LaSTUS/TALN at TASS 2019: Sentiment Analysis for Spanish Language Variants with Neural Networks

Lutfiye Seda Mut Altin, Àlex Bravo, and Horacio Saggion

LaSTUS-TALN Research Group, DTIC
Universitat Pompeu Fabra
C/Tànger 122-140, 08018 Barcelona, Spain
{name.surname}@upf.edu

Abstract. This paper describes the participation of LaSTUS/TALN team in the shared task Sentiment Analysis at SEPLN (TASS) organized in the context of IberLEF 2019. TASS focuses on the classification of tweets written in the Spanish language (from Spain, Peru, Costa Rica, Uruguay and Mexico) with respect to their polarity or sentiment. This year TASS proposes two sub-tasks: monolingual and cross-lingual sentiment analysis. This paper presents a deep learning approach based on bidirectional LSTM (biLSTM) models to face both sub-tasks. The paper reports and discusses the official results achieved by our team.

Keywords: Natural Language Processing · Neural Networks · Sentiment Analysis · Spanish Language

1 Introduction

Sentiment analysis is the process of detecting subjective information of a given text such as whether the text expresses a positive, negative or neutral opinion. Sentiment analysis is widely used in several application areas. For instance, private companies or political organizations are interested in knowing what their clients think about their product or services [7, 11]. The number of users of micro-blogging platforms such as Twitter grows day by day, making data from these sources very useful for opinion mining and sentiment analysis.

TASS at IberLEF 2019[1] focuses on the evaluation of polarity classification systems of tweets written in the Spanish language spoken in Spain, Peru, Costa Rica, Uruguay and Mexico [1]. The task consists of two sub-tasks:

– **Subtask 1:** Monolingual Sentiment Analysis: Training and test using each InterTASS dataset (ES-Spain, PE-Peru, CR-Costa Rica, UR-Uruguay and MX-Mexico).

[1] https://sites.google.com/view/iberlef-2019/

– **Subtask 2:** Cross-lingual Sentiment Analysis: Training in a combination of datasets while using a different dataset to test. Since the languages spoken in different Spanish-speaking countries differ considerably one-another, this is a very challenging problem.

This paper describes a neural network for sentiment analysis of Tweets in Spanish. The rest of the paper is organized as follows: In section 2, we present an overview of the related work for sentiment analysis, specifically on Spanish. In Section 3, we describe our model. In Section 4, we provide the results and discuss the performance of the system. Lastly, in Section 5, we give the conclusions.

2 Related Work

Previous research on Twitter sentiment analysis can be considered in two categories: supervised approaches and lexicon-based approaches. Where supervised methods are of concerned, the algorithms used are based on classifiers such as Random Forest, Support Vector Machine, Naive Bayes with diverse features such as Part-Of-Speech (POS) tags, N-grams, hashtags, retweets, emoticons [2, 5, 3]. In lexicon-based approaches, dictionaries of words with their sentiment orientations have been used [9, 8, 13]. Deep learning methods have recently gained popularity in this area [4, 14]. Tang et. al gave an overview for sentiment analysis and stated that many studies with machine learning approach focused on building powerful feature extractor with domain expert and feature engineering; however deep learning approaches emerged as powerful computational models that discover complex semantic representations of texts automatically from data without feature engineering.[12] Moreover, recent sentiment analysis shared tasks on various languages also showed that top ranked systems used deep learning approaches or deep learning ensembles.[10]

In the previous edition of TASS (in 2018) [6], the Task 1 also promoted the development and evaluation of systems able to automatically detect the polarity of tweets written in Spanish. Five system were presented and most of them used deep learning algorithms, combining different ways of obtaining word embeddings combining them with hand-crafted linguistic features.

3 Data and Methodology

The participants were provided with a training and a development corpora and several test corpora. All the corpora are annotated with 4 different levels of opinion intensity as positive, negative, neutral or none (P, N, NEU, NONE).

We address the problem with a neural network based on two bidirectional LSTM (biLSTM) models with two dense layers at the end. In Figure 1 a simplified schema of our shared model can be seen.

First, the tweets were preprocessed removing punctuation marks and keeping emojis and full hashtags since they can contribute to define the meaning of a tweet, and then, the tweets were tokenized.

Second, the embedding layer transforms each element in the tokenized tweet into a low-dimension vector. The embedding layer was randomly initialized from a uniform distribution (between -0.8 and 0.8 values and with 100 dimensions). In addition, the initialized embedding layer was updated with the corresponding word vectors related to Spanish variant to predict, which were updated during the training. These word vectors are included in a pre-trained model from Regional Embeddings [2], which provides FastText word embeddings for Spanish language variations.

Then, two subsequent biLSTM layers get high-level features from previous embeddings with 128 and 64 units, respectively. A disadvantage of LSTM models is that they compress all information into a fixed-length vector, causing the incapability of remembering long tweets. To overcome the limitation of fixed-length vector keeping relevant information from long tweet sequences, after biLSTMs, we added an attention layer producing a weight vector and merge word-level features from each time step into a tweet-level feature vector, by multiplying the weight vector [15]. Next, the tweet-level feature vector produced by the previous layers is decreased by a fully-connected layer with a ReLU as activation function and an output of 64 elements. Finally, the output produced by the previous layer is used for classification task by a fully-connected layer with Softmax as activation function.

Moreover, to be able to mitigate overfitting problem we applied dropout regularization. Dropout operation sets randomly to zero a proportion of the hidden units during forward propagation, creating more generalizable representations of data. In the model, we employ dropout on the embeddings and biLSTM layers. The dropout rate was set to 0.5 in all cases. Finally, the model was compiled using the Adam optimizer and the categorical cross-entropy as loss function.

4 Results

In the Subtask 1 (monolingual sentiment analysis), we used the training and test dataset for each language (ES-Spain, PE-Peru, CR-Costa Rica, UR-Uruguay and MX-Mexico). For this Subtask, our results have been ranked between third and fifth positions depending on the Spanish variant (see Table 1).

On the other hand, in the Subtask 2 (cross-lingual sentiment analysis), we trained our model using all datasets other than the test dataset. For example, to predict results in Spanish (ES), we trained with the data for the following Spanish variants: PE-Peru, CR-Costa Rica, UR-Uruguay and MX-Mexico. In this case, we have achieved better results, between the second and third positions depending on the Spanish variant (see Table 2).

5 Conclusions

In this paper, we presented our results for the participation to TASS task of IberLEF 2019. We described and evaluated our system which is based on two

[2] https://github.com/INGEOTEC/

Table 1. Ranking and the results in the Subtask 1 (monolingual sentiment analysis)

Ranking	F1 Score	Precision	Recall	Language Variety
First System (1)	0.533034	0.443634	0.484242	ES
Last System (8)	0.087046	0.250000	0.129130	ES
Our Approach (3)	0.463609	0.470095	0.457299	ES
First System (1)	0.461602	0.446265	0.453804	PE
Last System (8)	0.320949	0.303982	0.312235	PE
Our Approach (3)	0.420549	0.437464	0.404894	PE
First System (1)	0.587928	0.454011	0.512363	CR
Last System (8)	0.057327	0.250000	0.250000	CR
Our Approach (4)	0.455959	0.453540	0.458405	CR
First System (1)	0.640615	0.519142	0.573517	UY
Last System (6)	0.345662	0.388452	0.365810	UY
Our Approach (4)	0.468353	0.473410	0.463403	UY

Table 2. Ranking and the results in the Subtask 2 (cross-lingual sentiment analysis)

Ranking	F1 Score	Precision	Recall	Language Variety
First System (1)	0.455952	0.465083	0.460472	ES
Last System (8)	0.242442	0.226824	0.234373	ES
Our Approach (2)	0.458758	0.455505	0.462058	ES
First System (1)	0.468214	0.480260	0.474161	PE
Last System (7)	0.271690	0.254126	0.262615	PE
Our Approach (3)	0.447784	0.441971	0.453752	PE
First System (1)	0.478759	0.469602	0.474136	CR
Last System (7)	0.191637	0.179832	0.185547	CR
Our Approach (2)	0.464940	0.472122	0.457973	CR
First System (1)	0.516860	0.510278	0.513548	UY
Last System (5)	0.218466	0.239803	0.218466	UY
Our Approach (3)	0.469439	0.449887	0.490768	UY

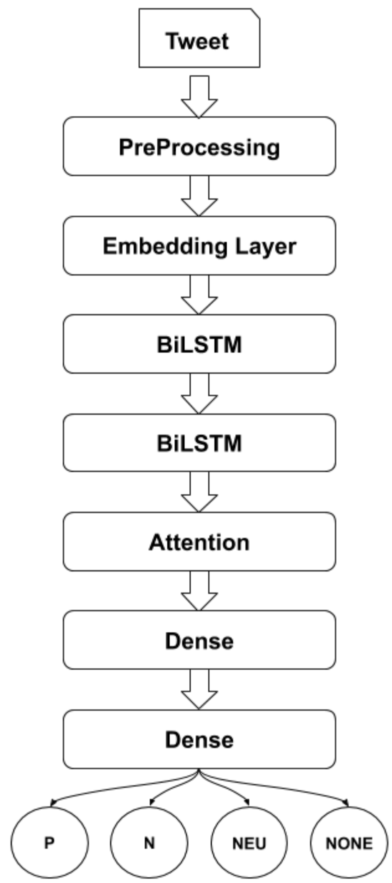

Fig. 1. Simplified schema of the model

biLSTM models with an Attention layer, to classify the tweet in 4 different levels of opinion intensity (P, N, NEU, NONE). Regarding the results of the TASS task, we have achieved better results in the cross-lingual sub-task, although the model has been trained with different Spanish variants, there was more data to learn the classification than the monolingual task. In Table 1 and Table 2, we can also observe the best system of the task. Our results are usually close to the winning system, indicating the difficulty of the task. Due to time constraints, we were not able to perform an error analysis, for that reason, in future work, we will work in a detailed error analysis in order to understand the limitations of our approach. Furthermore, more detailed analyses on integration of linguistic annotations into neural network and other models (such as convolution) can be considered in order to improve the performance of the model.

Acknowledgements

Our work is partly supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502). We thank to reviewers for their constructive comments.

References

1. Díaz-Galiano, M.C., et al.: Overview of tass 2019. CEUR-WS, Bilbao, Spain (2019)
2. Hagen, M., Potthast, M., Büchner, M., Stein, B.: Twitter sentiment detection via ensemble classification using averaged confidence scores. In: European Conference on Information Retrieval. pp. 741–754. Springer (2015)
3. Jianqiang, Z.: Combing semantic and prior polarity features for boosting twitter sentiment analysis using ensemble learning. In: 2016 IEEE First International Conference on Data Science in Cyberspace (DSC). pp. 709–714. IEEE (2016)
4. Jianqiang, Z., Xiaolin, G., Xuejun, Z.: Deep convolution neural networks for twitter sentiment analysis. IEEE Access **6**, 23253–23260 (2018)
5. Jianqiang, Z., Xueliang, C.: Combining semantic and prior polarity for boosting twitter sentiment analysis. In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). pp. 832–837. IEEE (2015)
6. Martínez Cámara, E., Almeida Cruz, Y., Díaz Galiano, M.C., Estévez-Velarde, S., García Cumbreras, M.Á., García Vega, M., Gutiérrez, Y., Montejo Ráez, A., Montoyo, A., Muñoz, R., et al.: Overview of tass 2018: Opinions, health and emotions (2018)
7. Martínez-Cámara, E., Martín-Valdivia, M.T., Urena-López, L.A., Montejo-Ráez, A.R.: Sentiment analysis in twitter. Natural Language Engineering **20**(1), 1–28 (2014)
8. Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M.T., Ureña-López, L.A.: A knowledge-based approach for polarity classification in t witter. Journal of the Association for Information Science and Technology **65**(2), 414–425 (2014)
9. Paltoglou, G., Thelwall, M.: Twitter, myspace, digg: Unsupervised sentiment analysis in social media. ACM Transactions on Intelligent Systems and Technology (TIST) **3**(4), 66 (2012)
10. Rosenthal, S., Farra, N., Nakov, P.: Semeval-2017 task 4: Sentiment analysis in twitter. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). pp. 502–518 (2017)
11. Saggion, H., Funk, A.: Extracting opinions and facts for business intelligence. Revue des Nouvelles Technologies de lInformation (RNTI) **E-17**, 119–146 (2009)
12. Tang, D., Qin, B., Liu, T.: Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **5**(6), 292–303 (2015)
13. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology **63**(1), 163–173 (2012)
14. Wehrmann, J., Becker, W.E., Barros, R.C.: A multi-task neural network for multilingual sentiment classification and language detection on twitter. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. pp. 1805–1812. ACM (2018)

15. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). vol. 2, pp. 207–212 (2016)