Overcoming Transformer Fine-Tuning process to improve Twitter Sentiment Analysis for Spanish Dialects

Daniel Palomino[1]

[1]Department of Computer Science, Universidad Católica San Pablo

Motivation

Is there an effective Spanish Sentiment Analysis algorithm? The aim of this paper is to answer this question. The task is challenging because there are, at least, 21 Spanish speaking countries in the world.
In this paper we propose a system which allow us to achieve state-of-the-art results regarding multidialect benchmark datasets.

Introduction

Deriving an effective algorithm for Spanish Twitter sentiment analysis has been long pursued from the research community in this language [1]. Nowadays, despite recent advances in algorithms and word embeddings, the basic polarity detection task has not been completely solved. Thus, whereas it is usually claimed that a transfer learning approach can smoothly solve any classification tasks in NLP [2]; this is not usually the case when we applied it to Spanish. Moreover, the task becomes harder when several language dialects are considered. In fact, low F1-macro values were obtained on previous multi dialect benchmarks [1].
Our proposed classifier relies on three components: a multidialect Spanish corpus, a general language model and a data augmentation step. Thus, we start by training a general language model based on BERT [3] using a multidialect Spanish corpus. Such language model allows us to learn general features of the Spanish language. The final prediction is enhanced using an unsupervised data augmentation process and an enhanced fine-tuned process.

Contact Information

- Web: linkedin.com/in/dpalominop
- Email: daniel.palomino.paucar@ucsp.edu.pe

System Overview

We propose a system to perform polarity classification in small datasets based on the high performance Language Model (LM) called BERT [3]. In addition, by using a clever variation of this LM, we are able to produce contextual data augmentation [4]. Finally, we build the polarity classifier and propose a fine-tune process using groups of layers. Those choices allow us to develop a robust approach that could be used indistinctly in different Spanish dialects and get competitive results in several sentiment analysis tasks. A general view of our system is depicted in Figure 1. Further details are given as follows.

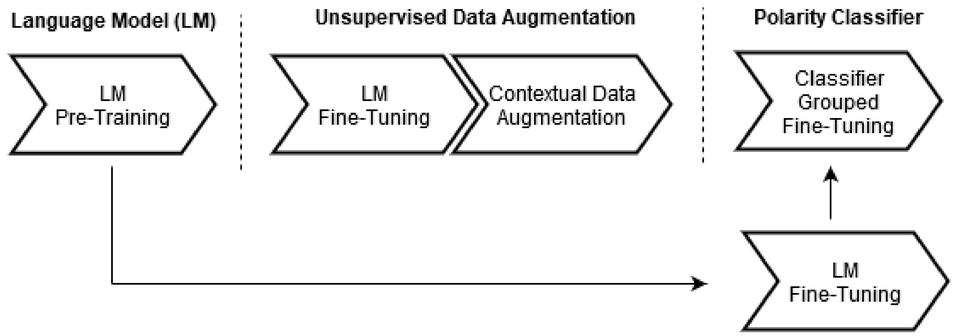

Figure: Overview of system pipeline. General pre-trained LM is used to build unsupervised data augmentation and a robust classifier system using two LM fine-tuning processes in the way. The final classifier is built using the previous steps and grouped fine-tuning technique.

Results I

Our proposal was ranked 1st in all variants of Spanish presented in TASS 2019 competition (Table 1) among all submissions (F1-macro score) and boost the state-of-the-art in 4-labeled sentiment analysis.

Country	Team	F1-macro
CR	**Our proposal**	**0.5529**
	RETUYT-InCo	0.5120
ES	**Our proposal**	**0.5409**
	ELiRF-UPV	0.5070
MX	**Our proposal**	**0.5143**
	ELiRF-UPV	0.5010
PE	**Our proposal**	**0.4862**
	Atalaya	0.4540
UY	**Our proposal**	**0.5609**
	ELiRF-UPV	0.5150

Table: Top 2 results on TASS 2019.

Results II

According to the ablation presented in Table 2, even when every fine-tuning process helps to increase the performance in similar way, it is our proposal of fine-tuning the classifier using unfreezing in groups of layer what was more relevant.

	CR	ES	MX	PE	UY
Our proposal					
All steps	**0.5529**	**0.5409**	**0.5143**	**0.4862**	**0.5609**
wo/ Unsupervised Data Augmentation	0.5293	0.5240	0.5166	0.4741	0.5333
wo/ Classifier: LM fine-tuning	0.5286	0.5371	0.4986	0.4741	0.5273
wo/ Classifier: grouped fine-tuning	0.5252	0.4523	0.4967	0.4329	0.5210

Table: Comparative analysis according to F1-macro metric on TASS 2019 Task 1 - Monovariant test dataset (F1-macro Score) removing one step at once. "wo/" denotes "without".

Conclusion

We have presented a robust sentiment classification system based on BERT that includes several fine-tuning steps such as LM fine-tuning, grouped fine-tuning in the classifier and unsupervised data augmentation. The system has been applied to sentiment analysis on Spanish tweets and its dialects. Despite its simplicity, this approach allowed us to be ranked 1st on the **TASS 2019 Task 1 - Monovariant** [1] and overcome the state-of-the-art.
Furthermore, the ablation experiments have shown that a careful choice of the fine-tuning steps in the LM and classifier can improve the results drastically even more than data augmentation regarding each Spanish dialect.

References

[1] Manuel Diaz-Galiano, Manuel Garcia-Vega, Edgar Casasola, Luis Chiruzzo, Miguel Garcia-Cumbreras, Eugenio Martinez, Daniela Moctezuma, Arturo Montejo, Marco Sobrevilla, Eric Tellez, Mario Graff, and Sabino Miranda.
Overview of tass 2019: One more further for the global spanish sentiment analysis corpus.
In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019) co-located with 35th Conference of the Spanish Society for Natural Language Processing (SEPLN 2019)*, pages 550–560, 2019.

[2] Jeremy Howard and Sebastian Ruder.
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)*, pages 328–339, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.
BERT: pre-training of deep bidirectional transformers for language understanding.
CoRR, abs/1810.04805, 2018.

[4] Xing Wu, Shangwen Lv, Liangjun Zang, Jizhong Han, and Songlin Hu.
Conditional BERT contextual augmentation.
CoRR, abs/1812.06705, 2018.