LaSTUS/TALN at IroSvA: Irony Detection in Spanish Variants

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. The *Irony Detection in Spanish Variants* (IroSvA'19) shared task aims at investigating the recognition of irony in Twitter messages in three different Spanish variants: from Spain , Mexico, and Cuba. This paper describes our approach to the shared task: a Neural Network system based on a simple bidirectional LSTM (biLSTM) model. Since we have developed the system in the context of other related IberLEF 2019 shared tasks, we train multiple models simultaneously sharing some of the layers of the neural architecture between them. Furthermore, we also have applied a method for data augmentation given the scarcity of resources which were available.

Keywords: Irony Detection · biLSTM · neural network · language variety.

1 Introduction

Figurative language is an important device for communication in social media allowing people to express themselves in unexpected ways. Due to its significance, investigating figurative language such as irony has gathered considerable attention from various disciplines.

The main problem with automatic detection of irony is that ironic messages aim to convey the opposite meaning of what is literally said or written. This effect is sometimes achieved using humor as a key device. For various natural language processing (NLP) applications, irony detection has a great potential, especially where semantic analysis is of concern.

IroSvA focuses on the detection of irony in short messages (tweets and news comments) written in Spanish and with respect to a given context [6]. Tree sub-tasks are proposed:

– **Subtask A:** Irony detection in Spanish tweets from Spain

– **Subtask B:** Irony detection in Spanish tweets from Mexico
– **Subtask C:** Irony detection in Spanish news comments from Cuba

Therefore, in addition to detect if a text is ironic or not depending on the context, the problem is also related to the way irony is expressed in distinct Spanish variants. This paper describes a neural network for irony detection for different dialects. The rest of the paper is organized as follows: In section 2, we present an overview of the related work for irony detection. In Section 3, we describe our model and the differences between different runs for each sub-task. 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 automatic irony detection follows two main approaches: (i) rule-based or (ii) machine-learning based methods [4], sometimes combining both. Rule-based approaches try to identify irony through a set of rules, such as looking for a positive verb and a negative situation phrase in a sentence [8].

Apart from rule based systems, there are machine learning approaches with handcrafted feature engineering and deep-learning systems. Most of the machine learning approaches start by calculating the features with various methods to then use a machine learning algorithm to classify the text as ironic or not. González-Ibánez et. al investigated the contribution of linguistic and pragmatic features and found that pragmatic features such as positive emoticons (e.g. smileys), negative emoticons (e.g. frowning faces) and user mentions – indicating that a message is addressed to some specific entity – are distinctive features [3]. Barbieri et al. used Random Forest and Decision tree classifiers utilizing a group of features (e.g. frequency, written-spoken style, gap between positive and negative terms) [1].

More recently, among the systems that participated in a shared task on irony detection in Italian, EVALITA 2018, innovative deep learning approaches showed high performance, with the best performing system based on a deep learning approach [2]. In another shared task at SemEval-2018, the best performing system's architecture consisted of densely connected LSTMs; based on pre-trained word embeddings, sentiment features and syntactic features (e.g. PoS-tag features). In addition, at this shared task most frequently preferred features were lexical features such as n-grams, punctuation, emoji presence and sentiment or emotion-lexicon features [9].

On the other hand, some previous works focusing only on detecting the language variety of the tweets have highlighted the challenges researchers face [7, 9].

3 Data and Methodology

The corpus that was provided by the shared task organizers consist of 9,000 short messages about different topics written in Spanish as 3,000 from Cuba, 3,000 from Mexico and 3,000 from Spain. The messages are annotated according to being ironic or not. Roughly, 80% of the corpus is given for training purposes and the remaining 20% is given for testing purpose.

In this work, we presented a neural network based on a simple bidirectional LSTM (biLSTM) model with two dense layers at the end to detect ironic messages based on our previous work [5]. We considered each Spanish variant as a different task, therefore the corpus provided by the organizers may not be enough to train a neural network (2,400 instances per Spanish variant in the training dataset) and it could present overfitting or underfitting problems during the training. For that reason, we have used data from different tasks in order to train more examples in the model. In the context of the IberLEF 2019, we have selected three additional task to train with IroSvA at the same time:

- From MEX-A3T, we used the Aggressiveness Identification track, which focuses on the detection of aggressive comments in tweets from Mexican users.
- From HAHA, we used the classification task related to identify if a Spanish tweet is a joke or not.
- The TASS 2019 focuses on the evaluation of polarity classification systems of tweets written in Spanish. We used the data related to this task, tweets written in the Spanish language spoken in Spain, Peru, Costa Rica, Uruguay and Mexico, which were annotated with 4 different levels of opinion intensity (Positive, Negative, Neutral and Nothing).

In this scenario, we defined an Embedding layer for each Spanish variant. Classification tasks with the same Spanish variant used the same Embedding layer during the training process. For instance, the embedding layer related to the Spanish from Mexico was used by the MEX-A3T task, the Mexican part of the TASS 2019 task and the tweets written in the Spanish language spoken in Mexico from IroSvA. Furthermore, all task shared the biLSTM layer during training.

In Figure 1 a simplified schema of our shared model can be seen. In the following we explain how the model works in one specific classification task. In order to train all task at the same time, we have divided each data set into the same number of batches. Then, during the training, a batch of data is randomly selected and it is used to train its specific model (sharing the embedding and BiLSTM layers with other models). In this sense, we consider one epoch when all batches from all task were trained.

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

Second, the embedding layer transforms each element in the tokenized tweet into a low-dimension vector. The embedding layer, composed of the vocabulary of the task, was randomly initialized from a uniform distribution (between -0.8 and 0.8 values and with 100 dimensions). The initialized embedding layer was updated with the word vectors included in a pre-trained model from Regional Embeddings, which provides FastText word embeddings for Spanish language variations.

Then, a biLSTM layer gets high-level features from previous embeddings. A disadvantage of seq2seq models (such as LSTM) 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, 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. Finally, the tweet-level feature vector produced by the previous layers is used for classification task by two fully-connected (dense) layers.

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. The dropout rate was set to 0.5 in all cases.

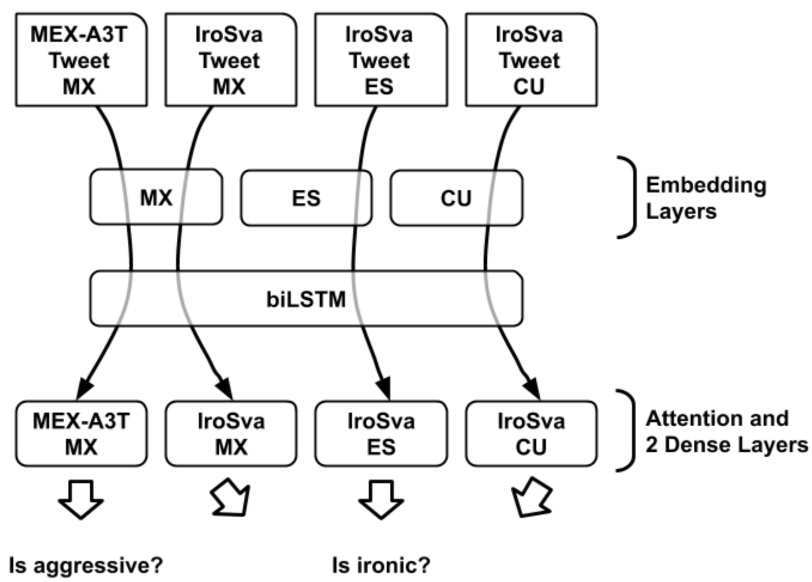

Fig. 1. Simplified schema of the model

4 Results

We generated 2 submissions per Spanish variant in this task. In the first submission, we trained 4 models at the same time: for IroSvA from Cuba, Spain, Mexico and for the MEX-A3T task (Figure 1). In the second one, we trained 10 models at the same time: three Spanish variants from the Irosva task, one from the MEX-A3T and one HAHA tasks, and five from the TASS task.

Results of our submissions for each task are given in the Table 1 together with the baselines proposed by the organizers.

Table 1. Results for different submissions and baselines in F-score

Result/Baseline	ES	MEX	CU	AVG
LASTUS-UPF-submission1	0.6606	0.5933	0.6017	0.6185
LASTUS-UPF-submission2	0.6493	0.6218	0.5737	0.6149
LDSE	0.6795	0.6608	0.6335	0.6579
W2V	0.6823	0.6271	0.6033	0.6376
Word nGrams	0.6696	0.6196	0.5684	0.6192
MAJORITY	0.4000	0.4000	0.4000	0.4000

5 Conclusion

In this paper, we have presented our results from the participation in the IroSvA task from the IberLEF 2019. We have investigated multi-task learning on neural networks with different tasks. Our results were close to the baselines presented by the organizers. As commented before, each Spanish variation data set from IroSva only includes 2,400 tweet for training, in comparison with other tasks (MEX-A3T or HAHA with 7,700 and 30,000 tweets, respectively). We divided each data set into the same number of batches, then, batches related to the IroSva task contain less information to train than batches from MEX-A3T or HAHA. In this case, the updates in the models produced by the IroSva batches could be diminished by other batches. In any case, our model related to each Spanish variant from IroSva and mostly trained with other tasks achieved results compared with the baselines, opening new research lines with the purpose of improving the relevance of the IroSva task during the multi-task training (e.g. data augmentation in the IroSva task). In addition, we want to test different types of neural networks (e.g. convolutions or combinations of convolutions and LSTM layers) and share more layers between task. Finally, we also consider that the integration of linguistic features (e.g. word frequency, POS tags and word shape) and metadata (e.g. whether a tweet is a response to another tweet) can represent useful contextual information to improve our performance.

References

1. Barbieri, F., Saggion, H.: Modelling irony in twitter. In: Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics. pp. 56–64. Association for Computational Linguistics, Gothenburg, Sweden (Apr 2014). https://doi.org/10.3115/v1/E14-3007
2. Cignarella, A.T., Frenda, S., Basile, V., Bosco, C., Patti, V., Rosso, P.: Overview of the EVALITA 2018 task on irony detection in italian tweets (ironita). In: Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech

Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018. (2018)
3. González-Ibánez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in twitter: a closer look. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2. pp. 581–586. Association for Computational Linguistics (2011)
4. Joshi, A., Bhattacharyya, P., Carman, M.J.: Automatic sarcasm detection: A survey. CoRR **abs/1602.03426** (2016)
5. Mut Altin, L.S., Bravo, A., Saggion, H.: Lastus/taln at semeval-2019 task 6: Identification and categorization of offensive language in social media with attention-based bi-lstm model. In: Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019). pp. 668–673. Association for Computational Linguistics, Minneapolis, Minnesota, USA (2019)
6. Ortega-Bueno, R., Rangel, F., Hernández Farías, D.I., Rosso, P., Montes-y-Gómez, M., Medina Pagola, J.E.: Overview of the Task on Irony Detection in Spanish Variants. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019), co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2019). CEUR-WS.org (2019)
7. Rangel, F., R.P.F.M.: A low dimensionality representation for language variety identification. In: Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing16). pp. pp. 156–169. Springer-Verlag, LNCS(9624), (2018)
8. Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., Huang, R.: Sarcasm as contrast between a positive sentiment and negative situation. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 704–714. Association for Computational Linguistics, Seattle, Washington, USA (Oct 2013)
9. Van Hee, C., Lefever, E., Hoste, V.: SemEval-2018 task 3: Irony detection in English tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 39–50. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). https://doi.org/10.18653/v1/S18-1005