Breaking the State of the Art in dialects of Spanish Sentiment Analysis

Daniel Palomino [1]

Abstract

Recent advances in neural language models such as ULMFit and BERT have shown impressive results on several Natural Language Processing (NLP) Tasks. However, this same performance for the Spanish language has not been equaled until now. In this work, we present a combined approach based on training a BERT language model within a modified ULMFit pipeline that allows us to obtain state-of-the-art results Spanish sentiment analysis for several dialects. In order to reinforce our approach we have performed our tests on relevant challenges (Martinez-Camara et al., 2017; 2018).

1. Introduction

Spanish is the third language most used on the Internet[1]. However, the development of Natural Language Processing (NLP) techniques for this language have not had the same trend. In particular, this research gap can be observed in Spanish *sentiment analysis*. In this context, the main issue that we aim to address is how to build a polarity detection system that can be interchangeably used across several dialects of Spanish. It is challenging to obtain the same performance when classifying texts written in Spanish from different dialects such as Peru, Argentina, Mexico and so on. While there are a lot of similarities among dialects, there are also several ways to express positive or negative sentiments.

2. Related Work

In recent years, a plethora of works have been published about how Language Modelling and Transfer Learning techniques have improved several NLP tasks including polarity classification in Sentiment Analysis. There are two existing strategies for applying pre-trained language representations in NLP: feature-based and fine-tuning. For the first group, we can include (Arora et al., 2017) (SIF) that propose an innovative sentence embedding approach, and the more recent Embeddings from Language Model (ELMo) (Peters et al., 2018) that has enhanced the state-of-the arts in a lot of NLP tasks. For the second one, it is further important to consider the actual state-of-the art (Howard & Ruder, 2018; Devlin et al., 2018) from which we have been mostly influenced.

Lately, there have been several Deep Learning architectures used for Spanish Twitter Sentiment Analysis, ranging from Multilayer Perceptron (Hurtado et al., 2017), Recurrent Neural Networks (Garcia-Vega et al., 2017) and Convolutional Neural Networks (Segura-Bedmar et al., 2017) and several combinations of them. We refer to Ochoa-Luna & Ari (2018) and Palomino & Ochoa-Luna (2019) in order to get in-depth reviews of several deep Learning approaches for the Spanish language before 2019.

Our proposal is also based on deep learning but, unlike previous approaches, it plans to use a transformer-based language model and contextual data augmentation to improve the polarity detection task on different dialects of Spanish. This setup is novel for the Spanish language.

3. Methodology

While a Deep Learning approach for Spanish Sentiment Analysis would be a natural choice, in the context of similar and at the same time different word meanings among dialects a careful design is needed.

Our proposal is inspired by the success of transfer learning approaches in several text classification tasks for the English language. In particular, we resort to BERT language model (Devlin et al., 2018) and Conditional BERT Contextual Augmentation (Wu et al., 2018) as the extension of the first. In this sense, it has been obtained surprising preliminary results using a mix setup for the Spanish language and its dialects, even higher than the results obtained by previous works (Palomino & Ochoa-Luna, 2019) using another important language model like ULMFiT (Howard & Ruder, 2018).

A general view of all stages of our system is presented in Figure 1 and will be explained in detail in the next paragraphs.

[1]Department of Computer Science, San Pablo Catholic University, Arequipa, Peru. Correspondence to: Daniel Palomino <daniel.palomino.paucar@ucsp.edu.pe>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[1]http://www.internetworldstats.com/stats7.htm

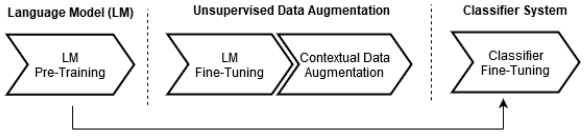

Figure 1. Stages of our system pipeline.

Using a modified version of the pipeline defined by Howard & Ruder (2018):

1. Taking BERT base as principal language model (LM), it is trained on a general domain corpus to capture general features of the language in different layers. To do so, we have learned a LM for the Spanish language using the big Spanish Unannotated Corpora (Cañete, 2019).

2. The full LM is used to create other language model as described by Wu et al. (2018) and re-train it using the novel Conditional Masked Language Model task (C-MLM), presented by the author, on the complete target dataset. In our case, the target dataset is made up of Spanish Tweets. Next, using the second language model, it is used to extend the original labeled Spanish Tweets.

3. An additional classifier layer is located on top of the first language model and is fine-tuned on the target task using gradual unfreezing, discriminative fine-tuning (Discr), and slanted triangular learning rates (STLR) to preserve low-level representations and adapt high-level ones (shaded: unfreezing stages; black: frozen). In our context, the sentiment analysis classifier is fine-tuned using the original labeled Spanish tweets and the extended ones got before.

4. Experiments

A complete description about the hardware and software requirements for reproducing this paper are described in the public repository of the project. In addition, we show some preliminary results.

4.1. Benchmark Datasets

In order to train our algorithms we are using benchmark datasets provided by the TASS competition at SEPLN workshops (Martinez-Camara et al., 2018; 2017). Those datasets comprise several collections of Spanish Tweets including different dialects such as Spanish from Spain, Peruvian and Costa Rican. In addition, those datasets will allow us to compare our approach against recent Deep Learning approaches for Spanish sentiment analysis.

4.2. Results

The result for InterTASS (Task1) Competition 2017 (Martinez-Camara et al., 2017), which only included the Spanish variant from Spain, was better than expected as shown in Table 1, achieving the best result and 6.6% more than the second place, according to M-F1 metric (the ELiRF-UPV team reached a M-F1 score of 0.493).

Furthermore, results on InterTASS (Task1 / Sub-task 2) Competition 2018 (Martinez-Camara et al., 2018), which included dialects such us Spanish from Spain, Peruvian and Costa Rican, are shown in Tables 2, 3 and 4 respectively. Moreover, in addition to have obtained the best result in each dataset, we have reached a difference up to 5.6% more than second place, overcoming the actual state-of-the-art result for this competition.

Table 1. InterTASS 2017 - Spanish dialect.

Team	M-F1	Acc.
Our proposal	**0.559**	**0.668**
ELiRF-UPV-run1	0.493	0.607
RETUYT-svm_cnn	0.471	0.596
ELiRF-UPV-run3	0.466	0.597
ITAINNOVA-model4	0.461	0.476

Table 2. InterTASS 2018 - Spanish dialect.

Team	M-F1	Acc.
Our proposal	**0.559**	**0.668**
elirf-es-run-1	0.503	0.612
retuyt-lstm-es-1	0.499	0.549
retuyt-lstm-es-2	0.498	0.514
retuyt-combined-es	0.491	0.602

Table 3. InterTASS 2018 - Peruvian dialect.

Team	M-F1	Acc.
Our proposal	**0.493**	**0.528**
retuyt-cnn-pe-1	0.472	0.494
atalaya-pe-lr-50-2	0.462	0.451
retuyt-lstm-pe-2	0.443	0.488
retuyt-svm-pe-2	0.441	0.471

5. Future Works

Currently, we are working for including another brand new training task in BERT mechanism in order to improve the contextual representation used in this approach.

Table 4. InterTASS 2018 - Costa Rican dialect.

Team	M-F1	Acc.
Our proposal	**0.556**	**0.643**
retuyt-lstm-cr-2	0.504	0.537
retuyt-svm-cr-2	0.499	0.577
retuyt-svm-cr-1	0.493	0.567
elirf-cr-run-2	0.482	0.561

References

Arora, S., Liang, Y., and Ma, T. A simple but tough-to-beat baseline for sentence embeddings. *International Conference on Learning Representations*, 2017.

Cañete, J. Compilation of large spanish unannotated corpora, May 2019. URL https://doi.org/10.5281/zenodo.3247731. Data retrieved from public repository, https://doi.org/10.5281/zenodo.3247731.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

Garcia-Vega, M., Montejo-Raez, A., Diaz-Galiano, M. C., and Jimenez-Zafra, S. M. Sinai in tass 2017: Tweet polarity classification integrating user information. In *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, pp. 91–96, 2017.

Howard, J. and Ruder, S. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 328–339, Melbourne, Australia, July 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/P18-1031.

Hurtado, L.-F., Pla, F., and Gonzalez, J.-A. Elirf-upv at tass 2017: Sentiment analysis in twitter based on deep learning. In *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, pp. 29–34, 2017.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Martinez-Camara, E., Diaz-Galiano, M., Garcia-Cumbreras, M. A., Garcia-Vega, M., and Villena-Roman, J. Overview of tass 2017. In *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, pp. 13–21, 2017.

Martinez-Camara, E., Almeida-Cruz, Y., Diaz-Galiano, M., Estévez-Velarde, S., Garcia-Cumbreras, M. A., Garcia Vega, M., Gutiérrez, Y., Montejo-Ráez, A., Montoyo, A., Munoz, R., Piad-Morffis, A., and Villena-Roman, J. Overview of tass 2018: Opinions, health and emotions. In *Proceedings of TASS 2018: Workshop on Sentiment Analysis at SEPLN*, pp. 13–27, 2018.

Ochoa-Luna, J. and Ari, D. Deep neural network approaches for spanish sentiment analysis of short texts. In Simari, G. R., Fermé, E., Gutiérrez Segura, F., and Rodríguez Melquiades, J. A. (eds.), *Advances in Artificial Intelligence - IBERAMIA 2018*, pp. 430–441, Cham, 2018. Springer International Publishing. ISBN 978-3-030-03928-8.

Palomino, D. and Ochoa-Luna, J. Advanced transfer learning approach for improving spanish sentiment analysis. In Martínez-Villaseñor, L., Batyrshin, I., and Marín-Hernández, A. (eds.), *Advances in Soft Computing*, pp. 112–123, Cham, 2019. Springer International Publishing. ISBN 978-3-030-33749-0.

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1202. URL https://www.aclweb.org/anthology/N18-1202.

Segura-Bedmar, I., Quiros, A., and Martínez, P. Exploring convolutional neural networks for sentiment analysis of spanish tweets. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 1014–1022. Association for Computational Linguistics, 2017. URL http://aclweb.org/anthology/E17-1095.

Wu, X., Lv, S., Zang, L., Han, J., and Hu, S. Conditional BERT contextual augmentation. *CoRR*, abs/1812.06705, 2018. URL http://arxiv.org/abs/1812.06705.