GuillemGSubies at IDPT2021: Identifying Irony in Portuguese with BERT

Guillem García Subies[1]

Instituto de Ingeniería del Conocimiento, Francisco Tomás y Valiente st., 11 EPS, B Building, 5th floor UAM Cantoblanco. 28049 Madrid, Spain
guillem.garcia@iic.uam.es

Abstract. This paper describes a system created for the IDPT 2021 shared task, framed within the IberLEF 2021 workshop. We present an approach mainly based on fine-tuned BERT models using a Grid-Search and Data Augmentation with MLM substitution. Our models far outperform the baselines and achieve results close to to the state-of-the-art.

Keywords: Irony Detection · BERT · Transformers · Data Augmentation · BERTimbau

1 Introduction

Although irony can be relatively easy to identify for humans, it is not so easy to detect for NLP models [5], mainly because the information can be implicit and usually doesn't use the literal meaning of the words used. This makes the task of irony detection prefect to evaluate the evolution of NLP systems.

The IDPT (Irony Detection in Portuguese) shared task proposes, during this third edition of the IberLEF [15] workshop, a corpus to detect irony in tweets and news written in Portuguese [1]. This article summarizes our participation in all the IDPT tasks.

Given the success of Transformer-inspired language models [23], both in academia and industry [24], we decided to use already pre-trained BERT [6] models. Furthermore, their ability to understand contextual information can be very useful for the irony detection task. Specifically, we will use BERTimbau [21] with hyperparameters Grid-Search. To address the problem of small data, we will use Data Augmentation techniques.

1.1 Task Description

There are two corpora, one for tweets and one for news (task1 and task2 respectively). For both of them, the problem is binary classification, where the sample can be ironic or not.

IberLEF 2021, September 2021, Málaga, Spain.

The tweets corpus has 15212 tweets and the news one has 18494 news for their train splits. The data is collected from various preexisting sources [8] [4] [14]. Then, the test splits are composed of 300 tweets and 300 news gathered and annotated by the organizers of the tasks. This will help to create models that generalize very well.

The metric used to evaluate the results is the balanced accuracy. This is mainly because both datasets are very unbalanced as we can see in the table 1. It is also notable the difference between tasks; most of the tweets are ironic while most of the news are not ironic.

Tweets		**News**	
Class	**Nº of samples**	**Class**	**Nº of samples**
ironic	12736	ironic	7222
not ironic	2476	not ironic	11272

Table 1. Distribution of Samples

1.2 Goals

This work is focused on proving that it is possible to use open source resources and relatively small language models (compared to the newest models like GPT-3 [2]) to obtain state-of-the-art results. Specifically, the main goal is to obtain a Portuguese language model that can detect the irony in the text and meet the requirements explained before.

1.3 Summary of the proposal

To achieve the goals explained above we will fine-tune a BERTimbau model [21] with Grid-search optimized parameters. Along with this model, the data is first preprocessed with simple heuristics and then, augmented with Masked Language Model word masking.

In the next section, we will briefly see some previous work related to this topic. Then, in Section 3, we will explain the main ideas behind the proposed models and the experiments we did. In Section 4 we will present a summary of the results we got. Finally, in Section 5 we will expose the main conclusions of our work and results, and we will also propose some ideas for future work.

2 Related Work

There is an extensive bibliography on Sentiment Analysis and irony detection in social media given the high scientific interest in solving such a difficult problem. Some early attempts to create corpora in this field were for sarcasm detection, for instance Davidov et al. used Amazon Mechanical Turk to create a corpus with 5.9 million tweets [5] and Riloff et al. explore the identification of sarcastic tweets that have a positive word or comment followed by an undesirable situation [20].

There have also been some attempts to create irony detection corpora in other languages than English. For example, Ptacek et al. [19] created a Czech sarcasm binary classification dataset for tweets and also propose a n-gram and heuristics based embeddings that are feed intro classic machine learning models. Liebrecht et al. [12] collect a Dutch sarcasm dataset from tweets that included the hashtag *#sarcasm* and hypothesize about that hashtag being the digital equivalent of non-verbal expressions in live interactions. Bilal et al. [10] collect irony detection datasets in different languages in order to show that good models can be trained even when the data for some language is scarce.

Following this trend, there have been a lot of irony detection competitions these last years. For instance, IDAT [9] proposed a binary classification problem to detect irony in tweets written in Arabic. The best model was a feature based one with classic machine learning models, outperforming even BERT models. IroSvA, [16], proposed a binary irony classification problem for Spanish tweets in different Spanish dialects. This time, the best model fed a Word2Vec embeddings into a Transformer model as a weights initialization.

For the Portuguese language, Carvalho et al. [3] detect irony in newspaper comments using simple glossaries, proving that complex linguistic features do not work for irony. Following the same trend Freitas et al. [7] create a list of relevant patterns to detect irony in Portuguese tweets.

It is notable that some of the models used in these works still use linguistic features and heuristics to detect the irony. However, we will focus on the potential of language models to solve this task without any linguistic features.

3 Models

3.1 Data Preprocessing

We performed a simple preprocessing where we substituted some expressions with a more normalized form:

- Every URL was replaced with the token "[URL]", so we don't get strange tokens when the tokenizer tries to process a URL. Furthermore, no semantic information about irony can be inferred from a URL, the only information relevant for the model is that there is a URL in that token.
- The hashtag characters ("#") were deleted ("#example" → "example") because the base language models we will use, are trained in generic text and might not understand their meaning. Furthermore, most of the hashtags are used the same way as normal words.
- We replaced every username with the generic token "[USER]" because the exact name of a user does not really add any information about the irony. The only relevant feature is knowing if someone was mentioned or not, but not who.
- Finally, we normalized every laugh ("jasjajajajj" → "haha"), so we minimize the noise of the misspellings, common in social networks.

3.2 Baselines

We created some baselines, so we can compare our models properly. We selected a HashingVectorizer + RandomForest. This way, we can compare our models to a classic feature extraction model.

3.3 Language Models

We used BERTimbau [21], a Portuguese BERT model that outperforms mBERT and the previous state-of-the-art. Specifically, we used the large model, *bert-large-portuguese-cased*. For the fine-tuning process, we carried out a Grid-search optimization over the main parameters of the neural network: learning rate, batch size and dropout rate. The search was performed with a 5-fold stratified cross-validation with the following grid: Learning rate, $(1e-6, 1e-5, 3e-5, 5e-5, 1e-4)$; batch size, $(8, 16, 32)$ and dropout rate, $(0.08, 0.1, 0.12)$. The best parameters for both models were: learning rate, $1e-5$; batch size, 16 and dropout rate, 0.1.

3.4 Data Augmentation

As the dataset is relatively small, we decided to run Data Augmentation techniques. The selected strategy was the Data Augmentation through the masking of words with a Masked Language Model, BERTimbau. For every sample in the dataset, we randomly masked 15% of the tokens and used BERTimbau to predict them, creating a modified sample. With this method, we obtained double the amount of the original samples.

4 Experiments and Results

4.1 Experimental Setup

The software we used was Python3.8, transformers 4.5.1 [24], pytorch 1.8.1 [17], scikit-learn 0.24.1 [18] and nlpaug 1.1.3 [13].

4.2 Results

In the Table 2 we can see the results for our models in the test set of the first task. Our runs for this task are *BERTimbau* and *BERTimbau-aug*, without data augmentation and with data augmentation, respectively as explained in Section 3.3.

We can see that the language model far outperforms classic methods like hashing tricks and a random forest. We can also see that, although the Data Augmentation does not provide a great performance boost, it is still useful in order to have better models. All in all, our models obtain great results given their simplicity, proving that finding the right parameters for the model is crucial for optimizing the performance. These results are placed fourth among all the participating teams, which proves that our approach, given it's simplicity and the lack any linguistic analysis, is very good.

Model	bacc
HV+RF	0.3316
BERTimbau	0.4912
BERTimbau-aug	0.5000

Table 2. Results for task1

For the second task, the results were not as good as the ones obtained in the first task. In the Table 3 we can look at them in more detail. It looks like BERTimbau did not behave so well with the news dataset.

Model	bacc
HV+RF	0.5423
BERTimbau	0.7804
BERTimbau-aug	0.7858

Table 3. Results for task2

5 Conclusions and Future Work

Through this shared task, we have seen that NLP can be of great help in detecting irony from natural language in social networks and there is still a long way to go. The results obtained by our systems are very promising given their great performance and their simplicity. This compilation of methods is very significant because it could lead to much better results when combined with other improvements from the state-of-the-art. Particularly, the Data Augmentation approach with the Grid-search have proven to work really well in this context. We therefore consider that we have achieved our goals for this shared task.

However, we believe that our results could improve a lot using specific language models trained only with corpora from social networks. Another interesting approach would be to use a general language model and further pre-train it with corpora from the same domain [22] as the final task. Finally, we have proven that good hyperparameters are also key for a good neural network, so a better search, like the Population Based Training [11], would further improve the model.

Acknowledgments

This work has been partially funded by the Instituto de Ingeniería del Conocimiento (IIC) and the hardware used was also provided by the IIC.

References

1. Brisolara Corrêa, U., Pereira dos Santos, L., Coelho, L., A. de Freitas, L.: Overview of the idpt task on irony detection in portuguese at iberlef 2021. Procesamiento del Lenguaje Natural **67** (2021)
2. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners (2020)
3. Carvalho, P., Sarmento, L., Silva, M.J., De Oliveira, E.: Clues for detecting irony in user-generated contents: oh...!! it's" so easy";-. In: Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion. pp. 53–56 (2009)
4. DA SILVA, F.R.A.: Detecção de ironia e sarcasmo em língua portuguesa: Uma abordagem utilizando deep learning (2018)
5. Davidov, D., Tsur, O., Rappoport, A.: Semi-supervised recognition of sarcastic sentences in twitter and amazon. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning. p. 107–116. CoNLL '10, Association for Computational Linguistics, USA (2010)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2018)
7. Freitas, L., Vanin, A., Vieira, R., Bochernitsan, M.: Some clues on irony detection in tweets. In: WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web (05 2013). https://doi.org/10.1145/2487788.2488012
8. de Freitas, L.A., Vanin, A.A., Hogetop, D.N., Bochernitsan, M.N., Vieira, R.: Pathways for irony detection in tweets. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing. pp. 628–633 (2014)
9. Ghanem, B., Karoui, J., Benamara, F., Moriceau, V., Rosso, P.: Idat at fire2019: Overview of the track on irony detection in arabic tweets. In: Proceedings of the 11th Forum for Information Retrieval Evaluation. pp. 10–13 (2019)
10. Ghanem, B., Karoui, J., Benamara, F., Rosso, P., Moriceau, V.: Irony detection in a multilingual context. In: Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F. (eds.) Advances in Information Retrieval. pp. 141–149. Springer International Publishing, Cham (2020)
11. Jaderberg, M., Dalibard, V., Osindero, S., Czarnecki, W.M., Donahue, J., Razavi, A., Vinyals, O., Green, T., Dunning, I., Simonyan, K., Fernando, C., Kavukcuoglu, K.: Population based training of neural networks (2017)
12. Liebrecht, C., Kunneman, F., van den Bosch, A.: The perfect solution for detecting sarcasm in tweets #not. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 29–37. Association for Computational Linguistics, Atlanta, Georgia (Jun 2013), https://www.aclweb.org/anthology/W13-1605
13. Ma, E.: Nlp augmentation. https://github.com/makcedward/nlpaug (2019)
14. Marten, G.S., de Freitas, L.A.: The construction of a corpus for detecting irony and sarcasm in portuguese. Brazilian Journal of Development **7**(5), 47973–47984 (2021)
15. Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Ángel Álvarez Carmona, M., Álvarez Mellado, E., de Albornoz, J.C., Chiruzzo, L., Freitas, L., Adorno, H.G.,

Gutiérrez, Y., Zafra, S.M.J., Lima, S., de Arco, F.M.P., Taulé, M.: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021). In: CEUR Workshop Proceedings (2021)
16. Ortega-Bueno, R., Rangel, F., Hernández Farıas, D., 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. vol. 2421, pp. 229–256 (2019)
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
19. Ptáček, T., Habernal, I., Hong, J.: Sarcasm detection on Czech and English Twitter. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. pp. 213–223. Dublin City University and Association for Computational Linguistics, Dublin, Ireland (Aug 2014), https://www.aclweb.org/anthology/C14-1022
20. 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), https://www.aclweb.org/anthology/D13-1066
21. Souza, F., Nogueira, R., Lotufo, R.: BERTimbau: pretrained BERT models for Brazilian Portuguese. In: 9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear) (2020)
22. Sun, C., Qiu, X., Xu, Y., Huang, X.: How to fine-tune bert for text classification? (2020)
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need (2017)
24. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 38–45. Association for Computational Linguistics, Online (Oct 2020), https://www.aclweb.org/anthology/2020.emnlp-demos.6