A Spanish dataset for Targeted Sentiment Analysis of political headlines

Tomás Alves Salgueiro[1] [a], Emilio Recart Zapata[1,3] [a], Damián Furman[2] [b], Juan Manuel Pérez[2] [b], and Pablo Nicolás Fernández Larrosa[1] [b]

[1] Instituto de Fisiología, Biología Molecular y Neurociencia, CONICET, UBA,
{talvessalgueiro, fernandezlarrosa}@fbmc.fcen.uba.ar
[2] Instituto de Ciencias de la Computación, CONICET, UBA
{jmperez, dfurman}@dc.uba.ar
[3] Facultad de Psicología, UBA
recartemilio@psi.uba.ar

Abstract. Subjective texts have been especially studied by several works as they can induce certain behaviours in their users. Most work focuses on user-generated texts in social networks, but some other texts also comprise opinions on certain topics and could influence judgement criteria during political decisions. In this work, we address the task of Targeted Sentiment Analysis for the domain of news headlines, published by the main outlets during the 2019 Argentinean Presidential Elections. For this purpose, we present a polarity dataset of 1,976 headlines mentioning candidates in the 2019 elections at the target level. Preliminary experiments with state-of-the-art classification algorithms based on pre-trained linguistic models suggest that target information is helpful for this task. We make our data and pre-trained models publicly available.

1 Introduction

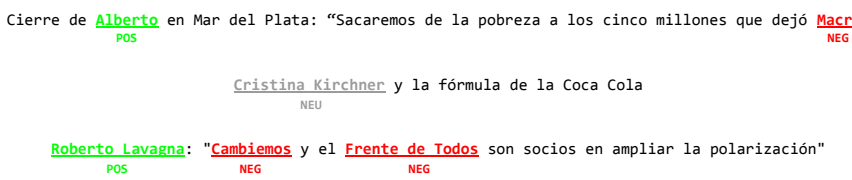
Extracting opinions from subjective texts has attracted a lot of the Interest since the eclosion of Internet and Social Networks, given the unprecedented availability of opinion-rich resources [1]. Most works for opinion mining are directed towards user-generated texts from social networks; however, some other texts –such as news headlines– also convey subjective content about certain topics or entities.

Particularly, it is of interest to analyze the role of the media and campaigns in the formation of judgment criteria during political decisions [2]. The rise of social networks marked a great dynamism in the management and transfer of large volumes of data, playing a fundamental role in the transmission of information in political and candidate campaigns at a massive level [3].

The current work is part of a research project to evaluate the cognitive processes underlying the presidential elections, and assessing the impact of greater exposure to positive content in news headlines associated with the candidates. For this purpose, we are interested in analyzing headlines mentioning candidates by the main national written media.

[a], [b] These authors contributed equally to this work

Fig. 1. Examples of the dataset



In this paper, we present an approach to the task of Targeted Sentiment Analysis for the domain of newspaper headlines. To the best of our knowledge, no Spanish dataset is available for this task. To bridge this gap, we present a novel dataset of headlines mentioning candidates in the 2019 elections in Argentina, having annotations at target level instead of assigning a single polarity to the whole sentence. Preliminary experiments with state-of-the-art techniques suggest that classifiers that consume both the headline and the target improve their performance for the task over those that only consume the headline, giving indications that both sources of information are useful. We make our dataset and the pre-trained models available for further research.

2 Previous work

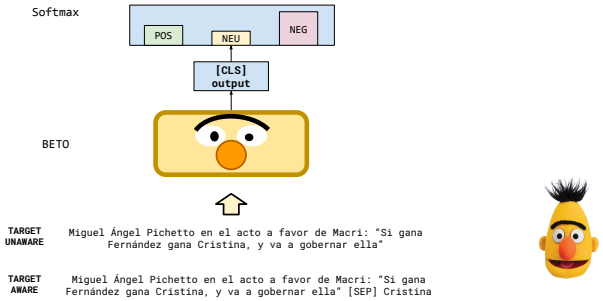
Sentiment analysis and opinion mining have been one of the most popular applications of NLP, and several workshops were dedicated to this topic [4, 5]. This task usually consists of the prediction of the polarity of a text (positive, negative, or neutral) or a Likert-scale rating from negative to positive [6]. One variant of this problem is *Aspect-based sentiment analysis* (ABSA), in which the prediction is performed on a text and a particular aspect [7], something useful to extract multiple opinions from a text. Similarly, *Targeted sentiment analysis* analyses the polarity for a given entity mentioned in the text [8]. As far as we know, in spite of the many resources available for Spanish [9, 5], no dataset is available for this language in this particular subtask.

Pretrained language models based on transformers [10] are state-of-the-art for most NLP tasks. BERT [11] and variants [12] are among the most popular classification techniques nowadays and have top performances on language-understanding benchmarks such as GLUE [13]. Several models have been pre-trained in Spanish [14, 15, 16]. Nozza et al. [17] provide a good overview of pre-trained models across several domains and languages.

3 Data

We collected news headlines published between 21 September and 27 October 2019 by the main national news outlets: Ambito, Clarin, El Cronista, INFOBAE, La Nacion, Pagina 12, Perfil, Popular, La Izquierda Diario, Prensa

Fig. 2. Classification models for the task. Target-unaware classifier consumes only the headline, while the target-aware version consumes both headline and target



Obrera, Tiempo Argentino. We selected only those headlines mentioning one of the contending parties or candidates in the national election.

Three annotators were hired to label the headlines. For each pair of headline and target (a political party or candidate), the annotator assigned a polarity to the pair. To diminish political biases as much as possible, we masked the targets to the annotators (e.g. Mauricio Macri was shown as `[MASK]`).

Each instance was annotated by the three workers as the task is subjective. The agreement was measured using Krippendorff's Alpha [18] and turned out to be $\alpha = 0.62$ (moderate/substantial agreement). A majority voting scheme was used to aggregate the labels, discarding those targets for which the three raters assigned different polarities.

The resulting dataset consists of 1,976 headlines and 2,439 targets. 1,567 headlines have exactly one target, and the remaining have two or more. Among these, 165 headlines feature mixed polarities, mostly in the negative/positive form. Figure 1 illustrates some examples of the dataset.

4 Method

To assess whether algorithms can leverage target information, we proposed a classification experiment with target-aware and target-unaware models:

1. Target-unaware: a classifier only consuming the headline and predicting an overall polarity
2. Target-aware: a classifier consuming both the headline and the target and predicting a polarity for the pair

Our algorithms are based on BETO [14]. The only difference between the two versions is the input: in the target-unaware classifier the input consists of the headline alone, while in the target-aware version it is made of the headline and the target with the special token `[SEP]` in-between them [19].

Table 1. Results of the classification experiments, expressed as the mean ± standard deviation of the 15 runs of the experiments.

Metric		Target	
		Unaware	Aware
Negative	Precision	62.17 ± 3.7	70.23 ± 3.2
	Recall	62.04 ± 4.1	69.77 ± 3.7
	F1	62.01 ± 3.0	69.93 ± 2.6
Neutral	Precision	59.62 ± 3.9	61.04 ± 4.8
	Recall	55.59 ± 8.6	57.09 ± 7.3
	F1	57.07 ± 5.0	58.54 ± 3.5
Positive	Precision	65.95 ± 3.9	70.36 ± 3.9
	Recall	68.97 ± 4.3	73.62 ± 5.4
	F1	67.25 ± 2.0	71.73 ± 2.4
Macro	Precision	62.58 ± 1.7	67.21 ± 1.3
	Recall	62.20 ± 2.0	66.82 ± 1.4
	F1	62.11 ± 2.1	66.73 ± 1.4

Figure 2 illustrates the two algorithms and sample input. We followed standard BERT training for classification tasks, fine-tuning with LR = 10^{-5} for 5 epochs, selecting the best model in terms of Macro F1 for the validation split.

Instead of performing a single train/dev/split, we used a Monte Carlo cross-validation [20] with 15 folds. The splits were performed at the headline level to avoid overestimating the performance.

5 Results

Table 1 displays the results of the classification experiments, expressed as the mean and its standard deviation for the 15 runs. For all metrics, the target aware version performs above the target unaware model.

These differences are statistically significant ($p < 0.001$, one-sided Wilcoxon Mann-Whitney U-test, FDR corrected) [21] for all cases but for the neutral class. This is in line with our data: as neutral targets mostly do not mix with positive or negative polarities, predicting a general polarity is good enough.

6 Conclusion and future work

We presented a dataset for targeted sentiment analysis in Spanish news headlines, showing that state-of-the-art classifiers can leverage its target information. We make the dataset available at the huggingface hub [4].

As future work, we plan to enlarge the dataset with new sources and other events, and explore new classification techniques.

[4] pysentimiento/spanish-targeted-sentiment-headlines

Bibliography

[1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, p. 1–135, jan 2008. [Online]. Available: https://doi.org/10.1561/1500000011

[2] M. J. Kushin and M. Yamamoto, "Did social media really matter? college students' use of online media and political decision making in the 2008 election," in *New Media, Campaigning and the 2008 Facebook Election*. Routledge, 2013, pp. 63–86.

[3] F. Reiter and J. Matthes, ""the good, the bad, and the ugly": A panel study on the reciprocal effects of negative, dirty, and positive campaigning on political distrust," *Mass Communication and Society*, pp. 1–24, 2021.

[4] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 task 4: Sentiment analysis in Twitter," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017, pp. 502–518. [Online]. Available: https://aclanthology.org/S17-2088

[5] M. García-Vegaa, M. C. Díaz-Galianoa, M. Á. García-Cumbrerasa, F. M. P. del Arcoa, A. Montejo-Ráeza, S. M. Jiménez-Zafraa, E. M. Cámarab, C. A. Aguilarc, M. Antonio, S. Cabezudod *et al.*, "Overview of tass 2020: introducing emotion detection," 2020.

[6] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, Oct. 2013, pp. 1631–1642. [Online]. Available: https://aclanthology.org/D13-1170

[7] I. Pavlopoulos, "Aspect based sentiment analysis," *Athens University of Economics and Business*, 2014.

[8] M. Mitchell, J. Aguilar, T. Wilson, and B. Van Durme, "Open domain targeted sentiment," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1643–1654.

[9] C. H. Miranda and J. Guzman, "A review of sentiment analysis in spanish," *Tecciencia*, vol. 12, no. 22, pp. 35–48, 2017.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," pp. 4171–4186, Jun. 2019. [Online]. Available: https://aclanthology.org/N19-1423

[12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[13] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018, pp. 353–355. [Online]. Available: https://aclanthology.org/W18-5446

[14] J. Canete, G. Chaperon, R. Fuentes, and J. Pérez, "Spanish pre-trained bert model and evaluation data," *PML4DC at ICLR*, vol. 2020, 2020.

[15] J. D. L. Rosa, E. G. Ponferrada, M. Romero, P. Villegas, P. G. de Prado Salas, and M. Grandury, "Bertin: Efficient pre-training of a spanish language model using perplexity sampling," *Procesamiento del Lenguaje Natural*, vol. 68, no. 0, pp. 13–23, 2022. [Online]. Available: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6403

[16] J. M. Pérez, D. A. Furman, L. Alonso Alemany, and F. M. Luque, "Robertuito: a pre-trained language model for social media text in spanish," in *Proceedings of the Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, June 2022, pp. 7235–7243. [Online]. Available: https://aclanthology.org/2022.lrec-1.785

[17] D. Nozza, F. Bianchi, and D. Hovy, "What the [mask]? making sense of language-specific bert models," *arXiv preprint arXiv:2003.02912*, 2020.

[18] K. Krippendorff, *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[19] C. Sun, L. Huang, and X. Qiu, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," in *Proceedings of NAACL-HLT*, 2019, pp. 380–385.

[20] K. Gorman and S. Bedrick, "We need to talk about standard splits," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 2786–2791. [Online]. Available: https://aclanthology.org/P19-1267

[21] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in statistics*. Springer, 1992, pp. 196–202.