information extracted from the embedded tweets themselves.

Another notable observation is the superior performance achieved by training the model on both Spanish and English data simultaneously. This, reasonably explained by the fact that larger training datasets typically lead to better results, also validates TwHIN-BERTS's true multilingual capabilities, as it could effectively leverage and transfer knowledge from both language corpora during the training process.

English					
	Rank	ICM	F1 - Macro	F1 - True	F1 - False
Method 3	17 of 33	0,0880	0,7568	0,5865	0,9270
Method 1	18 of 33	0,0831	0,7626	0,6248	0,9005
Method 2	19 of 33	0,0548	0,7523	0,6063	0,8982
Method 4	20 of 33	0,0299	0,7281	0,5356	0,9206
Spanish					
	Rank	ICM	F1 - Macro	F1 - True	F1 - False
Method 3	9 of 33	0,1394	0,7932	0,6383	0,9481
Method 1	10 of 33	0,1344	0,7928	0,6531	0,9326
Method 4	18 of 33	0,1058	0,7757	0,6068	0,9446
Method 2	20 of 33	0,1029	0,7801	0,6330	0,9271
Bilingual					
	Rank	ICM	F1 - Macro	F1 - True	F1 - False
Method 3	16 of 33	0,1126	0,7736	0,6096	0,9375
Method 1	17 of 33	0,1083	0,7769	0,6369	0,9167
Method 2	19 of 33	0,0783	0,7654	0,6180	0,9128
Method 4	20 of 33	0,0661	0,7500	0,5675	0,9325

Table 3
English, Spanish and Bilingual performance extracted from the official DIPROMATS results.

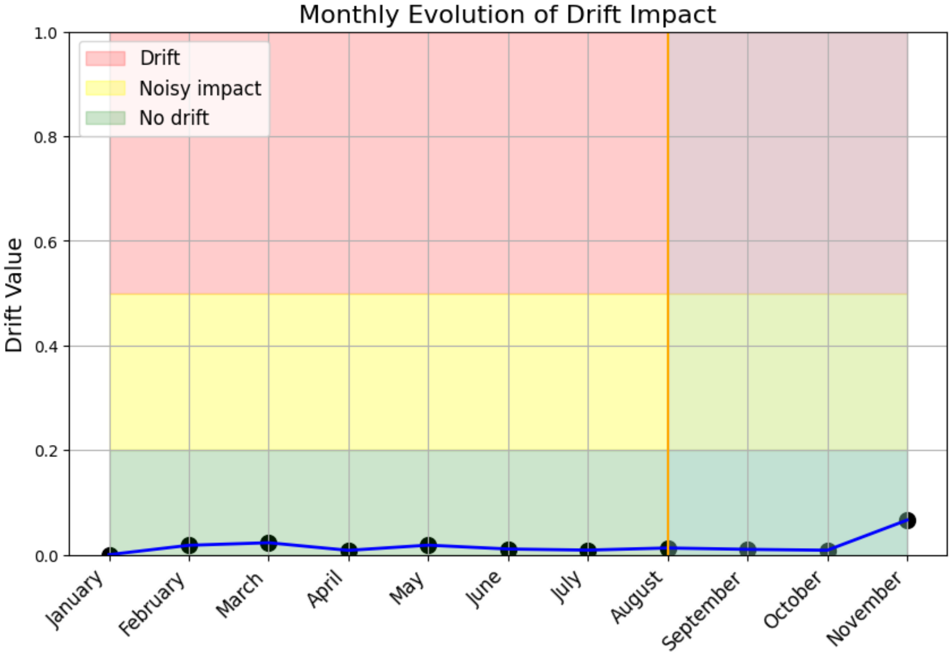


Figure 5: Monthly drift evolution for the English training set. The fine-tuned TwHIN-BERT encoder has been trained on data from January until August. The different impact thresholds are based on [23] work.

Since contextual data has not been able to notably improve the classification capabilities of our model, we have conducted an study about the concept drift present in the dataset. Following the work of [23], we attempted to detect drift by tracking the evolution of the propaganda and non-propaganda cluster centers. We used a fine-tuned TwHIN-BERT model trained on data from January until August to generate sentence embeddings for each tweet in the entire English training set, and then computed the centers of the propaganda and non-propaganda embeddings. Then, we measured the distance between the monthly cluster centers using cosine semantic similarity distance, as [23]. Figure 5 shows the evolution of the drifts between consequent months. This analysis revealed that the TwHIN-BERT embeddings did not exhibit any significant drift over time, which could be an explanation, why the addition of context features did not improve the model's performance.

6. Conclusion

We developed a propaganda detection model for the DIPROMATS shared challenge that achieved a good performance across languages, with particularly strong results for Spanish data. Our methodology effectively leveraged the multilingual capabilities of the pre-trained TwHIN-BERT encoder, placing us in the mid-tier of the rankings.

However, our primary goal of investigating concept drift as a method to uncover the veiled political intentions behind the use of propaganda proved less successful. The model performed similarly with and without contextual features, likely due to the lack of concept drift reflected from the tweet embeddings over time. Also more generally, the inclusion of additional features did not offer large performance improvements.

Despite this, we think that approaching a more holistic view of propaganda in propaganda detection by including contextual information about the producers of propaganda, the potential audience and the sociopolitical situation, could improve the performance of models and make them more resistant against spurious correlations and especially concept drift. Further research is needed on the additional data that could provide useful for the detection of propaganda in text and in which settings this could make a text based classifier more resilient against spurious correlations and topic drift, and in which setting the additional data is not useful. Incorporating different contextual features could help uncover the subtle intentions behind political social media speech, and bring us closer to the possibility of a real-time automated propaganda detector.

Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 101034328.

References

[1] G. D. S. Martino, S. Cresci, A. Barron-Cedeno, S. Yu, R. Di Pietro, P. Nakov, A Survey on Computational Propaganda Detection, 2020. URL: http://arxiv.org/abs/2007.08024, arXiv:2007.08024 [cs].

[2] A. Barrón-Cedeño, G. D. S. Martino, I. Jaradat, P. Nakov, Proppy: A System to Unmask Propaganda in Online News, 2019. URL: http://arxiv.org/abs/1912.06810, arXiv:1912.06810 [cs].

[3] G. Da San Martino, S. Yu, A. Barrón-Cedeño, R. Petrov, P. Nakov, Fine-Grained Analysis of Propaganda in News Article, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 5635–5645. URL: https://www.aclweb.org/anthology/D19-1565. doi:10.18653/v1/D19-1565.

[4] G. Da San Martino, S. Shaar, Y. Zhang, S. Yu, A. Barrón-Cedeño, P. Nakov, Prta: A System to Support the Analysis of Propaganda Techniques in the News, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 287–293. URL: https://www.aclweb.org/anthology/2020.acl-demos.32. doi:10.18653/v1/2020.acl-demos.32.

[5] P. Vijayaraghavan, S. Vosoughi, TWEETSPIN: Fine-grained Propaganda Detection in Social Media Using Multi-View Representations, in: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Seattle, United States, 2022, pp. 3433–3448. URL: https://aclanthology.org/2022.naacl-main.251. doi:10.18653/v1/2022.naacl-main.251.

[6] K. Hristakieva, S. Cresci, G. D. S. Martino, M. Conti, P. Nakov, The Spread of Propaganda by Coordinated Communities on Social Media, in: 14th ACM Web Science Conference 2022, 2022, pp. 191–201. URL: http://arxiv.org/abs/2109.13046. doi:10.1145/3501247.3531543, arXiv:2109.13046 [cs].

[7] P. Moral, G. Marco, J. Gonzalo, J. Carrillo-de Albornoz, I. Gonzalo-Verdugo, Overview of DIPROMATS 2023: automatic detection and characterization of propaganda techniques in messages from diplomats and authorities of world powers (2023).

[8] P. Moral, J. Fraile, G. Marco, A. Peñas, J. Gonzalo, Overview of DIPROMATS 2024: Detection, characterization and tracking of Propaganda in messages from diplomats and authorities of world powers, Procesamiento del Lenguaje Natural 73 (2024).

[9] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024.

[10] V. Nerino, Overcome the fragmentation in online propaganda literature: the role of cultural and cognitive sociology, Frontiers in Sociology 8 (2023) 1170447. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10366602/. doi:10.3389/fsoc.2023.1170447.

[11] C. Sparkes-Vian, Digital Propaganda: The Tyranny of Ignorance, Critical Sociology 45 (2019) 393–409. URL: https://doi.org/10.1177/0896920517754241. doi:10.1177/0896920517754241, publisher: SAGE Publications Ltd.

[12] C. M. Garcia, R. S. Abilio, A. L. Koerich, A. d. S. Britto Jr., J. P. Barddal, Concept Drift Adaptation in

Text Stream Mining Settings: A Comprehensive Review, 2023. URL: http://arxiv.org/abs/2312.02901. doi:10.48550/arXiv.2312.02901, arXiv:2312.02901 [cs].

[13] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019. URL: http://arxiv.org/abs/1810.04805. doi:10.48550/arXiv.1810.04805, arXiv:1810.04805 [cs].

[14] M. Casavantes, M. Montes-y Gómez, D. I. Hernández-Farías, L. C. González, A. Barrón-Cedeño, PropaLTL at DIPROMATS: Incorporating Contextual Features with BERT's Auxiliary Input for Propaganda Detection on Tweets (2023). URL: https://ceur-ws.org/Vol-3496/dipromats-paper2.pdf.

[15] D. Q. Nguyen, T. Vu, A. T. Nguyen, BERTweet: A pre-trained language model for English Tweets, 2020. URL: http://arxiv.org/abs/2005.10200, arXiv:2005.10200 [cs].

[16] J. M. Pérez, D. A. Furman, L. Alonso Alemany, F. M. Luque, RoBERTuito: a pre-trained language model for social media text in Spanish, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7235–7243. URL: https://aclanthology.org/2022.lrec-1.785.

[17] L. Tian, X. Zhang, M. M.-H. Kim, J. Biggs, Efficient Text-based Propaganda Detection via Language Model Cascades., in: IberLEF@ SEPLN, 2023. URL: https://ceur-ws.org/Vol-3496/dipromats-paper4.pdf.

[18] X. Zhang, Y. Malkov, O. Florez, S. Park, B. McWilliams, J. Han, A. El-Kishky, TwHIN-BERT: A Socially-Enriched Pre-trained Language Model for Multilingual Tweet Representations, 2022. URL: http://arxiv.org/abs/2209.07562. doi:10.48550/arXiv.2209.07562, arXiv:2209.07562 [cs].

[19] M. Grootendorst, BERTopic: Neural topic modeling with a class-based TF-IDF procedure, 2022. URL: http://arxiv.org/abs/2203.05794, arXiv:2203.05794 [cs].

[20] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino, L. Rodés-Guirao, A global database of COVID-19 vaccinations, Nature Human Behaviour 5 (2021) 947–953. URL: https://www.nature.com/articles/s41562-021-01122-8. doi:10.1038/s41562-021-01122-8, publisher: Nature Publishing Group.

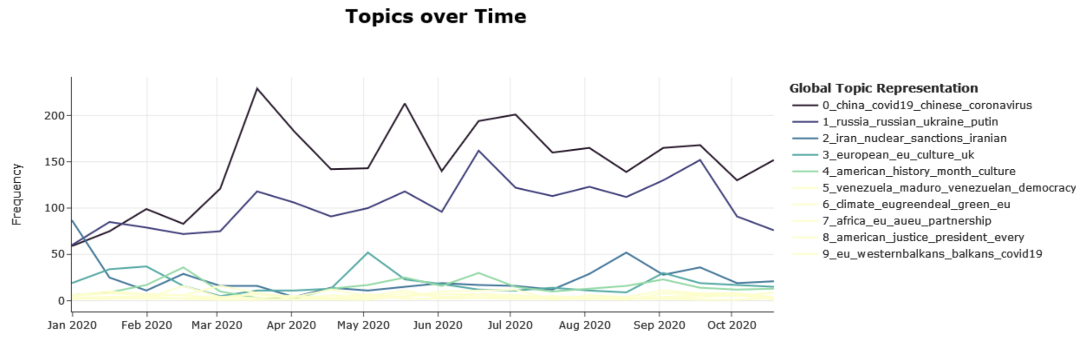
[21] ACLED | Bringing Clarity to Crisis, 2024. URL: https://acleddata.com/.

[22] E. Amigo, A. Delgado, Evaluating Extreme Hierarchical Multi-label Classification, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 5809–5819. URL: https://aclanthology.org/2022.acl-long.399. doi:10.18653/v1/2022.acl-long.399.

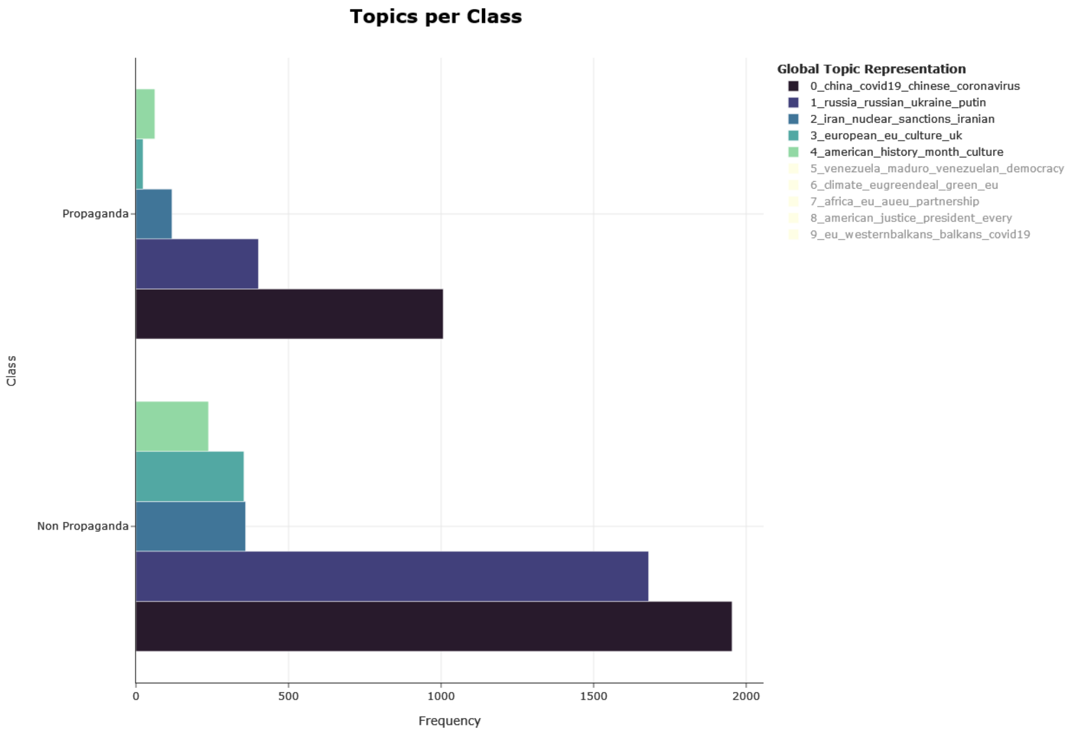
[23] P. Li, L. He, H. Wang, X. Hu, Y. Zhang, L. Li, X. Wu, Learning From Short Text Streams With Topic Drifts, IEEE Transactions on Cybernetics 48 (2018) 2697–2711. URL: https://ieeexplore.ieee.org/document/8039425/. doi:10.1109/TCYB.2017.2748598.

A. Topic Modeling

Topic modeling is an essential technique when attempting to uncover the underlying themes and narratives present in the tweet data. We assert that identifying the contextual features through topic modeling can significantly aid our model in overcoming concept drift. All topic modeling has been performed using the high-performance BERTopic repository [19]. Figure 6a shows how topics in the English tweets of diplomats have evolved over time. This figure traces the rise and fall of different themes, illustrating how diplomatic focus oscillate across different periods. Meanwhile, Figure 6b zooms in on the five most prominent topics within these tweets. It not only highlights their dominance but also explores their distinct tendencies towards propagandistic content.



(a) Evolution of topics frequency over time.



(b) Propaganda frequency of top five topics.

Figure 6: Topic modelling visualization on the English DIPROMATS dataset.