RETUYT-InCo at TASS 2019: Sentiment Analysis in Spanish Tweets

Marcos Pastorini, Mauricio Pereira, Nicolás Zeballos,
Luis Chiruzzo, Aiala Rosá, and Mathias Etcheverry

Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay,
{marcos.pastorini,mauricio.pereira,nicolas.zeballos,
luischir,aialar,mathiase}@fing.edu.uy

Abstract. This paper presents three approaches for classifying the sentiment of tweets for different Spanish variants in the TASS 2019 challenge. The classifiers are based on Multilayer Perceptron (MLP), Long Short Term Memory networks (LSTM), and transfer learning using BERT.

Keywords: Sentiment Analysis, Machine Learning, Neural Networks, Word Embeddings

1 Introduction

Sentiment analysis in tweets is an interesting task due to the large volume of information generated every day, the subjective nature of most messages, and the easy access to this material for analysis and processing. Specific tasks related to this field have been organized for several years now: the International Workshop on Semantic Evaluation (SemEval) includes a task on Tweets Sentiment Analysis since 2013 [17], and, exclusively for Spanish, the TASS workshop, organized by the SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural), exists since 2012 [20].

As in many NLP areas, in the last years most of the works on sentiment analysis have incorporated techniques based on Deep Learning and Word Embeddings, in search of improving results. In recent editions of the TASS shared tasks (2017 and 2018), the majority of participating systems rely on different neural network models and on the use of word embeddings [10,12]. However, approaches based on classic machine learning models (like SVM), when including word embedding based features, remain competitive, reaching the top positions for some test corpora [12].

In TASS 2018 (task 1) the best results were obtained by systems which used deep learning [4,7], SVM [4], and genetic algorithms combined with SVM [15].

All of them used word embeddings for words and tweets representation. Results for monolingual experiments (using a single Spanish variant) were better than results for crosslingual experiments. As in previous TASS editions, neutral tweets are the most difficult to recognize.

In TASS 2017 [13] the most used methods were deep learning and word embeddings as well. The best results were obtained by: [9], who experimented with different deep neural network architectures, using as input domain-specific and general-domain sets of embeddings; [3], who presented an ensemble of SVM and Logistic Regression classifiers; [18], who presented an SVM classifier based on the centroid of the tweets embeddings, a deep neural network (CNN), and a combination of both; and [14], who combined an SVM classifier with genetic programming. On the other hand, in the TASS editions prior to 2017, most of the participants presented machine learning systems based on hand crafted features.

SemEval 2018 [16], has included for the first time a dataset for Spanish tweets sentiment analysis. The best results for Spanish were obtained by systems based on deep neural networks (Convolutional Neural Networks and Recurrent Neural Networks) and SVM, based on word embeddings [11,19,1,8]. Some of them extended the training set by translating English tweets [11,19]. Other systems used subjective lexicons (Spanish lexicons and translated English lexicons).

In this paper we describe three different approaches for Spanish tweet classification presented by the RETUYT-InCo team for the TASS 2019 [6] sentiment analysis challenge : a classifier based on Multilayer Perceptron (MLP), a Long Short Term Memory (LSTM) network, and transfer learning using the BERT model.

2 Approaches

2.1 Approach 1: Sentence Word Vectors Mean

In this approach we considered multiple variants to perform the classification of the sentence word vectors mean. This kind of sentence representation is not aware of the sentence order but it can give surprisingly good results. We performed the classification through layered fully connected neural networks and support vector machines.

Model Features We attempted to improve the classification performance including the following additional features.

- **Sentiment Lexicon:** We considered a sentiment lexicon constituted by two sets of words: positive and negative. We added two dimensions to the input with the amount of words in the tweet of each set.
- **Tweet split:** We split the input tweets into phrases and obtained the mean vector through the mean vector of each phrase.
- **Dimensionality Reduction:** We reduced the input vector dimensionality using Principal Component Analysis (PCA).
- **Crosslingual exclusion:** In the crosslingual task we excluded an additional country from the training data.

Experiments Performed We trained multiple classifiers to achieve a good model setting observing an overall better performance with MLP than SVM. Regarding dimensionality reduction we used PCA trying many numbers of principal components, however, we could not obtain any considerable performance improvement with this approach.

In the crosslingual experiments, we trained models excluding each country and we observed a performance improvement when excluding the Peru corpus from the training data (for all the Spanish variants). Note also that the results obtained for the Peru corpora are lower than the ones obtained for the remaining variants.

Regarding the MLP we considered batch sizes of 100 and 200 and different configuration of hidden layer sizes.

2.2 Approach 2: BERT Transfer Learning

This approach relies on the transfer learning from a pretrained Spanish BERT [5] model to the sentiment detection task. The BERT model consists mainly in three parts: word embeddings, encoder, and classifier. The input in term of its word embeddings is given to the encoder and the encoder output is used to perform the classification through a fully connected layer with a softmax output.

Pre-processing In this approach the following pre-processing steps were considered to process the input tweet before it is given to the model:

- **User normalization:** Each user id (e.g. @tass) was replaced by the token *User*.
- **URLs:** Each url was replaced by the token *URL*.
- **Hashtags:** The hash symbols (#) were deleted.
- **Laugh normalization:** The laugh tokens (e.g. *ja*, *jajj*, *jajjajaj*, etc.) were replaced by *jaja*.
- **Numbers:** The numbers (e.g. *1*,*2*,*3*,...) were replaced by its word names (*uno*, *dos*, *tres*, ...)

Fine Tuning In the experiments we performed four fine tuning strategies: only the encoder, only the classifier, both sequentially (first encoder and then classifier) and both jointly. The best results were obtained when the encoder and classifier are fine tuned jointly.

The best performances were obtained when the model was trained for 5 epochs balancing the training corpus according to its polarity. When the training corpus is not balanced, the model tends to overfit and give the majority class.

This approach performed better in the crosslingual than in the monolingual tasks. Probably, this was because of the reduced training data for the monolingual task.

2.3 Approach 3: FastText LSTM

The third approach is based on the use of fastText embeddings [2] as input for an LSTM neural network.

The tweets were normalized, as in [18] and [4], including abbreviation normalization (que, porque, etc.) and emojis substitution (:) :D etc).

FastText was used to train 300 dimension vectors, using the training corpus of each variant for monolingual experiments, and the whole training corpus for crosslingual experiments.

The best results on the development corpora were obtained by an LSTM network with two LSTM layers, and the following configuration:

- drop out: 0.5
- number of words in input: 20
- number of neurons: 256
- batch size: 256
- embeddings adjust: TRUE
- using the early stopping and model checkpoint techniques

3 Results

Five different corpora considering five Spanish variants were used for this task: Spain (ES), Costa Rica (CR), Peru (PE), Uruguay (UY), and Mexico (MX). Furthermore, the systems could be trained with training data for the corresponding Spanish variant (monolingual case), or they could be trained using data from other variants (crosslingual case). We decided to submit the best results for each approach on each of the variants and training combinations. Despite we performed some experiments on the Uruguayan dataset, we decided not to send our results on this corpus because we participated in the corpus annotation process. The results we obtained on the test corpora are shown in table 1.

Table 1. Results for test.

	System 1			System 2			System 3		
Corpus	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Mono ES	0.431	0.456	0.443	0.087	0.250	0.129	0.416	0.364	0.388
Mono CR	0.588	0.454	0.512	0.032	0.250	0.057	0.363	0.337	0.350
Mono PE	0.437	0.439	0.438	0.422	0.397	0.409	0.321	0.304	0.312
Mono MX	0.487	0.485	0.486	0.396	0.378	0.387	0.417	0.401	0.407
Cross ES	0.431	0.456	0.443	0.456	0.465	0.460	0.386	0.379	0.383
Cross CR	0.486	0.453	0.469	0.452	0.446	0.449	0.408	0.392	0.400
Cross PE	0.367	0.443	0.401	0.436	0.417	0.427	0.389	0.400	0.394
Cross MX	0.456	0.470	0.462	0.455	0.474	0.465	0.387	0.381	0.384

For the monolingual datasets, the best results were achieved by the first approach, based on MLP. On the other hand, for crosslingual datasets, the system based on the BERT model performed better (except for the case of Costa Rica).

Comparing to the systems developed by the other teams participating in the TASS share task, we obtained some interesting results. Our MLP approach was ranked first on the corpus from Costa Rica for the monolingual task, and our BERT based approach was ranked first on the corpus from Spain for the crosslingual task.

Concerning the Uruguayan datasets, we had a good performance evaluating on the development corpus, reaching a Macro-F of 0.50 for the monolingual task (BERT based approach) and 0.48 for the crosslingual task (MLP based approach).

4 Conclusions

We presented three approaches for TASS 2019 about classifying the sentiment of tweets in different Spanish variants. The approaches we used are: MLP using word embedding centroids and manually crafted features, transfer learning based on the BERT model, and LSTM using fastText word embeddings.

Our MLP based approach achieved good results in monolingual experiments while the BERT based system performed better in the crosslingual task. Probably this approach needs a bigger dataset for training.

References

1. Abdou, M., Kulmizev, A., Ginés i Ametllé, J.: Affecthor at semeval-2018 task 1: A cross-linguistic approach to sentiment intensity quantification in tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 210–217. Association for Computational Linguistics (2018), http://aclweb.org/anthology/S18-1032
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
3. Cerón-Guzmán, J.A.: Classifier ensembles that push the state-of-the-art in sentiment analysis of spanish tweets. In: Proceedings of TASS (2017)
4. Chiruzzo, L., Rosá, A.: RETUYT-InCo at TASS 2018: Sentiment Analysis in Spanish Variants using Neural Networks and SVM. In: Proceedings of TASS (2018)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2018)
6. Díaz-Galiano, M.C., et al.: Overview of tass 2019. CEUR-WS, Bilbao, Spain (2019)
7. González, J., Pla, F., Hurtado, L.: ELiRF-UPV at TASS 2018: Sentiment Analysis in Twitter based on Deep Learning. In: Proceedings of TASS (2018)
8. González, J.Á., Hurtado, L.F., Pla, F.: Elirf-upv at semeval-2018 tasks 1 and 3: Affect and irony detection in tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 565–569. Association for Computational Linguistics (2018), http://aclweb.org/anthology/S18-1092
9. Hurtado, L.F., Pla, F., González, J.: ELiRF-UPV at TASS 2017: Sentiment Analysis in Twitter based on Deep Learning. In: Proceedings of TASS (2017)

10. y Eugenio Martínez-Cámara y M. Ángel García Cumbreras y Manuel García Vega y Julio Villena Román, M.C.D.G.: The democratization of deep learning in tass 2017. Procesamiento del Lenguaje Natural **60**(0), 37–44 (2018), http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5556
11. Kuijper, M., Lenthe, M., Noord, R.: Ug18 at semeval-2018 task 1: Generating additional training data for predicting emotion intensity in spanish. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 279–285. Association for Computational Linguistics (2018), http://aclweb.org/anthology/S18-1041
12. Martínez-Cámara, E., Almeida Cruz, Y., Díaz-Galiano, M.C., Estévez Velarde, S., García-Cumbreras, M.A., García-Vega, M., Gutiérrez Vázquez, Y., Montejo Ráez, A., Montoyo Guijarro, A., Muñoz Guillena, R., Piad Morffis, A., Villena-Román, J.: Overview of TASS 2018: Opinions, health and emotions. In: Martínez-Cámara, E., Almeida Cruz, Y., Díaz-Galiano, M.C., Estévez Velarde, S., García-Cumbreras, M.A., García-Vega, M., Gutiérrez Vázquez, Y., Montejo Ráez, A., Montoyo Guijarro, A., Muñoz Guillena, R., Piad Morffis, A., Villena-Román, J. (eds.) Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN (TASS 2018). CEUR Workshop Proceedings, vol. 2172. CEUR-WS, Sevilla, Spain (September 2018)
13. Martínez-Cámara, E., Díaz-Galiano, M.C., García-Cumbreras, M.A., García-Vega, M., Villena-Román, J.: Overview of tass 2017. In: Villena Román, J., García Cumbreras, M.A., Martínez-Cámara, Eugenio, D.G.M.C., García Vega, M. (eds.) Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN (TASS 2017). CEUR Workshop Proceedings, vol. 1896. CEUR-WS, Murcia, Spain (September 2017)
14. Moctezuma, D., Graff, M., Miranda-Jiménez, S., Tellez, E., Coronado, A., Sánchez, C., Ortiz-Bejar, J.: A genetic programming approach to sentiment analysis for twitter: Tass'17. In: Proceedings of TASS (2017)
15. Moctezuma, D., O.B.J.T.E.M.J.S.G.M.: INGEOTEC solution for Task 1 in TASS'18 competition. In: Proceedings of TASS (2018)
16. Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: Semeval-2018 task 1: Affect in tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 1–17 (2018)
17. Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., Wilson, T.: SemEval-2013 Task 2: Sentiment Analysis in Twitter. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). pp. 312–320 (2013)
18. Rosá, A., Chiruzzo, L., Etcheverry, M., Castro, S.: RETUYT en TASS 2017: Análisis de Sentimientos de Tweets en Español utilizando SVM y CNN. In: Proceedings of TASS (2017)
19. Rozental, A., Fleischer, D.: Amobee at semeval-2018 task 1: Gru neural network with a cnn attention mechanism for sentiment classification. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 218–225. Association for Computational Linguistics (2018), http://aclweb.org/anthology/S18-1033
20. Villena Román, J., Lana Serrano, S., Martínez Cámara, E., González Cristóbal, J.C.: TASS - Workshop on Sentiment Analysis at SEPLN (2013-03)