Exploring BERT for Aspect Extraction in Portuguese Language

Emerson Lopes
Federal University of Pelotas
Pelotas, Brazil

Ulisses B. Corrêa
Sul-Rio-Grandense Federal Institute of
Education, Science, and Technology
Charqueadas, Brazil

Larissa A. de Freitas
Federal University of Pelotas
Pelotas, Brazil

Abstract

Sentiment Analysis is the computer science field that comprises techniques that aim to automatically extract opinions from texts. Usually, these techniques assign a Sentiment Orientation to the whole document (Document Level Sentiment Analysis). But a document can express sentiment about several aspects of an entity. Methods that extract those aspects, paired with the sentiment about them, operate in the Aspect Level. Aspect-Based Sentiment Analysis approaches can be split into two stages: Aspect Extraction and Aspect Sentiment Classification. The literature presents works mainly focused on reviews about hotels, smartphones, or restaurants. In this work, we present an approach for Aspect Extraction based on Multilingual (Google's) and Portuguese (BERTimbau) BERT pre-trained models. Our experiments show that Aspect Extraction based on BERT pre-trained for Portuguese achieved Balanced Accuracy of up to 93% on a corpus of reviews about the accommodation sector.

Introduction

According to Liu (2015), people are influenced by other's opinions. Even when simple decisions have to be taken individuals want to be backed up by the previous experience of others.

Opinions and sentiments towards a product or topic also are valuable sources of information for companies, that can use those opinions to take business decisions (Liu, 2015).

Due to the increase of User-Generated Content (UGC), manually classifying those reviews became very hard, thus generating the need for ways to automatically extract useful information from this UGC data (Liu, 2012).

Sentiment Analysis (SA) is the Computer Science field that solve this problem. Sentiment Analysis studies the opinions, sentiments, attitudes, and emotions of people about entities and their attributes in texts (Liu, 2015). Sentiment Classification can be done in different levels. Freitas (2015) argues that SA literature is mainly focused at three levels of granularity: document level, sentence level, and aspect level (Freitas, 2015).

Although most works focus on Document Level, it is clear that to assign a single sentiment orientation to the whole review may be not enough to express the experience of the individual about a product or service. A review can present different sentiments towards parts of the same entity. A guest can complains about the quality of the food at the hotel's restaurant, but can be satisfied with the service.

Liu (2012) proposes to analyze sentiment presented in a document towards each aspect, part or attribute, of the target entity. This kind of approach is called Aspect-Based Sentiment Analysis (ABSA). Aspect Extraction (AE) is the first part of the ABSA. It is a key task, as errors in this step generate errors in the next step (Aspect Sentiment Classification) (Pereira, 2020). This phase consists of the identification of the parts of the text that represent entity's aspects.

BERT is a self-supervised methodology to generate pre-trained language models proposed by Devlin et al. (2019). It is trained on a large amount of raw text, avoiding the need of previews text annotation. These models allow fine-tuning to specific Natural Language Processing (NLP) tasks obtaining good results, despite demanding less annotated data. It was released originally in English and Multilingual versions (Devlin et al., 2019), and since then it has been trained and released for different languages, including Portuguese (Souza, Nogueira, and Lotufo, 2020).

In this paper, we analyze the performance of BERT models on AE applied to the Portuguese language. Our experiments are conducted over a corpus of hotel reviews annotated in aspect level (Freitas, 2015).

The remainder of this paper is organized as follows. In the next section, we show the background. In the "Related Works" section, we describe some works that use BERT for the Aspect Extraction task, and some works that use different methods focused on the Portuguese language. In the "Methodology" section, we discuss the dataset format, the preprocessing done, and how our experiments are made. In the "Results" sections, we show the parameters used and our results. Finally, in the "Conclusion and Future Works" section, we present our conclusion and future works.

Background

Sentiment Analysis

Sentiment Analysis is the Computer Science field that analyze the opinions expressed towards some entity in a text, and classify the sentiment orientation of this opinion (Liu, 2015).

Opinion Formal Definition

Opinions are formally defined as a quintuple (Liu, 2015):

$$Opinion = (e, a, s, h, t) \quad (1)$$

where:

- e – entity: an entity may be a product or topic;
- a – aspect of an entity: an aspect is a set of attributes or components of an entity;
- s – the sentiment orientation of opinion holder at opinion time about an aspect of an entity – a sentiment orientation indicates whether the opinion is positive, negative or neutral;
- h – opinion holder: an opinion holder may be a person expressing the opinion;
- t – opinion time: an opinion time is a date.

Levels of Sentiment Analysis

On the document level, each document is considered to have only one sentiment, usually classified as positive, negative, or neutral. On the sentence level, there is one sentiment in each sentence of the document. On the aspect level, it is needed to find which aspects of an entity exist in the text, and determine the sentiment for each aspect.

Aspect-based Sentiment Analysis

Aspect Based Sentiment Analysis (ABSA) is a level of granularity of Sentiment Analysis that focuses not only on classifying the sentiment orientation of an opinion towards an entity but also on what aspect of that entity this sentiment is expressed. This level of granularity consists mainly of two tasks: Aspect Extraction (AE) and Aspect Sentiment Classification.

Aspect Extraction This task consists of determining which aspects of some entity are being considered in some text.

For example: in the sentence "Hotel com boa **localização**" ["Hotel with good **localization**"], the aspect **'localização'** [**'localization'**] should be identified.

Aspect Sentiment Classification This task consists of determining the sentiment orientation for each aspect in some text.

For example: "Hotel com boa **localização**" ["Hotel with good **localization**"], the sentiment orientation about an aspect **'localização'** [**'localization'**] should be classified as **positive**.

Related Works

Since the release of BERT, many works have used it to solve NLP tasks. For AE or End-to-End ABSA we can cite works of Hoang, Bihorac, and Rouces (2019) and Sun, Huang, and Qiu (2019), that model the problem as a sentence pair classification task, using an auxiliary sentence; Xu et al. (2019) use extra post-training with unsupervised domain data, showing great improvement on performance. All of these works use English datasets.

In other languages, Yanuar and Shiramatsu (2020) use BERT for AE, classifying five different aspects in the Indonesian language. Winatmoko, Septiandri, and Sutiono (2020) also use BERT for AE on the Indonesian dataset.

To the best of our knowledge, there are no works that use BERT for ABSA on the Portuguese dataset. For this task on Portuguese, other approaches have been taken. Aires et al. (2018) used one approach in which a SVM classifier was trained for each aspect, and a similar one, one in which a LSTM network was trained for each aspect. Freitas and Vieira (2019) use an ontology based approach to the problem.

Methodology

Dataset

Our experimental results were obtained using a *corpus* of aspect-level annotated hotel reviews written in Portuguese (Freitas and Vieira, 2015). The *corpus* contains 194 reviews about 10 hotels. Reviews were collected from TripAdvisor[1] and manually annotated in aspect level. Each aspect referred has a sentiment assigned to it. A total of 17 unique aspects were detected.

Proposed Approach

In our proposed approach we used the Sentence Pair Classifier from BERT to predict if the aspect is related to the text, using the review and the aspect as inputs, and labels 'related' and 'unrelated'. As the dataset contains only the aspects that are present in each review, we annotated all other aspects as 'unrelated', and for the aspects that were already annotated, the label was changed from the polarity to the 'related' label. By doing this, the dataset became very unbalanced. To compensate for that, we use weighted labels, using the implementation of Tayyar Madabushi, Kochkina, and Castelle (2019).

The models were trained using the pre-trained models 'base-multilingual-cased' (Devlin et al., 2019) and 'base-portuguese-cased' (Souza, Nogueira, and Lotufo, 2020). Usually, the 'cased' models get worse results, however, we decided to use it as there are no Portuguese uncased models available. As the models used are the cased versions, there is no need for any preprocessing for converting characters to lowercase, so all preprocessing done was fixing spelling errors.

We test a few scenarios, with different hyperparameters, and with a preprocessed dataset, and another dataset without any preprocessing. Every experiment was made on both datasets. These hyperparameters were based on the recommendations from Devlin et al. (2019), with some changes, especially in the number of epochs, to try to obtain better results.

To evaluate the performance of the models, we use 10-fold cross validation, and measure the Precision, F1-Measure,

[1] http://www.tripadvisor.com.br

Model	Epoch	Precision (%)	Acc (%)	F1-Measure (%)	BAcc (%)
Portuguese	4	**90.41**	87.99	75.02	81.39
Portuguese	8	90.15	**91.25**	**83.75**	**87.47**
Multilingual	4	89.16	81.91	57.27	70.45
Multilingual	8	83.02	85.79	71.98	79.35

Table 1: Results of Portuguese and Multilingual w/ preprocessing.

Max Sequence Length	Epochs	Precision (%)	Acc (%)	F1-Measure (%)	BAcc (%)
512	8	90.22	**94.73**	**90.97**	**93.85**
512	4	**92.15**	93.69	88.64	91.25
256	8	90.41	93.91	89.42	92.34
256	4	90.47	92.17	85.61	89.03

Table 2: Results of Max Sequence Length w/ preprocessing.

Model	Epoch	Precision (%)	Acc (%)	F1-Measure (%)	BAcc (%)
Portuguese	4	**92.36**	90.45	81.36	85.29
Portuguese	8	88.30	**91.78**	**85.32**	**89.11**
Multilingual	4	90.69	80.09	48.80	66.82
Multilingual	8	86.24	85.97	71.35	78.73

Table 3: Results of Portuguese and Multilingual without preprocessing.

Max Sequence Length	Epoch	Precision (%)	Acc (%)	F1-Measure (%)	BAcc (%)
512	8	91.21	**94.5**	**90.38**	**93.07**
512	4	90.1	93.11	87.84	90.97
256	8	89.25	93.63	89.09	92.32
256	4	**91.76**	92.19	85.44	88.65

Table 4: Results of Max Sequence Length without preprocessing.

Accuracy (Acc). Still, as the data is unbalanced, we also measure the Balanced Accuracy (BAcc).

In Table 5 we show the hyperparameters used for the tests that compare the Portuguese and the Multilingual models. The Learning Rate and Class Weights are the same for all the tests. In the tests that use Max Sequence Length, this parameter and the Batch Size are changed. For all tables, the tests were made with 4 and 8 Epochs.

Parameter	Value
Learning Rate	2e-5
Batch Size	16
Max Sequence Length	128
Class Weights	20.1

Table 5: Hyperparameters of tests used in Table 1 and 3.

Results

In Table 1 we compare the results of training, with the Portuguese and Multilingual models, during 4 and 8 Epochs.

In Table 3 we compare the results of training with greater Max Sequence Length, using the same Learning Rate, Class Weights and Epochs, Sequence Length of 256 and 512, and Batch Size of 6 and 3, respectively.

The results of Table 2 show that the Portuguese pre-trained models can get very good results, 87.47% of BAcc, even with a smaller Max Sequence Length. This also shows that the pre-training focused on one language get significantly better results than the Multilingual version, even for languages without many resources, like Portuguese.

The results of Table 3 show that, for this dataset, a Max Sequence Length results in better performance. This is dependent on the size of sentences on the dataset.

In Table 4 and 5 we repeat the tests, but using the dataset without any preprocessing, to compare the efficiency of the preprocessing on these tests. We can see that the preprocessing had almost no influence on the results.

Conclusion and Future Works

These results show that BERT can work well for AE in a Portuguese language dataset of hotel reviews, especially with models pre-trained for the language.

As future work, we plan to test how BERT works on Aspect Sentiment Classification, and in a full ABSA system, for the Portuguese language. We also intend to investigate the performance on other dataset written in Portuguese and manually annotated in aspect-level.

References

Aires, J. P.; Padilha, C.; Quevedo, C.; and Meneguzzi, F. 2018. A deep learning approach to classify aspect-level sentiment using small datasets. In *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Freitas, L., and Vieira, R. 2015. Exploring resources for sentiment analysis in portuguese language. In *Proceedings of the 4th Brazilian Conference on Intelligent Systems, BRACIS, Natal, Brazil, November 04-07*.

Freitas, L., and Vieira, R. 2019. Ontology-based feature-level sentiment analysis in Portuguese reviews. *International Journal of Business Information Systems* 32(1):30–55.

Freitas, L. 2015. *Feature-level sentiment analysis applied to brazilian portuguese reviews*. Ph.D. Dissertation, Programa de Pós-Graduação em Ciência da Computação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre. 2015.

Hoang, M.; Bihorac, O. A.; and Rouces, J. 2019. Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 187–196. Turku, Finland: Linköping University Electronic Press.

Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Pereira, D. A. 2020. A survey of sentiment analysis in the Portuguese language. *Artificial Intelligence Review* 1–29.

Souza, F.; Nogueira, R.; and Lotufo, R. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *Proceedings of the 9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23*.

Sun, C.; Huang, L.; and Qiu, X. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 380–385. Minneapolis, Minnesota: Association for Computational Linguistics.

Tayyar Madabushi, H.; Kochkina, E.; and Castelle, M. 2019. Cost-sensitive BERT for generalisable sentence classification on imbalanced data. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, 125–134. Hong Kong, China: Association for Computational Linguistics.

Winatmoko, Y. A.; Septiandri, A. A.; and Sutiono, A. P. 2020. Aspect and opinion term extraction for hotel reviews using transfer learning and auxiliary labels.

Xu, H.; Liu, B.; Shu, L.; and Yu, P. S. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis.

Yanuar, M. R., and Shiramatsu, S. 2020. Aspect extraction for tourist spot review in indonesian language using bert. In *Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 298–302.