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

Anonymous Author(s)
Affiliation
Address
email

Abstract

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 several dialects for the Spanish Language. Thus, identically written words could have several meanings and polarities regarding Spanish speaking countries. To tackle this multidialect issue we rely on a transfer learning approach. To do so, we train a BERT language model to "transfer" general features of the Spanish language. Then, we fine-tune the language model to specific dialects. BERT is also used to generate contextual data augmentation aimed to prevent overfitting. Finally, we build the polarity classifier and propose a fine-tuning step using groups of layers. Our design choices allow us to achieve state-of-the-art results regarding multidialect benchmark datasets.

1 Introduction

Deriving an effective algorithm for Spanish Twitter sentiment analysis has been long pursued from the research community in this language [3]. Nowadays, despite recent advances in algorithms (Deep Learning [7]) and word embeddings [1]), 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 [4]; 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 [5, 3].

2 Methodology

We propose a system to perform polarity classification in small datasets based on the high performance Language Model (LM) called BERT [2]. In addition, by using a clever variation of this LM, we are able to produce contextual data augmentation [6]. Finally, we build the polarity classifier and propose a fine-tune process using groups of layers.

To build the classifier, we take the original pre-trained LM and fine-tune this using the target dataset and using again the MLM task in order to enhance the semantic relationship to be extracted later from this. Next, we put a single classification layer on top of the previous fine-tuned LM. Then, we propose fine-tune this classifier using a novel grouped fine-tuning process that consists in freezing a whole architecture and defrost not layer by layer but groups of layers at once and retrain the whole model at each step on the target dataset. Moreover, with the aim of preventing overfitting, we use the new enhanced dataset created before using the aforementioned data augmentation process. A general view of our system is depicted in Figure 1.

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

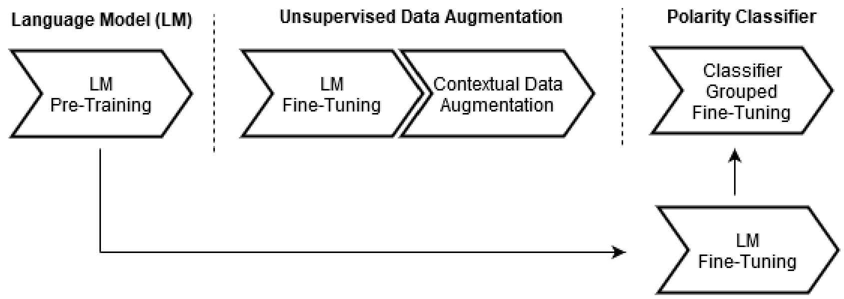

Figure 1: Overview of system pipeline. General pre-trained LM is used to build unsupervised data augmentation and a robust classifier system using two LM fine-tuning processes in the way. The final classifier is built using the previous steps and grouped fine-tuning technique.

3 Results

Our proposal was ranked 1st in all variants of Spanish presented in TASS 2019 competition among all submissions (F1-macro score) and boost the state-of-the-art in 4-labeled sentiment analysis in these dialects of Spanish adding a difference of even up to 4.6 percent with respect to the second place. A detailed report of results for **TASS 2019 Task 1 - Monovariant** is presented in Table 1.

Country	Team	F1-macro
CR	**Our proposal**	**0.5529**
	RETUYT-InCo	0.5120
	ELiRF-UPV	0.4960
ES	**Our proposal**	**0.5409**
	ELiRF-UPV	0.5070
	Atalaya	0.4840
MX	**Our proposal**	**0.5143**
	ELiRF-UPV	0.5010
	GTH-ETSIT-UPM	0.4870
PE	**Our proposal**	**0.4862**
	Atalaya	0.4540
	ELiRF-UPV	0.4470
UY	**Our proposal**	**0.5609**
	ELiRF-UPV	0.5150
	Atalaya	0.4990

Table 1: Top 3 results on TASS 2019.

Furthermore, in order to get a better understanding of these outstanding results, we perform ablation experiments removing every step of our pipeline at once and testing the performance of the remaining system. The outputs of these experiments are shown in Table 2 where we can observe the real impact of every step of our proposal. According to this, 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 2: 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".

References

[1] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

[2] 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.

[3] 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.

[4] 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.

[5] Eugenio Martinez-Camara, Yudivián Almeida-Cruz, Manuel Diaz-Galiano, Suilan Estévez-Velarde, M. Angel Garcia-Cumbreras, Manuel Garcia Vega, Yoan Gutiérrez, Arturo Montejo-Ráez, Andres Montoyo, Rafael Munoz, Alejandro Piad-Morffis, and Julio Villena-Roman. Overview of tass 2018: Opinions, health and emotions. In *Proceedings of TASS 2018: Workshop on Sentiment Analysis at SEPLN*, pages 13–27, 2018.

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

[7] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis : A survey. *CoRR*, abs/1801.07883, 2018.