Abstract
Aspect-Based Sentiment Analysis (ABSA) is a Natural Language Processing (NLP) task that extracts referred aspects from text and assigns polarities to opinions about those aspects. Most research on ABSA focuses on English. Only a few ABSA works deal with the Portuguese language. In this work, we used BERTimbau to create a Question-Answer approach to ABSA in Portuguese. First, we post-trained this model with text from the same domain as our target corpus. Then, we constructed an auxiliary sentence from the aspect and converted ABSA to a sentence-pair classification task, such as question answering (QA) and natural language inference (NLI). Our experiments show that ABSA based on BERT for Portuguese achieved Balanced Accuracy (BACC) of 77% on a corpus of reviews about the accommodation sector using a post-trained model with a QA approach.

Introduction
Sentiment Analysis (SA) is the field in Natural Language Processing (NLP) that automatically analyzes people’s sentiments or opinions towards some entity. These sentiments can be valuable sources of information about the consumer’s feelings about a particular product or idea, which can help in decisions by companies or governments (Liu, 2015).

Due to the significant increase in User-Generated Content (UGC), this automatic extraction and analysis of the sentiments became necessary, as manually analyzing them cannot cover this amount of generated data (Liu, 2012).

Sentiment Analysis can be done on different levels, focusing, mainly, on three possible granularity levels: document level, sentence level, and aspect level (Freitas, 2015).

We need Aspect-based Sentiment Analysis (ABSA) to tackle more complex sentences. This level of granularity makes it possible to analyze different opinions held towards different aspects of some entity or different entities in the same document or sentence.

Bidirectional Encoder Representations from Transformers (BERT) is a self-supervised methodology to generate pre-trained language models (Devlin et al., 2019). It can be pre-trained on large amounts of raw text and fine-tuned to more specific NLP tasks, obtaining good results with less labeled data. The seminal work of Devlin et al. (2019) produced two kinds of models. Multilingual models were made to support up to 104 languages, while the English language received particular attention with specific models. Approaches based on those English models rapidly became state-of-the-art in several NLP tasks. BERT’s performance aroused researchers’ attention to apply BERT in languages other than English. Currently, several works focus on developing BERT language models to specific languages, such as BERTimbau for Portuguese (Souza, Nogueira, and Lotufo, 2020) and BETO for Spanish (Cánerete et al., 2020).

This paper aims to develop a new solution for the aspect sentiment classification sub-task of ABSA, based on BERT and focused on Portuguese, by using auxiliary sentences. In our experiments, we used a corpus of hotel reviews, annotated in aspect level (Freitas and Vieira, 2015). We also test the efficiency of a post-training step on a small corpus Corrêa (2021) on the same domain.

Background
Sentiment Analysis
SA is the NLP field that aims to analyze the opinions expressed towards some entity in a text and classify the orientation of the sentiment on this opinion (Liu, 2015). These opinions are formally defined as a quintuple (Liu, 2015):

\[ \text{Opinion} = (e, a, s, h, t) \]  

where:
- \(e\) - entity: the target entity (may be a product or topic);
- \(a\) - aspect: the aspect of the entity \(e\) in which the opinion was given;
- \(s\) - sentiment: the sentiment orientation of the opinion - it indicates whether the opinion is positive, negative, or neutral;
- \(h\) - holder: the opinion holder (may be a user expressing the opinion);
- \(t\) - time: an opinion time is a date.

SA approaches can be split due to the level of granularity used in the analysis. Typically, SA works focus on three granularity levels: document, sentence, and aspect.

On the document level, it’s assumed that each document expresses opinions towards a single entity, from a single
opinion holder, without considering the different aspects of that entity.

Sentence-level approaches are similar to document-level ones, where sentences can be seen as documents, maintaining the aforementioned restrictions. It can handle more complex documents, where each sentence expresses an opinion towards some entity but still cannot handle multiple opinions, entities, or aspects in one sentence.

For ABSA, each document can have multiple opinions expressed towards different aspects of some entity.

**Aspect-based Sentiment Analysis**

On this granularity level, all opinions expressed towards any aspect of the entity are analyzed individually. This level allows a better understanding of the opinions and entities in the text. To accomplish the analysis on this level, the task used to be broken into two sub-tasks: Aspect Extraction (AE) and Aspect Sentiment Classification (ASC).

**Aspect Extraction** This task aims to determine which aspects of a given entity are considered in a given text.

**Aspect Sentiment Classification** This task consists of the classification of the polarity for each aspect that has been identified in the text.

For example, in the sentence “Hotel com boa localização” [“Hotel with good localization”], the goal of AE would be to identify the aspect ‘localização’ ['localization'], and the goal of ASC would be to classify this aspect as positive.

**BERT**

BERT is a self-supervised methodology for the pre-training of language models. It allows the training on a large corpus of unlabeled data, allowing fine-tuning on specific tasks with better results. The pre-training step consists of Masked Language Modeling (MLM) and Next Sentence Prediction (NSP).

**Masked Language Modeling** In the first task, some tokens of the input sentence are transformed into a [MASK] token or another random token from the vocabulary, and the task is to predict the original token.

**Next Sentence Prediction** In the second task, given two sentences A and B, this task consists of predicting if B is the sentence that follows A on the corpus, or is random.

The fine-tuning step is the step that trains the model for the final task. This step requires labeled data.

An extra step in this process is called post-training. This step consists of improving BERT to a specific domain.

It works like pre-training, as it uses the same tasks (or just one of them) and uses a corpus of unlabeled data. The main difference is that this step uses a corpus of the same domain of the final task (in this paper, a larger corpus of hotel reviews developed by López Barbosa (2015)).

**Related Works**

For ABSA, on English datasets, we can cite the works of Hoang, Bihorac, and Rouces (2019), that use an auxiliary sentence, modeling the problem as a sentence pair classification task, and the work of Sun, Huang, and Qiu (2019) that uses a similar approach like ours modeling the problem as question answering (QA) and natural language inference (NLI).

Xu et al. (2019) uses post-training with a large amount of unlabeled text from the same domain, showing performance improvement.

Freitas (2015) does ABSA in a dataset of Portuguese reviews from the accommodation sector, collected from TripAdvisor. The proposed methodology consists of four main parts: i) A preprocessing step is applied to the reviews. ii) A domain ontology is used to construct a hierarchy of explicit and implicit aspects from the preprocessed reviews. These aspects will be passed as features in the next step. iii) Several configurations using combinations of linguistic rules and sentiment lexicons are applied to determine the polarity classification of each aspect. iv) A summary of the aspects and their polarities is generated. The results of this method are bound to the quality of the resources used.

Corrêa (2021) conducted an extensive analysis of the applicability of Character-level Convolutional Neural Networks (CharCNN) for solving the ABSA task for Portuguese texts and constructing a new corpus for this task, composed of reviews of the accommodation sector. In his experiments, the best model for determining the ternary sentiment orientation (positive, negative, and neutral) achieved 65% of BACC. Moreover, data augmentation techniques such as back-translation were applied but with no gain in performance.

Aires et al. (2018) proposed two architectures to address the ABSA task. The first architecture comprises two models, one for AE and the other for ASC. The sentence is passed to the AE model, and if an aspect is identified, it passes sentence to the ASC model that gives an overall classification of the sentence polarity. This classification is considered the sentiment of the aspect. The second architecture is also composed of two parts (AE and ASC) but the second stage has separated models for each one of the possible present aspects of the reviews. Their experiments are conducted in three different datasets of user reviews, with ten annotated aspects related to cellphones.

Lopes, Corrêa, and Freitas (2021) use the BERT model in the same dataset as Freitas (2015); with the AE task. They use a Portuguese pre-trained BERT model (BERTimbau) to predict whether a given aspect is related or not to a given text. Using a multilingual BERT for comparison, the article presented that a language-focused BERT gets significantly better results, with 83.75% F1-score, 11.77% better than the multilingual model.

The work of Barros and De Bona (2021) presents a Deep Learning based framework for the Aspect Triplet Extraction (ASTE) task. This framework consists of 4 modules: sentence encoder, opinion-aspect representation, opinion-aspect tagging, and dependency parsing. It works as follows: The sentence encoder generates word vectors in the first module, which encode semantics and context. In the second module, opinion and context attributes are extracted. The output of the second module serves as input for the opinion-
aspect tagging module, which makes the classification between 3 classes: opinion, aspect, or none of the options. The last module, dependency parsing, checks the dependency between aspect and sentiment. Data from hotel and book reviews were used to carry out the experiments. The experiments demonstrated that the presented framework outperforms the baseline models for ASTE tasks.

Essebbar et al. (2021) present an ABSA using French pre-trained models (PTM): mBERT, Camem-BERT, and FlaubERT. In addition, three fine-tuning techniques were used: PTM-FC (Fully Connected), PTM-SPC (Sentence-Pair Classification) and PTM-AEN (Attention Encoded Network). The datasets of reviews of restaurants and museums, which are part of the SemEval2016 database, were used to carry out the experiments. The results demonstrate that the fine-tuning models achieve a better result than the state-of-the-art models. Experiments were also carried out with Out-Of-Domain data, which consists of museums review data, and the models used in this work also present a better result in this particular task.

Methodology

Datasets

The dataset used for training the model consists of hotel reviews written in Portuguese extracted from TripAdvisor. Aspects were manually labeled on a total of 194 reviews. The reviews talk about ten different hotels (Freitas and Vieira, 2015). Each aspect has an assigned sentiment to it, within three possibilities: negative (-1), neutral (0), and positive (+1). In this corpus, there are annotations about 17 different aspects.

For post-training, the dataset developed by Corrêa (2021) was used, which is a snippet of the dataset present in López Barbosa (2015). This dataset consists of 8067 Portuguese hotel reviews from TripAdvisor about New York, Las Vegas, and Paris hotels.

Proposed Approach

Our proposed approach was to use a pre-trained BERT model on BrWaC (Wagner et al., 2018), a large corpus in Portuguese, and train it using our dataset for ABSA tasks. The model receives two inputs for those tasks, the review and an auxiliary sentence, each one written in different formats.

QA-M: Returns whether that sentence is Positive, Negative or Neutral, using as auxiliary sentence a question, for example, for the review "Os quartos são confortáveis" ("The bedrooms are comfortable") the auxiliary sentence would be "Qual é a polaridade de quarto?" ("What is the polarity of bedroom?") the output should be "Positive".

QA-B: Returns False or True, and the auxiliary sentence is a question about a specific polarity of the aspect, thus generating three questions as auxiliary sentences for each aspect. For example, using the same review, "A polaridade de quarto é Positiva?" ("The polarity of bedroom is Positive?") is the auxiliary sentence and the output is "True".

NLI-M: It has a simpler architecture as an auxiliary sentence, looking like a pseudo-sentence, and returns whether that aspect is Positive, Negative, or Neutral. For example, using the same review, "Quarto" ("Bedroom") is the auxiliary sentence, and the output is "Positive".

NLI-B: Same as QA-B, but the auxiliary sentences are pseudo-sentences, only delivering the necessary information and returning True or False. For example, using the same review, the auxiliary sentence is "Quarto - Positivo" ("Bedroom - Positive") and the output is "True".

Even though the BERT model being trained in several NLP tasks, it has no domain in some fields, as is the case of opinion texts. After all, it was trained using data from Wikipedia, Book Corpus, and BrWaC. This ends up creating a problem in training, as there is little data to remove this bias from non-opinion texts, and to get around this, Xu et al. (2019) proposes the use of a method called post-training. This method consists of performing a new training with a different dataset, thus providing more examples of opinion texts to the model, consecutively decreasing bias.

In the tests, the learning rate and class weights hyperparameters remained fixed, but an exploration of sequence length and batch size was performed, in addition to testing models with 4 and 8 epochs. Due to the unbalanced data, we used F1-score and BACC metrics as evaluation, as they are metrics more sensitive to this kind of data.

Results

In Table 1 we show the base results for our experiments with each of the auxiliary sentence formats, without post-training, getting results of up to 70% of F1-score, using the QA-B auxiliary sentences.

The results shown in Table 2 compare the results of post-training on QA-B auxiliary sentences. The model fine-tuned with 8 epochs and no post-training achieved a slight increase in performance compared to four epochs, and, when experimented on a post-trained model, with just 5k and 10k steps of post-training, the results increased to 77% of F1-score.

After 10k post-training steps, the results get less stable (still being better than no post-training). This could be related to the small size of the post-training dataset, and future works with even more steps or with a more extensive dataset could help clarify.

Final Remarks and Further Works

These results show that the combination of transforming the ABSA task to a sentence-pair classification task and, more importantly, the post-training step can significantly improve the results of BERT, even when done with few examples of the end task domain.

For future works, we aim to test similar approaches for end-to-end ABSA and conduct more experiments to help clarify why the post-training past 10k steps have led to less stable results and consequently worse performance than fewer steps.

Acknowledgments

We thank CNPq Proc. Nº 117331/2021-0 and FAPERGS Proc. Nº 21/2551-0000974-8 for the financial support.
Table 1: Comparison of the Auxiliary Sentences.

| Type of Sentence | Epoch | Precision | Accuracy | Recall | F1   | BACC |
|------------------|-------|-----------|----------|--------|------|------|
| NLI-B            | 4     | 0.74      | 0.75     | 0.67   | 0.67 | 0.67 |
| NLI-M            | 4     | 0.59      | 0.56     | 0.56   | 0.55 | 0.56 |
| QA-B             | 4     | 0.76      | 0.77     | 0.70   | 0.70 | 0.70 |
| QA-M             | 4     | 0.61      | 0.58     | 0.58   | 0.58 | 0.58 |

Table 2: Comparison of the Post-Training.

| Model      | Epoch | Precision | Accuracy | Recall | F1   | BACC |
|------------|-------|-----------|----------|--------|------|------|
| Base       | 8     | 0.73      | 0.78     | 0.73   | 0.72 | 0.73 |
| 5k Steps   | 8     | 0.77      | 0.80     | 0.76   | 0.77 | 0.76 |
| 10k Steps  | 8     | 0.78      | 0.80     | 0.77   | 0.77 | 0.77 |
| 20k Steps  | 8     | 0.73      | 0.76     | 0.72   | 0.72 | 0.72 |
| 30k Steps  | 8     | 0.75      | 0.78     | 0.74   | 0.75 | 0.74 |

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).

Barros, J. M., and De Bona, G. 2021. A deep learning approach for aspect sentiment triplet extraction in portuguese. In Britto, A., and Valdivia Delgado, K., eds., Intelligent Systems, 343–358. Cham: Springer International Publishing.

Cañete, J.; Chaperon, G.; Fuentes, R.; Ho, J.-H.; Kang, H.; and Pérez, J. 2020. Spanish pre-trained bert model and evaluation data. In Practical ML for Developing Countries Workshop.

Corrêa, U. B. 2021. Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa. Ph.D. Dissertation, Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Pelotas, Pelotas.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. 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), 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Essebbar, A.; Kane, B.; Guinaudeau, O.; Chiesa, V.; Quénel, I.; and Chau, S. 2021. Aspect based sentiment analysis using french pre-trained models. In Proceedings of the 13th International Conference on Agents and Artificial Intelligence, Volume 1, 519–525.

Freitas, L., and Vieira, R. 2015. Exploring resources for sentiment analysis in portuguese language. In Proceedings of the 4th Brazilian Conference on Intelligent Systems.

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.

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.

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

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

Lopes, E.; Corrêa, U.; and Freitas, L. 2021. Exploring bert for aspect extraction in portuguese language. The International FLAIRS Conference Proceedings.

López Barbosa, R. R. 2015. Aplicación del análisis de sentimientos a la evaluación de datos generados en medios sociales. Ph.D. Dissertation, Programa de Doctorado en Ingeniería de la Información y del Conocimiento, Universidad de Alcalá, Alcalá.

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.

Xu, H.; Liu, B.; Shu, L.; and Yu, P. 2019. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, 2324–2335. Minneapolis, Minnesota: Association for Computational Linguistics.