UFSCar's Team at ABSAPT 2022: Using Syntax, Semantics and Context for Solving the Tasks

Fernanda Malheiros Assi[1,†], Gabriel Barbosa Candido[2,†], Lucas Nildaimon dos Santos Silva[1,†], Diego Furtado Silva[1,†] and Helena de Medeiros Caseli[1,†]

[1] *Computing Department, Federal University of São Carlos (UFSCar), Rod. Washington Luís, km 235, São Carlos - SP, Brazil, CEP 13565-905*

[2] *Electrical Engineering Department, Federal University of São Carlos (UFSCar), Rod. Washington Luís, km 235, São Carlos - SP, Brazil, CEP 13565-905*

Abstract

With the growth of social networks and commercial sites, an increasing number of users have shared their opinions and life experiences on websites. Aspect-Base Sentiment Analysis (ABSA) is a field dedicated to studying opinions and feelings expressed in text format, aiming to detect the polarity of feelings and relate them to entities. To achieve this goal, this task can be divided into two sub-tasks: Aspect Term Extraction and Sentiment Orientation Extraction. The first task aims to identify the aspect within a text review, while the Sentiment Orientation Extraction aims to infer the polarity of a text review concerning a single aspect. For the first one, spaCy was used as an approach for data pre-processing, tokenization, and feature extraction. Then the aspect terms were identified and extracted. For the second one, the approach consisted of inserting the relevant portion of the text review into the GoEmotions model and taking the top-3 predictions. If the top-1 prediction score was higher than a threshold, we would assign that polarity to the respective aspect. Otherwise, we would assign the top-2 and top-3 predictions, if they have the same polarity, to the respective aspect. In both cases, good results were achieved, yielding third place in the ABSAPT 2022 competition.

Keywords

Sentiment Analysis, Aspect Term Extraction, Sentiment Orientation Extraction

1. Introduction

When customers purchase a product or service, they often leave their opinion on the company's website as a review. These reviews can provide a sense of what to expect from a business

IberLEF 2022, September 2022, A Coruña, Spain

† These authors contributed equally.

fernanda.malheiros@estudante.ufscar.br (F. M. Assi); gcandido@estudante.ufscar.br (G. B. Candido); lucas.silva@estudante.ufscar.br (L. N. d. S. Silva); diegofs@ufscar.br (D. F. Silva); helenacaseli@ufscar.br (H. d. M. Caseli)

https://sites.google.com/site/lalicufscar/pessoas/fernanda-malheiros-assi (F. M. Assi); https://sites.google.com/site/lalicufscar/pessoas/lucas-nildaimon-dos-santos-silva (L. N. d. S. Silva); https://dc.ufscar.br/~diego (D. F. Silva); https://sites.google.com/site/lalicufscar/pessoas/helena-caseli (H. d. M. Caseli)

0000-0001-8820-964X (F. M. Assi); 0000-0002-2488-0087 (G. B. Candido); 0000-0002-5606-9416 (L. N. d. S. Silva); 0000-0002-5184-9413 (D. F. Silva); 0000-0003-3996-8599 (H. d. M. Caseli)

regarding quality and customer service to potential buyers. They can also help a company identify areas that might need improvement. These concerns are the subject of research on Sentiment Analysis and, more specifically, in Aspect-Based Sentiment Analysis (ABSA). In ABSA, instead of assigning a sentiment to the entire review, the aspects (entities, attributes) are identified and a sentiment is associated with each one. Therefore, ABSA is a valuable tool for companies since it automatically identifies and tracks the sentiment around specific topics. This helps businesses monitor and respond to negative emotions quickly and effectively.

Although millions of people speak Portuguese, there is a shortage of approaches and research available for ABSA in the Portuguese language. It is due to the lack of corpora written and annotated in Portuguese. Aiming to improve this scenario, professors and undergraduate students of the Pelotas Federal University (UFPel) and the University of São Paulo (USP), Brazil, proposed a shared task for ABSA in Portuguese (ABSAPT 2022) [1].

For this, they used the corpora previously developed by Freitas [2] and Corrêa [3], composed of Portuguese reviews about accommodation service companies published on the TripAdvisor website. The ABSA task was divided into two independent sub-tasks:

1. **Aspect Term Extraction (Task 1)** - Identifying and extracting aspects of a review;
2. **Sentiment Orientation Extraction (Task 2)** - Extract the sentiment orientation (polarity) of the review about a single aspect.

This paper will describe the approaches our team used to solve both tasks. We organized it as follows. Section 2 summarises some of the related works. The train and test datasets available to us are presented in section 3. Sections 4 and 5 explain our solutions for tasks 1 and 2, respectively. In section 6 we show the results achieved by both tasks. Finally, section 7 finishes the paper with the conclusion.

2. Related Work

Searching for inspiration in other bibliographical references, this article aims to solve, the two sub-tasks, based on [4] and [5].

In [4], the experiments were conducted using the filtered version of the GoEmotions dataset with 54,263 manually annotated sentences in 28 classes (27 emotions + neutral), extracted from the Reddit forum in English and then translated via the itranslate library. The paper analyzes the fine-tuning performance of the BERTimbau-base and BERTimbau-large models in the emotions classification task. It uses the Class Balanced Loss (CB) method to address the dataset imbalance issue, proven by the domination of the neutral class. Compared to the results provided by the authors of the GoEmotions dataset, this work achieved better performance, attributed to the robustness against imbalance. Based on it, we would extract the relevant text review as an input of the BERT model, already pre-trained with the GoEmotions dataset, classifying it into 27 emotions mapped to positive, negative, and neutral polarities.

In [5], begins with collecting information related to product reviews from e-commerce platforms to create a dataset. After the collection phase, data preprocessing, tokenization, feature extraction, classification, aspects extraction, and grouping classified review aspect wise was applied. In preprocessing, data cleaning was performed, stop-words removal, lemmatization, and tokenization. TF-IDF (Term Frequency - Inverse Document Frequency) transformation was used for the feature extraction. Finally, the algorithms used for review classification were support vector machine, logistic regression, and artificial neural network, besides the best result considering the metrics of precision and recall, the logistic regression with l1 regularization.

In [6], the authors developed a Question-Answer method for ABSA in Portuguese using BERTimbau. For this, they used a manually annotated collection of hotel reviews taken from TripAdvisor as their dataset. Then, for each review, they constructed auxiliaries sentences from the aspect, each one written in a different format. After that, they inputted the review and the auxiliary sentence into the pre-trained BERT model. They also used the dataset developed by Correa [3] to post-train their model. As a result, they obtained a bacc score of 0.77 with the post-trained model.

Based on the work described, to identify the aspect and the opinion in the text review, we use as NLP feature, POS (parts of speech) tag, and dependency parsing, as well as the lemmatization process to generate features for each token in the text review.

3. Corpora

For the training and test datasets, the task committee used corpora previously developed by Freitas [2] and Côrrea [3]. The training dataset comprises both Freitas and Côrrea corpora, while the test dataset contains only examples of the Corrêa's corpus. The annotation of aspects on both datasets followed the guidelines described in [2], and the aspects used were the concepts on the Accommodation Services Domain Ontology, HOntology [7]. In Table 1 it is possible to see some instances from the training corpus.

3.1. Training Corpus

The training dataset is the same for both tasks and is a CSV file with 3111 instances from 847 reviews about the accommodation sector. A single review can hold multiple sentiments from different aspects; however, each line consists of a single annotated aspect mentioned in a review. The columns presented in the dataset are:

- **id**: a sequential integer number indicating the id of the review. It goes from 0 to 3110;
- **review**: the complete review posted on TripAdvisor;
- **polarity**: the sentiment associated to the annotated aspect. It can be -1 for negative emotion, 0 for neutral and 1 for positive sentiment;
- **aspect**: the annotated aspect;
- **start_position**: an integer number indicating the start position of the aspect in the review;
- **end_position**: an integer number indicating the end position of the aspect in the review.

Of the 3111 instances, 2112 (67.9%) of them refer to aspects with positive sentiment, 527 (16.9%) have a negative emotion, and 472 (15.2%) reviews are labeled as neutral. Also, there were 77 different aspects identified, the most common being "hotel" (hotel), "quarto" (room) and "localização" (location). Table 1 shows the first three lines of the training dataset. We highlighted in bold the aspect annotated in each instance.

Table 1
First three lines of the training dataset

id	review	polarity	aspect	start	end
0	O hotel é perto de todos os pontos principais de Paris, cercado de estações de metrô, dá para fazer tudo a pé. O Hotel é charmoso, aconselho ficar nos **quartos** no piso superior, perto de telhado, típica arquitetura parisiense. O café da manhã é simples, mas na medida certa, tudo muito gostoso, croissant sensacional! O preço é um pouco salgado, mas se puder vale muito a pena. *The hotel is close to all the main points of Paris, surrounded by metro stations, you can do everything on foot. The Hotel is charming, I recommend staying in the* ***rooms*** *on the top floor, close to the roof, typical Parisian architecture. The breakfast is simple, but in the right measure, everything is very tasty, the croissant is sensational! The price is a little steep, but if you can, it's worth it.*	0	quarto *room*	152	158
1	Viajamos eu e minha irmã. O hotel tem uma extrutura excelente! **Elevadores** modernos e rápidos. O quarto bem climatizado e roupas de cama novas e limpas. O café da manhã farto e variado. O fato de ser ao lado da gare Montparnasse facilitou a ida a Versailles, pois pudemos pegar um trem lá. O ponto negativo é que o metrô não é muito perto, (6 quadras grandes) e a caminhada no frio muito angustiante! Mas retornaria com certeza! *I traveled with my sister. The hotel has an excellent structure!* ***Elevators*** *are modern and fast. The room was well air conditioned and the bedding was new and clean. The breakfast was plentiful and varied. The fact that it is next to the Montparnasse station made it easier to go to Versailles, as we were able to catch a train there. The negative point is that the subway is not very close, (6 big blocks) and the walk is in the very distressing cold! But would definitely return!*	1	elevador *elevator*	63	71
2	Estive por 8 dias hospedado neste hotel com minha familia e posso afirmar com todas as letras que este hotel cumpre o que promete em todos os quesitos oferecidos, e além de tudo é um dos único que oferecem um **café da manhã** aos moldes dos hotéis no Brasil, incluso na diária, sem contar que na recepção tem café, chocolate e biscoitos o dia inteiro..muita cortesia. Voltarei com certeza *I stayed at this hotel for 8 days with my family and I can honestly say that this hotel fulfills what it promises in all the aspects it offers, and in addition it is one of the only ones that offer a* ***breakfast*** *in the standards of the hotels in Brazil, included in the daily rate, not to mention that the reception has coffee, chocolate and cookies all day long..very complimentary. I will come back for sure*	1	café da manhã *breakfast*	209	222

3.2. Test Corpus

Each Task has its own test dataset, which is a CSV file. For task 1, the dataset has 257 instances, and each file line has two columns: id and review. The goal is to find a list of aspects for each review.

For task 2, the dataset consists of 686 instances. As in the training dataset, each review has an annotation for a single aspect, and each line has the following columns: id, review, aspect, start_position and end_position. The goal is to find the polarity (1, 0 or -1) of the annotated aspect for each review.

4. Aspect Term Extraction

An aspect can be defined as a concept for which the opinion is expressed in a given document [8]. Aspect Term Extraction (ATE) is a subtask of ABSA that aims to identify and extract aspects from a document.

In this section, we describe the proposed approach for extracting the aspect terms. Figure 1 shows the pipeline for identifying and extracting aspects.

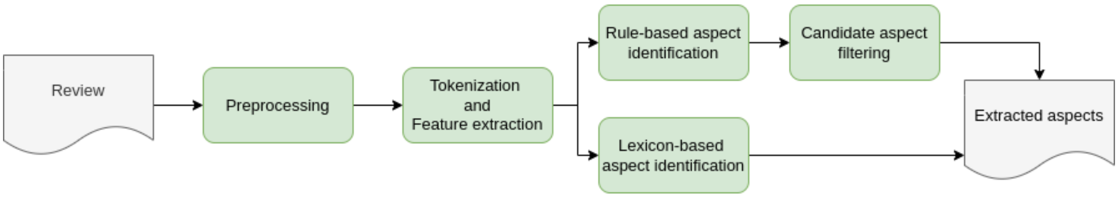

Figure 1: Pipeline for Aspect Term Extraction

The first step is data preprocessing. In this step, we begin by converting the review text to lowercase and removing all punctuation and special characters from it. Table 2 presents an example of our preprocessing method results.

Table 2
Example of preprocessing for review text

Original text	Preprocessed text
Hotel muito ruim, cheiro de mofo. Já no check-in fui informado que a lavanderia atrasou na entrega de toalhas!! Última opção... *Very bad hotel, musty smell. I was informed at check-in that the laundry service was late in delivering towels! Last option...*	hotel muito ruim cheiro de mofo ja no check in fui informado que a lavanderia atrasou na entrega de toalhas ultima opcao *very bad hotel musty smell i was informed at check in that the laundry service was late in delivering towels last option*

The second step involves tokenization and feature extraction. We used the large version of the Portuguese model from [9] to accomplish both tasks. First, the review text is tokenized, then linguistic features are extracted. In this work, we used dependency parsing (DEP), part-of-speech tagging (POS), and lemmatization to generate features for each token in the review text. Figure 2 shows an example of preprocessing and feature extraction for a review text. After preprocessing, we generate a quadruplet for each token in the review text, containing the token itself, its POS tag, DEP tag, and lemma.

The third step consists of the identification and extraction of aspect terms. First, we built a lexicon to identify aspects in the domain of accommodation services using the annotated aspects of the training and test corpora. To do so, we iterate over a review's tokens, checking if the token or its lemma is included in the aspects lexicon; if it is, we consider this token an aspect term of the provided review. In addition, we defined rules for identifying candidate aspects using the extracted linguistic features. Table 3 presents our rules for identifying a candidate aspect. Every noun in a review associated with a ROOT, NSUBJ, or OBL dependency tag is considered a candidate aspect. After applying these rules, we use HOntology [7] to filter the candidate aspects by extracting only those that are present in the ontology. Figure 2 shows how our second rule, that is, a noun with an NSUBJ dependency tag, identifies the aspect "lavanderia" (laundry). The aspect lexicon, on the other hand, identifies the aspects "hotel" and "toalha" (towel).

After identifying and extracting the aspects, we restore them to their proper surface forms, including punctuation.

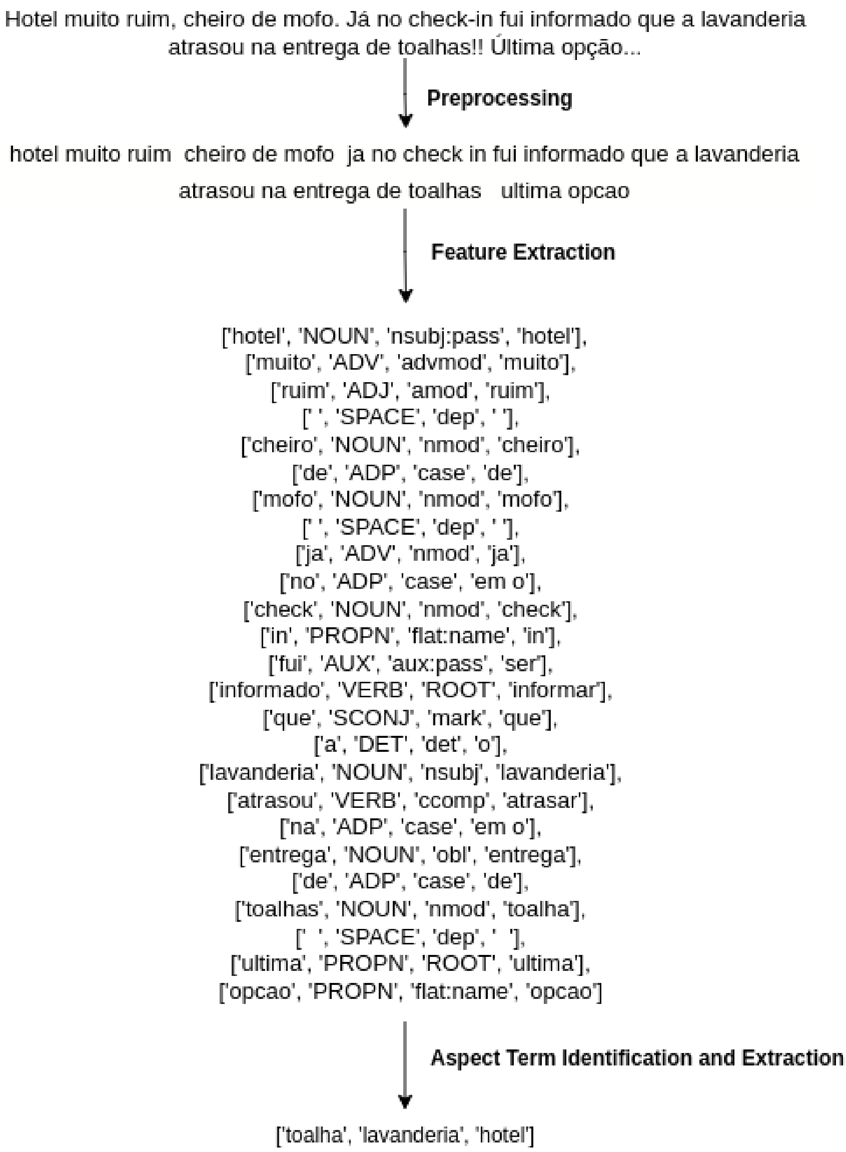

Figure 2: Example of preprocessing and feature extraction for a review text

Table 3
Rules for identifying candidate aspects

POS Tag + DEP Tag
NOUN + ROOT
NOUN + NSUBJ
NOUN + OBL

5. Sentiment Orientation Extraction

This section describes the process we used to extract the sentiment orientation of each aspect of a review and the results obtained. In this work, sentiment orientation can be described as the polarity (positive, negative, or neutral feeling) that an aspect express. Thus, our main goal is to correctly identify the polarity of an annotated aspect of a review. The methodology we used to achieve this objective is divided into two main steps: meaningful surroundings extraction (section 5.1) and sentiment extraction (section 5.2).

Figure 3 shows the pipeline we used for this task. We first used SpaCy [9] to help us find the meaningful surroundings; we could then input these surroundings into the pre-trained GoEmotions [4] for the Portuguese model. We then generated a polarity for each aspect based on the predictions and scores given by GoEmotions. Finally, we assigned polarity 0 to all aspects whose DEP tag in the SpaCy output was labeled as OBL.

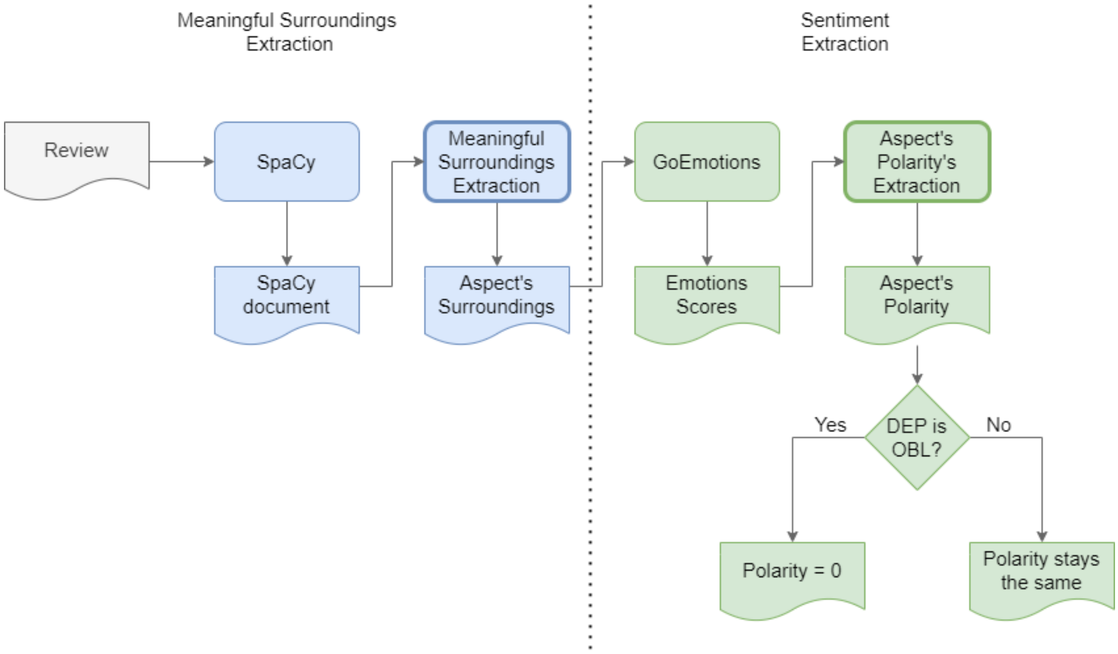

Figure 3: Pipeline for Sentiment Orientation Extraction

5.1. Meaningful Surroundings Extraction

An essential part of ABSA is identifying which part of the text is relevant to define the polarity of a given aspect. For example, in the following sentence: *"The internet is good, but the room was dirty"*, if we take the aspect *"internet"*, its polarity is positive. Notice that the second part of the sentence *"but the room was dirty"* is irrelevant for determining the emotion associated with *"internet"* and it carries a negative polarity. Therefore, we should not always take the entire sentence to identify the emotion of an aspect since it might contain noise.

Thus, in our work, we defined a meaningful surrounding as a text fragment containing words that relate directly or indirectly to the aspect in question. This meaningful surrounding should be long enough to identify the emotion associated with the aspect but not so long that it contains noise. To identify the meaningful surrounding of a review's aspect, we first used the Python library SpaCy [9] to extract the dependency parsing of the review. Then, we took all the tokens related to the aspect (their children and head) up to the second degree.

Figure 4 shows an example of part of a Dependency Tree provided by SpaCy for the aspect "quarto" *(room)*. To find the meaningful surrounding for this aspect, we first take all the tokens directly linked with "quarto", which are all the arrows that leave (children) and enter (head) it. With this, we get "ficar" *(staying)*, "nos" *(in)* and "piso" *(floor)*. In the next step, we do the same for all his children, getting the tokens "no" *(on)* and "superior" *(top)*. The last step is to put all these tokens together in the same order they appear in the original sentence. The final result is "ficar nos quartos no piso superior" *(staying in the rooms on the top floor)*.

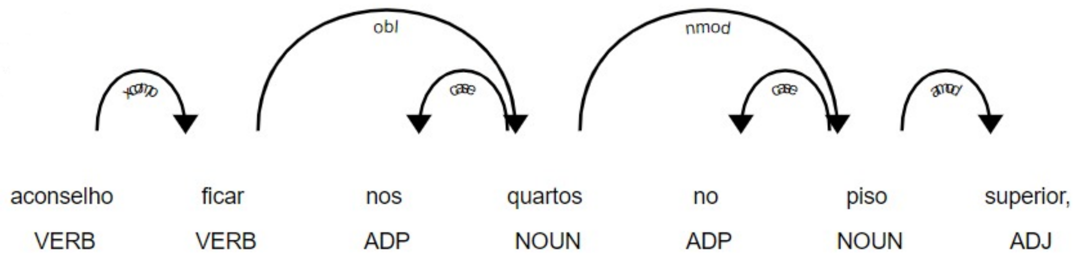

Figure 4: Example of a Dependency Tree

Table 4 shows three examples of the aspect sentence and the meaningful surrounding obtained. We highlighted in bold the meaningful surrounding in each instance.

5.2. Sentiment Extraction

In our approach, we used the Brazilian version of GoEmotions to identify the aspect sentiment in a review. GoEmotions for Portuguese [4] is a fine-tuning of the BERTimbau-base and BERTimbau-large models for the Sentence Classification task. The dataset used was the GoEmotions dataset [10], which was translated into Portuguese using automatic translation tools.

As output, the GoEmotions model gives a list of 27 emotions plus the Neutral one, and a score for each emotion. The higher the score an emotion gets, the more likely the text is to express that emotion. So, to identify the emotion of an annotated aspect, we first mapped each emotion into having a positive, negative, or neutral feeling. Table 5 shows the mapping of the GoEmotions emotions we did for this task.

We then took the top-3 predicted emotions in the scope of meaningful surrounding. If the top-1 prediction score was higher than a threshold, we associated its corresponding polarity to the aspect; otherwise, the assigned polarity was calculated based on the polarity associated

Table 4
Examples of Aspect Sentence and their Meaningful surroundings

Aspect	Aspect Sentence	Meaningful Surrounding
quartos *rooms*	O Hotel é charmoso, aconselho **ficar nos quartos no piso superior**, perto de telhado, típica arquitetura parisiense. *The Hotel is charming, I recommend* ***staying in the rooms on the top floor****, close to the roof, typical Parisian architecture.*	ficar nos quartos no piso superior *staying in the rooms on the top floor*
internet *internet*	O quarto e o banheiro sao espaçosos, novos, limpos **e a internet funciona muito bem**. *The room and bathroom are spacious, new, clean* ***and the internet works very well****.*	e a internet funciona muito bem *and the internet works very well*
localização *location*	**A localização é excelente, local privilegiado para compras** e para pontos agitados da cidade. ***The location is excellent, privileged place for shopping*** *and for the hot spots of the city.*	A localização é excelente, local privilegiado compras. *The location is excellent, privileged place shopping*

Table 5
Mapping of the GoEmotions for Portuguese Emotions

Positivo **Positive**	admiração, diversão, aprovação, zelo, curiosidade, desejo, entusisasmo, gratidão, alegria, amor, otimismo, orgulho, alívio, surpresa *admiration, amusement, approval, caring, curiosity, desire, excitement, gratitude, joy, love, optimism, pride, relief, surprise*
Negativo **Negative**	raiva, aborrecimento, confusão, decepção, desaprovação, nojo, constrangimento, medo, luto, nervosismo, remorso, tristeza *anger, annoyance, confusion, disappointment, desapproval, disgust, embarrassment, fear, grief, nervoussness, remorse, sadness*
Neutro **Neutral**	neutro, percepção *neutral, realization*

with the majority of the predictions. If a 3-way tied were to occur, the polarity would receive the top-1 prediction value. Therefore, for example, for a threshold of 0.8, if top-1 prediction was positive with a score of 0.5 and both predictions 2 and 3 were negative, the polarity would be negative.

Finally, we assigned polarity neutral for all those aspects that the SpaCy output labeled the DEP as being OBL. The OBL relation occurs when a nominal (noun, pronoun, and noun phrase) functions as an oblique argument or adjunct. It corresponds to an adverbial attaching to a verb, adjective, or other adverbs. When doing a manual analysis, we found that, in general, these aspects do not have any positive or negative polarity since they are not the main subject of the review's sentence.

To evaluate this proposed approach, we run some tests in the training dataset.

5.3. Testing Our Proposed Approach in the Training Dataset

The GoEmotions model we used to extract the sentiment associated with an aspect is already trained for this assignment. Therefore, we could use the training dataset to perform experimentations, using different techniques and approaches for the task, and then compare the results to decide which to apply to the test dataset.

Thus, in our experiments, we tested:

- inputting the whole aspect sentence versus only the aspect's meaningful surroundings to the GoEmotions model;
- calculating the polarity of the aspect in three different ways: (1) by assigning the value of prediction 1, (2) by selecting the value of the majority of the top-3 predictions, and (3) by combining these two approaches and setting the value of prediction 1 only if it was higher than a threshold;
- setting polarity 0 to all aspects that SpaCy tagged as OBL compared to not doing this post-processing.

Table 6 shows the results of the Balanced Accuracy Scores (Bacc) for the Meaningful Surroundings approach, and Table 7 shows the results for the Aspect Sentence approach. We can see that, as expected, the model predicts better when given only a short text fragment (Table 6), with the most important words related to the aspect, than the whole sentence (Table 7).

Table 6
Bacc scores on the train dataset for Meaningful Surroundings

	Not Annul OBL Aspects	Anull OBL Aspects
Prediction 1	0.59	0.59
Top 3 with treshold	0.64	**0.63**
Top 3 without treshold	0.62	0.63

Table 7
Bacc scores on the train dataset for Aspect Sentence

	Not Annul OBL Aspects	Anull OBL Aspects
Prediction 1	0.58	0.59
Top 3 with treshold	0.59	0.61
Top 3 without treshold	0.59	0.61

Besides that, the best way to calculate the aspect polarity is to take the majority of the top-3 predictions only if the score of prediction 1 is lower than the threshold. Lastly, the performance in the training dataset is mostly higher when setting the polarities of all the aspects whose DEP tag was OBL to 0.

Although the results of the investigated approaches are very close, it is worth mentioning that we did not run statistical tests to evaluate the significance of the difference between them. As we had to choose just one approach to send the results for the shared task, we decided to choose the one which determines the aspects' polarity based on the meaningful surrounding, taking into account the majority of the top-3 predictions only if the top-1 prediction is lower than the threshold of 0.9 and setting the polarities of all the aspects whose DEP tag was OBL to 0.

6. Results

This section reports the results of our approaches on Tasks 1 and 2 in the test dataset and also the final results sent to us by the Shared Task organizers.

6.1. Task 1 results

Table 8 shows the test corpus accuracy of our Aspect Term Extraction approach, as well as the final task outcomes reported by the shared task organisers.

Table 8
Results for Aspect Term Extraction

Dataset	Accuracy
Test	0.71
Final result	0.59

Our method is capable of identifying and extracting aspects, including domain multi-word aspects, such as *café da manhã* (breakfast) and *ar condicionado* (air conditioner).

Since the gold standard for the final results was not made available, it was not possible to carry on an error analysis.

6.1.1. Comparison With Previous Work

In [11], the authors used pre-trained BERT models to do a Sentence Pair Classifier to predict if a given aspect is related or not to the text. To do so, they labelled every annotated aspect presented in the review as “related” and every other aspect as “unrelated”. They also tested different hyperparameters and preprocessing approaches for the dataset. Their best result was achieved by the BERT model trained for Portuguese without preprocessing, with an Accuracy of 0.92.

In this work, we used a rule-based approach for Aspect Term Extraction, which is pretty much simpler and thus less costly. Although we achieved satisfactory results for this task, the previous work’s machine learning approach using pre-trained BERT models had a better performance.

6.2. Task 2 results

The approach we chose to send the results of Task 2 was the one we input the meaningful surroundings to the GoEmotions model, took the top-3 predictions with threshold, and annul the OBL aspects.

Table 9 shows the results provided by the ABSAPT 2022 committee for this task as a final result. The result obtained in the test dataset is compatible with what we found in the training dataset.

Table 9
Results for Sentiment Orientation Extraction

Dataset	Bacc	F1	Precision	Recall
Final result	0.63	0.61	0.65	0.63

6.2.1. Comparison With Previous Works

Over the years, researchers have used different approaches to solve the ABSA problem for Portuguese. They have, amongst others, used Character-level Convolutional Neural Networks (CharCNN) (bacc score of 0.65) [3], developed a question-answer method (bacc score of 0.77) [6], and used several combinations of linguistic rules and sentiment lexicons to determine the polarity of an aspect [2].

Compared with previous works for ABSA in Portuguese, our results were well performed, although achieving a slightly worse performance. The great advantage of our model is that it does not need to be trained to extract the polarity of an aspect. Therefore, we can apply the same model in various domains and still achieve good results with little effort.

7. Conclusion

In this work, we have described our approaches to solve the task of Aspect Term Extraction (Task 1) and Sentiment Orientation Extraction (Task 2) for ABSAPT 2022.

For Aspect Term Extraction, we start with preprocessing techniques to clean the text review, then tokenize it and extract linguistic features. The features we chose were the token lemma, part-of-speech tag, and dependency tag. We then use these features to identify and extract aspects based on a rule-based method and a lexicon of domain aspects.

For Sentiment Orientation Extraction, we divided the process into two stages: Meaningful Surroundings Extraction and Sentiment Extraction. In Meaningful Surroundings Extraction, we identify which part of the review text is relevant to define the polarity of a given aspect. Then, in Sentiment Extraction, we use GoEmotions for Portuguese to predict emotions scores for a given aspect surrounding. Next, we map each emotion score to a positive, negative, or neutral polarity. After that, we combine the top three emotions with the highest scores to get the polarity classification of the aspect. Additionally, we use dependency parsing to assign neutral polarity to all aspects with the OBL tag.

Our approach achieved satisfactory results on the train and test corpora, winning third place on both tasks' final leaderboards.

Acknowledgments

The experiments described in this paper are also part of the Scientific Initiation project, funded by FAPESP, “Automatic inference of affective charge in social media posts” (2021/07067-0), under development at the Federal University of São Carlos, Brazil.

References

[1] M. M. y Gómez, J. Gonzalo, F. Rangel, M. Casavantes, M. Ángel Álvarez Carmona, G. Bel-Enguix, H. J. Escalante, L. Freitas, A. Miranda-Escalada, F. Rodríguez-Sánchez, A. Rosá, M. A. Sobrevilla-Cabezudo, M. Taulé, R. Valencia-García, Proceedings of the iberian languages evaluation forum (iberlef 2022), 2022.

[2] Feature-level sentiment analysis applied to brazilian portuguese reviews, Ph.D. thesis, 2015. URL: http://tede2.pucrs.br/tede2/handle/tede/6031, faculdade de Informática.

[3] Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa, Ph.D. thesis, 2021. URL: http://guaiaca.ufpel.edu.br/handle/prefix/7433, centro de Desenvolvimento Tecnológico.

[4] L. Hammes, L. Freitas, Utilizando bertimbau para a classificação de emoções em português, in: Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, SBC, Porto Alegre, RS, Brasil, 2021, pp. 56–63. URL: https://sol.sbc.org.br/index.php/stil/article/view/17784. doi:10.5753/stil.2021.17784.

[5] E. Jana, V. Uma, Aspect extraction and sentiment analysis for e-commerce product reviews, in: D. J. Hemanth, V. D. A. Kumar, S. Malathi, O. Castillo, B. Patrut (Eds.), Emerging Trends in Computing and Expert Technology, Springer International Publishing, Cham, 2020, pp. 780–791.

[6] Émerson Philippe Lopes, L. Freitas, G. Gomes, G. Lemos, L. O. Hammes, U. Correa, Exploring bert for aspect-based sentiment analysis in portuguese language, The International FLAIRS Conference Proceedings 35 (2022). URL: https://journals.flvc.org/FLAIRS/article/view/130601. doi:10.32473/flairs.v35i.130601.

[7] M. Chaves, L. Freitas, R. Vieira, Hontology: A multilingual ontology for the accommodation sector in the tourism industry, 2012.

[8] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowledge-Based Systems 108 (2016) 42–49. URL: https://www.sciencedirect.com/science/article/pii/S0950705116301721. doi:https://doi.org/10.1016/j.knosys.2016.06.009, new Avenues in Knowledge Bases for Natural Language Processing.

[9] M. Honnibal, I. Montani, spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, 2017. To appear.

[10] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, S. Ravi, GoEmotions: A dataset of fine-grained emotions, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 4040–4054. URL: https://aclanthology.org/2020.acl-main.372. doi:10.18653/v1/2020.acl-main.372.

[11] Émerson Lopes, U. Correa, L. Freitas, Exploring bert for aspect extraction in portuguese language, The International FLAIRS Conference Proceedings 34 (2021). URL: https://journals.flvc.org/FLAIRS/article/view/128357. doi:10.32473/flairs.v34i1.128357.

A. Online Resources

The source code of our solution for both tasks is available in:

- https://github.com/LALIC-UFSCar/ABSAPT-2022