SiDi-NLP-Team at IDPT2021: Irony Detection in Portuguese 2021

Almeida Neto, Antônio Manoel dos Santos[1,2,3], Osti, Bruna Almeida[1,2,3], Azevedo, Camila de Araujo.[1], Nobrega, Fernando Antonio Asevedo[1,3,4], Brum, Henrico Bertini[1,2,3], Peinado, Luis Heriberto Osinaga[1], Bittencourt, Marciele de Menezes[1,2,3], da Silva, Nataly Leopoldina Patti[1,2,3], and Cortes, Priscila Osorio[1]

*Authors sorted by first name.
[1]The author contribute on the hypothesis and features.
[2]The author contribute on the experiments.
[3]The author contribute on writing.
[4]The author contribute on internal company procedures.

SiDi, Campinas SP 13.098-321, BRA (https://www.sidi.org.br/)

Abstract. This paper presents the submission of SiDi-NLP team in IDPT 2021 - Irony Detection in Portuguese (IberLEF 2021). Irony detection is a challenging semantic task similar to other tasks such as Sentiment Analysis. Due to these similarities we performed experiments using algorithms that achieved state-of-the-art results for similar semantic tasks in Brazilian Portuguese with linguistic feature representation and pre-trained BERT models applied to the two shared task datasets – Tweet Dataset and News Dataset. The pre-trained BERT models outperformed the other classifiers achieving 1.00 accuracy and F1 in the Tweet Dataset, and 0.903 accuracy/0.900 in F1 for the News Dataset. We also discuss the results considering the results obtained in the shared task.

Keywords: Irony · Sarcasm.

1 Context and Background

The Irony Detection in Portuguese (IDPT 2021) [4] is a shared task co-allocated at IberLEF 2021 [11] and presents competitors with datasets for identification of irony documents in two different domains – tweets and news. Some authors describe the ability to recognizing ironic sentences by humans as "relatively easy way although not always" [8]. The main goal of the task is to extract a label for a document as *False* when the whole document does not contain irony and

IberLEF 2021, September 2021, Málaga, Spain.

True otherwise. Due to the subjectivity of the task, there are several similarities between Irony Detection and other Natural Language Processing tasks, such as Sentiment Analysis and Hate Speech Detection. The main challenges in these scenarios are that the irony clues usually relate to different pragmatic contexts such as the period in which the text was published, external information, relation between writer and readers, and others.

Since the use of pragmatic feature for machine learning algorithms usually comprehends complex processes of modeling, we handled the problem as a text classification problem based on features related to Sentiment Analysis methods (Section 2.2). This idea was based on the hypothesis that ironic texts would share linguistic features with opinion texts, due to the nature of both tasks. Furthermore, in order to compare the efficiency of these features against most recent approaches for Natural Language Processing, we also performed experiments with a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model that was built for the Portuguese Language – BERTimbau [13].

We have observed that our proposed linguistic feature has not outperformed the results of BERTimbau, at least, at the test dataset that was shared by the shared task organizers. It is important to note that in the official report by the organizers, we may see a kind of rotation between the results of the participants over the two datasets. It could be a suggestion of a huge difference between the data distribution or, probably, different approaches were used and have presented good results over distinct contexts.

2 Experimental Setup

2.1 Corpora

The two corpora used in this work contain texts on different topics written in Brazilian Portuguese language and are publicly available.

The first corpora contain 15,212 tweets extracted from dataset used in an irony and sarcasm detection work [1]. The authors collected potentially ironic tweets containing the hashtags "#ironia" or "#sarcasmo" posted between August 10, 2014 and August 6, 2017. Others non-ironic tweets were collected considering random tweets about economics, politics, and education that do not contain the hashtags #ironia or #sarcasmo. Additionally, the authors included tweets collected by de Freitas et. al. [5] that were manually annotated by Portuguese language experts. This final dataset has many more ironic sentences than non-ironic sentences, as shown the distribution of classes in fig1. This dataset was free of words and expressions that could serve as tips for the model and interfere in the learning, such as links or "rt", "#ironia" or "#sarcastico" tags.

The second corpora contains 18,494 news extracted from Sensacionalista, The Piauí Herald, and Estadão websites. The news were labeled according to the source website: news from Sensacionalista and The Piauí Herald are sarcastic and, therefore, were labeled as ironic; news from Estadão were labeled as non-ironic [10]. The distribution of classes is shown in Fig. 2.

2.2 Machine Learning Features

There are several similarities between Irony Detection and Sentiment Analysis due to the nature of the tasks. Following this intuition, we ran baseline experiments in the datasets using a well known text representation and pipelines that were proposed at [2].

The features presented in [2] are the same we used for our ML methods since they enable the classifiers to observe several features that may indicate the semantic alignment of the sentence. The representation is not decisive on the classification, but merely as input for classifiers. We describe all the features as follows:

- **Bag-of-Words (BoW):** a BoW representation of the data with absence or presence as 0 and 1 respectively;
- **Presence of negations:** in Sentiment Analysis, the negation presence usually indicates the inversion of a polarity. We used a list of Brazilian Portuguese negations such as "não" and "nunca" in order to keep this aspect in

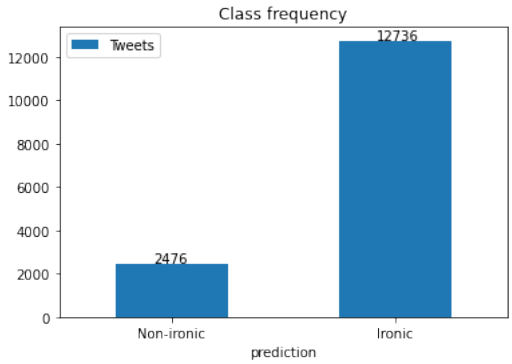

Figure 1: Frequency of Non-ironic and Ironic sentences

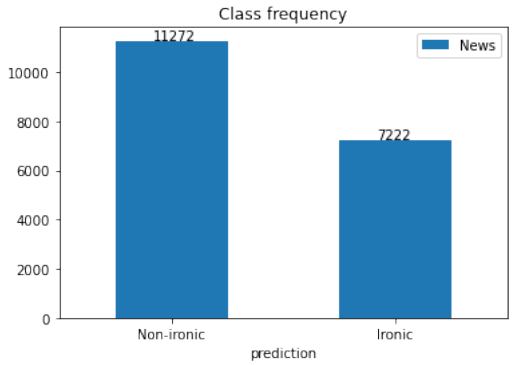

Figure 2: Frequency of Non-ironic and Ironic sentences

our experiments even though the feature must not be as important for Irony as it is for Sentiment Analysis;
- **Emoticons:** emoticon is a string that put together form a figure representation that contains semantic relevance for the sentence. We used a list of positive and negative emoticons to map the alignment of the token in the sentence;
- **Emojis:** similarly to the emoticons, we also represented the emojis in the documents. The main difference between them is that the emojis are alphanumeric characters that nowadays are interpreted by smartphones and browsers to be shown as figures. We also used a corpora with the polarity of each emoji in the document as a feature [9];
- **Sentiment Lexicon:** we also provided the ML methods with the count of positive and negative words in the sentence following Sentilex [12];
- **Part-of-Speech tagging:** we also counted the number of nouns, adverbs, verbs and adjectives using PoS tagged in NLPnet [7], the feature is specially relevant for identifying adjectives that are most frequent in opinion sentences and less frequent in factual information.

We used these features as input for five ML classifiers – a Support Vector Machine (SVM), a Logistic Regressor (LR), MultiLayer Perceptron (MLP), Random Forest (RF) and Naive Bayes (NB). We have chosen these methods considering [2] and the results obtained in Sentiment Analysis for Brazilian Portuguese tweets. The parameters were not grid-searched and are the same as the best fit the original work presents [3].

2.3 BERT and BERTimbau

BERTimbau[13] is a pre-trained BERT model trained on the Portuguese language [13]. BERT [6] is a Transformer encoder architecture that learns contextual relations between words in a text to generate a language model. BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. This pre-trained BERT model can be fine-tuned with just one additional output layer to create specific models for a wide range of tasks in NLP, without substantial architecture modifications [6].

To create a Portuguese version of BERT, the authors of BERTimbau used data from brWaC [14], the largest open Portuguese corpus, which contains 2.68 billion tokens from 3.53 million documents (web pages). They train BERTimbau models on two sizes: Base (12 layers, 768 hidden dimensions, 12 attention heads, and 110M parameters) and Large (24 layers, 1024 hidden dimensions, 16 attention heads and 330M parameters) [13].

In this work, we fine-tuned the base version of BERTimbau to classify irony sentences. The parameters were not grid-searched and are the same as in the original work.

3 Results

Since we have observed a big variation over the two datasets and that IDPT organizers also have indicated the submissions of the results separately, we will also report the results individually here.

Table 1 and Table 2 present the values of Accuracy and F-score obtained by each method in each dataset. To facilitate the comparison of the results, the best F-score for each dataset was highlighted in bold.

Table 1: Results obtained in Tweets dataset with hold-out validation.

Tweets dataset		
Method	Accuracy	F1
SVM	0.897	0.473
LR	0.897	0.473
MLP	0.9	0.474
RF	0.895	0.472
NB	0.867	0.509
lightgbm	0.99	**1.000**
Multilingual-BERT	1.000	**1.000**
BERTimbau	1.000	**1.000**

Table 2: Results obtained in News dataset with hold-out validation.

News dataset		
Method	Accuracy	F1
SVM	0.733	0.726
LR	0.767	0.764
MLP	0.787	0.783
RF	0.608	0.600
NB	0.557	0.538
lightgbm	0.700	0.720
Multilingual-BERT	0.823	0.819
BERTimbau	0.903	**0.900**

Multilingual-BERT and BERTimbau stood out positively in F-score and Accuracy in both datasets. These methods also correctly predicted all test samples from Tweets dataset. Analyzing this dataset, it was possible to notice that the non-ironic examples had a limited vocabulary related to economic news, and for that reason, it is possible to say the models were biased by training data. Since BERTimbau did better at predicting the News dataset, it was defined as a model to perform the final predictions.

4 Conclusion

It is interesting to note the teams that have participating of the IDPT 2021 have not presented the equivalent results over the two datasets from the competition. For instance, the submissions of the team "TeamBERT4Ever‘ have the highest results on news, but they have not performed well on dataset with tweets. The same behavior can be observed on our submission, in which our relative results over news are better than other ones. For instance, the models not based on deep learning approaches also have reached good results at the datasets with news than in the another one.

Acknowledgement

We acknowledge SiDi for all the support with the infrastructure for experiments and work environment for the development of the work.

References

1. Fabio Araujo da Silva. Detecção de ironia e sarcasmo em língua portuguesa: uma abordagem utilizando deep learning. Bachelor's thesis, Universidade Federal de Mato Grosso, February 2018.
2. Lucas V Avanço, Henrico B Brum, and MG Nunes. Improving opinion classifiers by combining different methods and resources. *XIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 25–36, 2016.
3. Henrico Bertini Brum and Maria das Graças Volpe Nunes. Semi-supervised sentiment annotation of large corpora. In *International Conference on Computational Processing of the Portuguese Language*, pages 385–395. Springer, 2018.
4. Ulisses Brisolara Corrêa, Leonardo Pereira dos Santos, Leonardo Coelho, and Larissa A. de Freitas. Overview of the IDPT Task on Irony Detection in Portuguese at IberLEF 2021. *Procesamiento del Lenguaje Natural*, 67, 2021.
5. Larissa Astrogildo de Freitas, Aline Aver Vanin, Denise Nauderer Hogetop, Marco Nemetz Bochernitsan, and Renata Vieira. Pathways for irony detection in tweets. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, SAC '14, page 628–633, New York, NY, USA, 2014. Association for Computing Machinery.
6. 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.
7. Erick R Fonseca, João Luís G Rosa, and Sandra Maria Aluísio. Evaluating word embeddings and a revised corpus for part-of-speech tagging in portuguese. *Journal of the Brazilian Computer Society*, 21(1):1–14, 2015.
8. Irazú Hernández-Farías, José-Miguel Benedí, and Paolo Rosso. Applying basic features from sentiment analysis for automatic irony detection. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 337–344. Springer, 2015.
9. Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.
10. Gabriel Schubert Marten and Larissa Astrogildo de Freitas. The construction of a corpus for detecting irony and sarcasm in portuguese. In *Proceedings of XVII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC-2020)*, pages 709–717, Rio Grande, Brazil, 2020.
11. Manuel Montes, Paolo Rosso, Julio Gonzalo, Ezra Aragón, Rodrigo Agerri, Miguel Ángel Álvarez Carmona, Elena Álvarez Mellado, Jorge Carrillo de Albornoz, Luis Chiruzzo, Larissa Freitas, Helena Gómez Adorno, Yoan Gutiérrez, Salud María Jiménez Zafra, Salvador Lima, Flor Miriam Plaza de Arco, and Mariona Taulé. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021)*. CEUR Workshop Proceedings, 2021.
12. Mário J Silva, Paula Carvalho, and Luís Sarmento. Building a sentiment lexicon for social judgement mining. In *International Conference on Computational Processing of the Portuguese Language*, pages 218–228. Springer, 2012.

13. Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. Bertimbau: Pretrained bert models for brazilian portuguese. In Ricardo Cerri and Ronaldo C. Prati, editors, *Intelligent Systems*, pages 403–417, Cham, 2020. Springer International Publishing.
14. Jorge A. Wagner Filho, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio. The brWaC corpus: A new open resource for Brazilian Portuguese. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).