Aspect-based sentiment analysis in comments on political debates in Portuguese: evaluating the potential of ChatGPT

Eloize R. M. Seno / Federal Institute of São Paulo
eloize@ifsp.edu.br

Lucas G. T. Silva and **Helena M. Caseli / Federal University of São Carlos**
lucasteixeira@estudante.ufscar.br, helenacaseli@ufscar.br

Fábio S. I. Anno and **Fabiano M. Rocha Junior / Federal Institute of São Paulo**
{fabio.seyiji, fabiano.j}@aluno.ifsp.edu.br

Abstract

This work presents a first study on the use of ChatGPT in two main tasks of aspect-based sentiment analysis applied in the political domain: aspect detection (AD) and aspect-oriented polarity classification (PC). ChatGPT was compared with traditional knowledge-based methods and with a fine-tuned BERT model for emotion detection in Portuguese. We found that a simple heuristic based on named entity recognition performed better than ChatGPT in the AD task. In the PC task, ChatGPT showed a significantly greater potential to associate polarity with aspect than the other investigated approaches. The highest efficiency achieved using ChatGPT on the PC task was a macro-average F-measure of 57.88%, while the second best approach combining the use of lexicon with the BERT model achieved a macro-average F-Measure of 39.30%.

1 Introduction

The automatic analysis of public opinion shared on social media, also known as Sentiment Analysis or Opinion Mining, has been the focus of attention of many studies in recent years (e.g. Jain et al., 2021; Pereira, 2021; Soni and Rambola, 2022; Hung and Alias, 2023), as these opinions can assist in social behavior analysis and decision- and policy-making for companies and government.

The most common sentiment analysis involves polarity classification, where the overall sentiment of the analyzed text (e.g. a review, an article, or a sentence) is assessed as either positive, negative, or neutral. However, for a more refined and accurate analysis, it is crucial to identify the opinion targets such as the entities (for example, individuals, organizations and products) or aspects (properties) of entities to which the opinion refers to. For instance, in the review "The Moto G6 camera is bad." there is a negative polarity derived from the word "bad" associated to the aspect "camera" of the entity "Moto G6". Entity-level and aspect-level sentiment analysis are commonly referred as aspect-based sentiment analysis (ABSA) (Schouten and Frasincar, 2016; Do et al., 2019).

This paper focus on the two main steps of ABSA task: aspect detection (AD) and polarity classification (PC). Thus, first the opinion targets are identified in the texts. Then, based on the sentiment words in the context of each opinion target, a polarity is assigned to each one (Tsytsarau and Palpanas, 2012). ABSA is considered a fine-grained sentiment analysis, and represents the most complex level of analysis, due to the complexity of modeling the semantic connections between a given target (aspect) and the words in its surrounding context (Zhang et al., 2018).

Although there is a vast literature on ABSA for English (e.g. Schouten and Frasincar, 2016; Zhang et al., 2018; Do et al., 2019; Soni and Rambola, 2022; Wu et al., 2023), according to Pereira (2021), there is a lack of research on the subject for Portuguese, despite advances in recent years (da Silva et al., 2022; Seno et al., 2023).

The first works for Portuguese focused on detecting aspects (e.g. Balage Filho, 2017; Vargas and Pardo, 2018; Costa and Pardo, 2020; Vargas and Pardo, 2020; Machado and Pardo, 2022), especially exploring the domain of reviewing products such as cameras, smartphones and books. Research involving the polarity association with each aspect is less common and focuses on hotel reviews (Assi et al., 2022; Gomes et al., 2022; Machado and Pardo, 2022) or general posts on the web (Saias et al., 2018). Some domains, such as politics, have practically not been explored on ABSA.

Given this context, in this study we investigated and evaluated different approaches for opinion target detection and target-oriented sentiment classification in comments on political debate in Portuguese. More specifically, we investigate the potential and limitations of ChatGPT[1] and compare it with a BERT model fine-tuned for emotion detection in Portuguese and with traditional knowledge-based approaches. In this sense, this work extends the previous one by Seno et al. (2023) by also considering the aspect detection task.

As public interest in pre-trained generative models like OpenAI's ChatGPT continues to grow, it is expected that these models will be used in various natural language processing tasks, including ABSA. In fact, several recent initiatives for the Portuguese language have emerged (e.g. de Fonseca et al., 2023; dos Santos and Paraboni, 2023; Seno et al., 2023; Oliveira et al., 2023). Thus, our ultimate goal is to find out if it is still useful to use knowledge-based methods combined with a fine-tuned BERT model for ABSA in comments about political debates in Portuguese or if ChatGPT is the best option for this subjective task.

The remainder of this paper is organized as follows: Section 2 describes related work. Section 3 presents the corpus used in our experiments and details its processing. The investigated approaches for the aspect detection and the polarity classification tasks are described in Sections 4 and 5, respectively. Experimental results are presented in Section 6. Finally, Section 7 finishes this paper with some conclusions.

2 Related Work

Previous approaches in ABSA use language rules, knowledge-based methods, statistical techniques or hybrid approaches (Cambria, 2016; Schouten and Frasincar, 2016; Pereira, 2021). Language rules typically rely on part-of-speech (PoS) tags and syntactic dependency relations to identify contextual patterns that capture the properties of terms and their relationships. Knowledge-based methods rely on linguistic resources built from corpora, such as lexicons, ontologies and wordnets, to identify words and expressions indicative of feelings in the input sentence. Besides relying on knowledge bases, these techniques also explore language rules to determine the context of words. Statistical methods use machine learning algorithms, which are trained from linguistic features extracted from texts. In general, these methods are based on high-frequency nouns and noun phrases in the input texts, some of which can reflect the sentiment polarity shown by the reviewer towards an aspect (e.g. Htay and Lynn, 2013; Perikos and Hatzilygeroudis, 2017). Statistical methods are usually simple and effective, but semantically weak and need a lot of data for training. On the other hand, approaches based on lexicons and ontologies are limited to non-exhaustive coverage of these resources. Combining knowledge with the use of rules appears to be a promising approach (Saias et al., 2018).

In Saias et al. (2018), for example, aspect detection on tweets and web comments in Portuguese was based on expressions having a relationship with the entity (opinion target) and possibly some polarized term. The relationship was identified using syntactic dependency and rules based on the PoS tags of the words in the surrounding context. Sentiment polarity was determined by a Maximum Entropy classifier, whose features include the entity mention, the aspect and its support text and sentiment lexicon-based polarity clues. The authors reported the following F-measure values for polarity classification: 66.0%, 74.0%, and 76.0% for positive, negative, and neutral class, respectively. The aspect detection task was not independently evaluated. In a similar manner, in this work we investigate the use of syntactic dependency and PoS tags for both aspect detection and polarity classification, as will be explained in Sections 4 and 5.

In Assi et al. (2022), aspect detection in the domain of hotel reviews is based on a domain-specific lexicon, built from corpus, and on rules based on PoS and syntactic dependency. In addition, a domain ontology is used to filter the candidate aspects extracted based on the rules, keeping only those that are present in the ontology. For polarity classification they used GoEmotion (Hammes and Freitas, 2021), a fine-tuning of the BERTimbau (Souza et al., 2020) for the classification of emotions in Portuguese, and then mapped each emotion to one of the three possible polarities. Following the approach of Assi et al. (2022), in this work we also investigate the use of the GoEmotions model for polarity classification (see Section 5).

Other important work for us is that of Catharin and Feltrim (2018). The authors evaluated three language rule-based approaches for aspect detection in Portuguese. The approaches were based on the well-known Centering Theory (Grosz et al., 1995), on morphosyntactic patterns and on heuristics that considered the subject of a sentence or proper names as the aspect. For the evaluation of the approaches Catharin and Feltrim (2018) used

[1]https://chat.openai.com/

SentiCorpus-PT (Carvalho et al., 2011), the same corpus used in this study (see Section 3). The heuristic which extracts all proper nouns of each sentence as aspect performed better than the other approaches, achieving 70.0% Precision, 61.0% Recall and 65.0% of F-measure. These results show that simple approaches based only on PoS tags may yield good results in this domain. Based on this intuition, in this work we investigate several heuristics using different POS tags and syntactic information for the aspect detection task (Section 4) and compared them with the GPT model.

3 Corpus Description and Preprocessing

In this study, the SentiCorpus-PT (Carvalho et al., 2011) was used as the research corpus, which consists of 1,082 comments (3,867 sentences) on television debates relating to the 2009 Portuguese Parliament elections. SentiCorpus-PT provides reference annotations for explicit opinion targets (aspects) in each sentence, along with the associated polarity for each one of them. 94.3% of the sentences has at least one annotated target, and 79% has exactly one target.

The opinion targets in this corpus are mostly human entities, namely politicians, media personalities (e.g. journalists) or users (commentators). Polarity is a value between -2 (the strongest negative value) and 2 (the strongest positive value). However, in our study polarity -2 was mapped to -1 (negative) and polarity 2 was mapped to 1 (positive), as will be explained in Section 5.

Table 1 shows an example of sentence extracted from SentiCorpus-PT with two distinct opinion targets (i.e., “Jerónimo” and “Louçã”) and their respective polarities (POL).[2]

The corpus preprocessing consisted of the following steps: tokenization, lemmatization, part-of-speech (PoS) tagging, syntactic dependency analysis and named entity recognition (NER). For the preprocessing we used UDPipe 2.0[3] and for NER we used SpaCy library[4]. 4.62% of the sentences in the entire corpus (approximately 178 sentences) could not be processed properly by the dependency parser due to words with capital letters (not necessarily proper nouns). Therefore, they were discarded. Of the remaining sentences, 7.60% of them did not have any target marked and were also discarded. Thus, 3,408 sentences were considered in this study.

4 Aspect Detection

Aiming at achieving our goal to define if ChatGPT outperforms knowledge-based methods, we carried out experiments with traditional knowledge-based methods, that combine the use of lexicons with syntactic and morphosyntactic heuristics, and compare them with the GPT model. The following sections describe the knowledge-based approaches and ChatGPT-based approaches investigated in this study.

4.1 Knowledge-based approaches

Considering the corpus characteristics (Section 3) we expected many targets to be proper noun, noun or named entities and to be related to a sentiment word in the input sentence. In addition, based on the notion that opinion targets would be relevant entities in the sentences, we expected many aspects to have the function of subject. Based on these intuitions, we implemented the following heuristics for aspect detection:

- `NE`: all named entities are considered aspects;
- `NE+NOUN`: all named entities and nouns are considered aspects;
- `NE+NOUN(Subj)`: named entities and nouns with subject function are considered aspects;
- `NE+NOUN(Pol)`: all named entities and nouns related to a sentiment word (via syntactic dependency) are considered aspects;
- `PROPN`: all proper nouns are considered aspects;
- `PROPN(Subj)`: all proper nouns with subject function are considered aspects;
- `PROPN+NOUN(Subj)`: proper nouns and nouns with subject function are considered aspects;
- `PROPN+NOUN(Pol)`: proper nouns and nouns related to a sentiment word (via syntactic dependency) are considered aspects.

For `NE+NOUN(Pol)` and `PROPN+NOUN(Pol)` we use SentiLex-PT02 (Silva et al., 2012) and LIWC (Balage Filho et al., 2013) as sentiment lexicons.

[2] ChatGPT’s translation for this example: It was indeed a cordial, civilized debate in which Jerónimo behaved like a gentleman and Louçã backed down.

[3] https://ufal.mff.cuni.cz/udpipe/2 (Accessed on: October 21, 2023).

[4] https://spacy.io/ (Accessed on: October 21, 2023).

Table 1: Example of a sentence extracted from SentiCorpusPT (for simplicity, other details of the annotation have been omitted).

< F ID ="1" TARG = "Jerónimo de Sousa" POL = "1"> Foi de facto um debate cordato, civilizado em que <TARG TYPE="NAME">Jerónimo</TARG> se mostrou um senhor e o Louçã meteu a viola no saco. </F> <F ID = "1" TARG ="Francisco Louçã" POL = "-1"> Foi de facto um debate cordato, civilizado em que Jerónimo se mostrou um senhor e o <TARG TYPE="NAME">Louçã</TARG> meteu a viola no saco. </F>

Considering that SentiLex-PT02 was designed for sentiment analysis on human entities and knowing that many opinion targets in the corpus are humans (politicians), first we check if the word has any polarity associated in SentiLex-PT02. Then, only when the word was not found in SentiLex-PT02, we consult LIWC. More details about these lexicons will be given in Section 5.

4.2 ChatGPT-based approach

To provide unbiased and scalable communication with ChatGPT, we used the OpenAI API, which gives us access to all the company's models via HTTP request. This approach gives us access to essential text analysis tools that are not normally available via GPT's conventional web service.

We developed a Python script based on the OpenAI library[5] and used the ChatCompletion method to make the API's requests. By doing so, it was possible to fine-tune the model's attributes according to our specific needs. The attributes chosen were:

- `Model`: "gpt-3.5-turbo"
- `Message Structure`: We used a two-part message structure. The first part, with the "system" role, was used to define the context of the conversation. The second part, with the "user" role, was used to present the user's sentence.
- `Maximum Tokens`: In line with the recommendations in the documentation, we set the maximum number of tokens at 1,024.
- `Temperature`: We set the temperature to 0 in order to get objective answers from the model.

ChatGPT is a prompt-based model. In general terms, it receives as input a string, called prompt, containing the description of the task to be performed by the system and generates the outputs as requested. The main challenge in dealing with the ChatGPT consists of defining a prompt that generates the expected outputs for a given task. The choice of prompt significantly impacts the outcome (Oliveira et al., 2023).

At the beginning, several prompt attempts were made using the temperature parameter set at 0.5 (empirically). However, the model varied greatly in responses and sometimes contradicted itself. After consulting the literature (e.g. Oliveira et al., 2023; de Fonseca et al., 2023; dos Santos and Paraboni, 2023), we changed the parameter to zero. We soon realized that the model became more stable and coherent in its responses. For this reason, we use temperature set at zero in both the aspect detection (AD) and polarity classification (PC) tasks.

In the AD task, the following prompt was provided for the model: *"Dada a seguinte sentença, responda no formato ["alvo1"] o(s) alvo(s) de opinião presente(s) na sentença."* ("Given the following sentence, answer in the format ["target1"] the opinion target(s).")

In addition to the aspect detection task, we also evaluate the GPT model on the named entity recognition (NER), in order to make a fairer comparison of the model with the heuristic that extracts named entities. Regarding NER task, the following prompt was used: *"Dada a seguinte sentença, responda no formato ["entidade1", "entidade2"] a(s) entidade(s) nomeada(s) presente(s) na sentença."* ("Given the following sentence, answer in the format ["entity1", "entity2"] the named entity(ies) present in the sentence.")

5 Polarity Classification

Aiming at achieving our goal to check if ChatGPT outperforms our approach for polarity classification, we compared the results from ChatGPT with our approach based on traditional lexical resources

[5] https://github.com/openai/openai-python

combined with a fine-tuned BERT model for emotion detection in Portuguese (a hybrid approach).

5.1 Our approach

In the approach we proposed for polarity classification the following lexical resources were used:

- SentiLex-PT02[6] (Silva et al., 2012) – a sentiment lexicon for Portuguese, made up of 7,014 lemmas, and 82,347 inflected forms. In our experiments we used only the single word entries of both, the lemmatized (`SentiLex-lem-PT02.txt`) and the inflected (`SentiLex-flex-PT02.txt`) versions with 6,344 and 47,411 entries, respectively. The adopted approach was: if the lemma of a word was not found in the lemmatized version we looked for its surface form in the full version.

- LIWC[7] (Balage Filho et al., 2013) – a Brazilian Portuguese version of LIWC[8] with around 127,000 entries. We considered 24,324 of them associated with the positive (`posemo`) or negative (`negemo`) polarity but not both[9]. In addition to the full word forms, we also considered the 2,665 truncated (with an * at the end) words associated to one of the mentioned polarities.

- OpLexicon v3.0[10] – a sentiment lexicon for the Portuguese language automatically created and revised by linguists based on Open Lexicon V2.1 (Souza and Vieira, 2012). In our experiments, we considered the 31,605 words associated with positive (1), negative (-1) or neutral (0) polarity.

- WordNetAffectBR[11] (Pasqualotti, 2015) – a lexicon with 289 words associated with negative (-) or positive (+) polarity.

- AffectPT-br[12] (Carvalho et al., 2018) – a Brazilian Portuguese affective lexicon based on the LIWC 2015 English dictionary. AffectPT-br has the same format as LIWC with words associated with the positive (`posemo`) or negative (`negemo`) polarity but not both at the same time. From AffectPT-br we were able to retrieve 510 full and 631 truncated (with an * at the end) word forms.

Besides the lexicons, we also used a fine-tuned BERT model for emotion detection in Portuguese[13] (Hammes and Freitas, 2021) in which the BERTimbau (Souza et al., 2020) was fine-tuned with a translated version of GoEmotions (Demszky et al., 2020) being able to detect 27 emotions plus a neutral class. In this case, we considered as positive polarity the emotions: "admiration", "amusement", "approval", "caring", "desire", "excitement", "gratitude", "joy", "love", "optimism", "pride" and "relief". We considered as negative polarity the emotions: "anger", "annoyance", "disappointment", "disapproval", "disgust", "embarrassment", "fear", "grief", "nervousness", "remorse" and "sadness". We considered as neutral the emotions: "confusion", "curiosity", "realization" and "surprise" besides the neutral class.

In order to have a bigger coverage we also experimented with the NILC embeddings[14] (Hartmann et al., 2017) by considering the polarity associated to the best neighbour of each word. Following this approach, if a word was not found in a lexicon, its best neighbour according to NILC embeddings was considered to the look up on that lexicon[15].

From these resources, we followed three approaches to attach the polarity to opinion targets. The first approach (`B`) takes into account all the polarity words or emotions detected in the whole sentence. The `NEG` inverts the polarity defined in the lexicon (`B`) approach if a negation word[16] occurs in the sentence. Finally, the (`D`) approach only considers the polarity words associated with an opinion target by means of a syntactic dependency relation.

5.2 ChatGPT-based approach

For this task we used the same parameters as for the aspect detection task (Section 4.2), just changing

[6] `https://b2share.eudat.eu/records/93ab120efdaa4662baec6adee8e7585f`

[7] `http://143.107.183.175:21380/portlex/images/arquivos/liwc/LIWC2007_Portugues_win.dic.txt`

[8] `http://www.liwc.net/`

[9] For example, the word "*desculpa*" (sorry) is associated with both `posemo` (code 126) and `negemo` (code 127).

[10] `https://github.com/marlovss/OpLexicon`

[11] `https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/wordnetaffectbr/`

[12] `https://github.com/LaCAfe/AffectPT-br/blob/master/AffectPT-br`

[13] `https://github.com/Luzo0/GoEmotions_portuguese`

[14] `http://nilc.icmc.usp.br/embeddings`

[15] We did experiments considering the top-3 best neighbours but the results were worse than when considering only the top-1 best neighbour.

[16] We considered the following negation words: *"não", "jamais", "nada", "nem", "nenhum", "nenhuma", "ninguém", "nunca", "tampouco", "zero"* that could represent the English words no, not, never, nothing, neither, none, nobody, zero.

the prompt and the entries.

For the polarity classification task we use the following prompt: *"Dada uma sentença e seus respectivos marcadores sobre o mesmo alvo de opinião responda apenas com o caractere (-1) se ela possui conotação negativa, (0) se for neutra ou (1) se for positiva"* ("Given a sentence and its respective markers about the same opinion target, only respond with the character (-1) if it has a negative connotation, (0) if it is neutral or (1) if it is positive"). And for the message, we sent the sentence and the corresponding set of terms that refers to the targets of that sentence.

6 Experiments and Results

In order to understand the potential of each heuristic and approach to detect and extract aspects, we evaluate aspect detection task independently of the polarity classification task. The next sections present the results obtained for each task.

6.1 Results for Aspect Detection

Following Catharin and Feltrim (2018), we considered that an aspect (opinion target) was correctly detected when the output of the strategy was equal to or contained within a reference target for the processed sentence.

Table 2 presents the precision (P), recall (R) and F-measure (F) values obtained for each heuristic and for the approaches based on the GPT model.

Table 2: Results for Aspect Detection

Strategy	P	R	F
NE_ChatGPT	**75.46%**	71.36%	**73.36%**
NE	60.57%	61.63%	**61.10%**
ChatGPT	62.13%	56.95%	**59.43%**
NE+NOUN(Subj)	52.02%	66.16%	58.25%
NE+NOUN(POL)	44.11%	64.44%	52.37%
PROPN	51.08%	48.36%	49.68%
NE+NOUN	24.26%	**74.69%**	36.63%
PROPN+NOUN(Subj)	43.48%	23.67%	30.65%
PROPN+NOUN	20.43%	63.50%	30.91%
PROPN(Subj)	**65.93%**	18.21%	28.53%
PROPN+NOUN(POL)	4.16%	6.08%	4.94%

As shown in Table 2, the strategies that consider all named entities of the sentence as aspects (NE_ChatGPT and NE) obtained the best results. Among them the best strategy is the one that uses ChatGPT for named entity recognition (NER). The NER task presents itself as a simpler task for the GPT model than the aspect detection task. It is important to note that the second best Precision value (65.93%) was achieved with the strategy that only considers proper nouns with subject function (PROPN(Subj)) as aspects. However, this strategy presented low recall. The highest Recall value was obtained with the strategy NE+NOUN (74.69%).

For the top three strategies with the best F-measure values, we performed a manual review to also consider those that partially matched the reference targets. Table 3 presents the results after human review. All strategies had an improvement in all assessment measures after review. The biggest gain was achieved by the NE heuristic, that is, an increase of around 9 percentage points in terms of precision and recall and approximately 8 percentage points for the F-measure. These gains in precision and recall are due to cases such as "Jerónimo!" and "Tvi!", automatically extracted, and which were not contained in the reference targets (i.e. "Jerónimo" and "TVI").

Table 3: Results for Aspect Detection after manual review

Strategy	P	R	F
NE_ChatGPT	**77.75%**	**73.52%**	**75.58%**
NE	69.35%	70.16%	69.76%
ChatGPT	65.60%	60.14%	62.75%

6.2 Results for Polarity Classification

In Table 4 we present the approaches which achieved the best values for precision (P), recall (R) and F-measure (F) for each class (Positive, Negative or Neutral) as well as the macro-average F-Measure (M-F) (henceforth, Macro-F) considering all the three classes.[17] The values presented here are those obtained when considering the top-1 best neighbour according to NILC word embeddings (as explained in section 5.1) even though the improvement when using the best neighbour was a small one (less than 1 percentage point in Macro-F).

As one can notice from Table 4, the best overall performance in terms of Macro-F was achieved by the GPT model (57.88%). The second best performance (i.e. 39.30%) was obtained using the polarity combination (sum) of SentiLex-PT02 (SL) polarity and GoEmotions (GE) without taking into account the syntactic dependency relation between the opinion target and the polarity word (SL-B+GE).

[17]We tested all possible combinations of lexical resources and GoEmotions and due to space limitations only the combinations with the best values for at least one of the evaluation measures in each class are presented here.

Table 4: Results for Polarity Classification

	Positive			Negative			Neutral			All
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**M-F**
ChatGPT	**62.54%**	**61.67%**	**62.10%**	80.57%	**73.68%**	**76.97%**	**29.87%**	41.02%	**34.57%**	**57.88%**
WN-B	25.25%	7.08%	11.06%	72.13%	13.94%	23.36%	**20.51%**	1.56%	2.90%	12.44%
WN-D	17.14%	0.83%	1.58%	**83.08%**	2.55%	4.95%	0.00%	0.00%	0.00%	2.18%
GE-B	**48.10%**	28.19%	35.55%	82.86%	16.68%	27.77%	17.16%	**83.79%**	28.49%	30.60%
SL-B+GE	33.51%	43.19%	37.74%	75.60%	40.12%	52.42%	19.35%	49.02%	27.75%	**39.30%**
SL-D+GE	43.89%	31.94%	36.97%	81.68%	22.12%	34.81%	17.86%	78.52%	**29.10%**	33.63%
LW-B+GE	26.44%	**56.81%**	36.09%	73.83%	29.87%	42.53%	16.08%	29.69%	20.86%	33.16%
AF-D+GE	44.72%	35.28%	**39.44%**	79.25%	20.94%	33.13%	17.69%	76.76%	28.75%	33.77%
⋆-B+GE	27.96%	45.97%	34.77%	69.69%	**42.91%**	**53.12%**	17.07%	28.71%	21.41%	36.43%

GoEmotions alone (GE-B) was the one with the best Precision for Positive class (48.10%) and the best Recall for the Neutral one (83.79%). In fact, GoEmotions has a tendency for the neutral class, as pointed out in previous work (Seno et al., 2023), what could explain that bigger Recall value. It is worth noticing that the combination of one or more lexicons with GoEmotions figured as 5 out of 8 best approaches.

The WordNetAffectBR (WN), with only 289 entries, was the one with the best Precision for Negative (83.08%) and Neutral (20.51%) classes when considered the syntactic dependency relations (D) or not (B), respectively. The best F-measure for Positive class (39.44%) was obtained with a combination (sum) of AffectPT-br (AF), taking into account the syntactic dependency relations (D) and GoEmotions (GE).

Finally, we tested a combination (⋆) of all lexicons[18] which led to the best Recall (42.91%) and F-measure (53.12%) for the Negative class when the syntactic dependency relations were not considered (B).

From the described results we can conclude that our lexicon and GoEmotions based approach is still far from the performance of ChatGPT on the same task of assigning the correct polarity for a given opinion target. We can also conclude that the simple approach we followed to take into account negation words did not impact positively in our results.

7 Conclusions and Future Work

In this paper we evaluate different approaches aiming to solve the two main tasks of aspect-based sentiment analysis (ABSA) applied in the political domain: aspect detection (AD) and polarity classification (PC). More specifically, for the first task of AD we investigate the potential of ChatGPT and compared it with traditional knowledge-based methods that combine the use of lexicons and morphosyntactic and syntactic heuristics.

In the experimental results, the heuristic that considers all named entities in the input sentence as opinion targets (aspects) performed better than ChatGPT when applied to the AD task (69.76% F-measure against 62.75% F-measure). However, when applying the ChatGPT named entity heuristic, this model obtained the best result (75.58% F-measure). Although it is not possible to do a direct comparison[19], when Catharin and Feltrim (2018) evaluated their aspect detection approaches using SentiCorpus-PT, the same corpus used in this research, the highest reported F-measure was 65.0%, achieved using a heuristic based in the extraction of proper names.

We also investigate the potential of ChatGPT in the polarity classification task. Besides the knowledge-based approaches, we also compared it to a fine-tuned BERT model for emotion detection in Portuguese. Results from an experimental evaluation indicated that ChatGPT has the potential to identify the polarity associated with each opinion target of an input sentence with a performance significantly superior to the performance of the other approaches investigated. However, it is worth mentioning that using ChatGPT presents some challenges, such as choosing the appropriate input prompt with the description of the task to be performed by the system, crucial for it to understand what we expect as an outcome, and the variability of responses given to the same input at

[18] The polarity of a word was assigned if it was found in one of these lexicons, in this order: SentiLexPT02, WordNetAffectBR, AffectPT-br, OpenLexicon v3.0, LIWC. This order was defined empirically based on the coverage and accuracy of the polarity in those resources.

[19] Catharin and Feltrim (2018) used only 50% of the sentences in the corpus to evaluate their approaches, which were randomly selected and were not available for comparison with other works.

different times.

Our results suggest the promising feasibility of using ChatGPT to associate polarity with targets in comments in the political domain in Portuguese. For the AD task, however, this model may not represent the ideal solution, since alternative methods, characterized by simplicity and low computational cost, have demonstrated comparable performance in the domain of the analyzed texts.

As future work we intend to compare the performance of ChatGPT with pre-trained large language models for Portuguese fine-tuned in both tasks: aspect detection and polarity classification. Regarding the aspect detection task, specifically, we also intend to investigate the identification of non-explicit aspects in the text (i.e. implicit aspects).

Acknowledgements

The work presented in this paper meets some goals of the FAPESP Grant #2022/03090-0. We also thank the Graduate Program in Computer Science (PPGCC) from UFSCar, the Federal Institute of São Paulo (IFSP) and the National Council for Scientific and Technological Development - CNPq for the financial support in the form of a PIBIC scholarship.

References

Fernanda Malheiros Assi, Gabriel Barbosa Candido, Lucas Nildaimon dos Santos Silva, Diego Furtado Silva, and Helena Medeiros Caseli. 2022. Ufscar's team at ABSAPT 2022: using syntax, semantics and context for solving the tasks. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)*, volume 3202 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Pedro P. Balage Filho. 2017. *Aspect extraction in sentiment analysis for portuguese language*. Ph.D. thesis, São Carlos - SP.

Pedro P. Balage Filho, Thiago A. S. Pardo, and Sandra M. Aluísio. 2013. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pages 215–219.

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107.

Flavio Carvalho, Gabriel dos Santos, and Gustavo Paiva Guedes. 2018. Affectpt-br: an affective lexicon based on liwc 2015. In *37th International Conference of the Chilean Computer Science Society (SCCC 2018)*, University Andres Bello, Campus Antonio Varas, Santiago – Chile.

Paula Carvalho, Luís Sarmento, Jorge Teixeira, and Mário J. Silva. 2011. Liars and saviors in a sentiment annotated corpus of comments to political debates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 564–568, Portland, Oregon, USA. Association for Computational Linguistics.

Leonardo Catharin and Valéria Delisandra Feltrim. 2018. *Finding opinion targets in news comments and book reviews: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings*, pages 375–384.

Raul Costa and Thiago Pardo. 2020. Métodos baseados em léxico para extração de aspectos de opiniões em português. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 61–72, Porto Alegre, RS, Brasil. SBC.

Felix L. V. da Silva, Guilherme da S. Xavier, Heliks M. Mensenburg, Rodrigo F. Rodrigues, Leonardo P. dos Santos, Ricardo M. Araújo, Ulisses Brisolara Corrêa, and Larissa A. de Freitas. 2022. ABSAPT 2022 at iberlef: Overview of the task on aspect-based sentiment analysis in portuguese. *Procesamiento del Lenguaje Natural*, 69:199–205.

Felipe de Fonseca, Ivandré Paraboni, and Luciano Digiampietri. 2023. Contextual stance classification using prompt engineering. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 33–42, Porto Alegre, RS, Brasil. SBC.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Hai Ha Do, PWC Prasad, Angelika Maag, and Abeer Alsadoon. 2019. Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*, 118:272–299.

Wesley dos Santos and Ivandré Paraboni. 2023. Predição de transtorno depressivo em redes sociais: Bert supervisionado ou ChatGPT zero-shot? In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 11–21, Porto Alegre, RS, Brasil. SBC.

Juliana Resplande Sant'Anna Gomes, Eduardo Augusto Santos Garcia, Adalberto Ferreira Barbosa Junior, Ruan Chaves Rodrigues, Diogo Fernandes Costa Silva, Dyonnatan Ferreira Maia, Nádia Félix Felipe da Silva, Arlindo Rodrigues Galvão Filho, and Anderson da Silva Soares. 2022. Deep learning Brasil at ABSAPT 2022: Portuguese transformer ensemble approaches.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Luiz Hammes and Larissa Freitas. 2021. Utilizando BERTimbau para a classificação de emoções em português. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 56–63, Porto Alegre, RS, Brasil. SBC.

Nathan Hartmann, Erick Fonseca, Christopher Shulby, Marcos Treviso, Jéssica Rodrigues, and Sandra Aluísio. 2017. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In *Anais do XI Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 122–131, Porto Alegre, RS, Brasil. SBC.

Su Htay and Khin Lynn. 2013. Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal*, 2013:394758.

Lai Hung and Suraya Alias. 2023. Beyond sentiment analysis: A review of recent trends in text based sentiment analysis and emotion detection. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 27:84–95.

Praphula Kumar Jain, Rajendra Pamula, and Gautam Srivastava. 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41:100413.

Mateus T. Machado and Thiago A. S. Pardo. 2022. Evaluating methods for extraction of aspect terms in opinion texts in Portuguese - the challenges of implicit aspects. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3819–3828, Marseille, France. European Language Resources Association.

Amanda Oliveira, Thiago Cecote, Pedro Silva, Jadson Gertrudes, Vander Freitas, and Eduardo Luz. 2023. How good is chatgpt for detecting hate speech in portuguese? In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 94–103, Porto Alegre, RS, Brasil. SBC.

Paulo Roberto Pasqualotti. 2015. *WordNet Affect BR – uma base de expressões de emoção em Português*. Novas Edições Acadêmicas.

Denilson Alves Pereira. 2021. A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*, 54(2):1087–1115.

Isidoros Perikos and Ioannis Hatzilygeroudis. 2017. Aspect based sentiment analysis in social media with classifier ensembles. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 273–278.

José Saias, Mário Mourão, and Eduardo Oliveira. 2018. Detailing sentiment analysis to consider entity aspects: An approach for portuguese short texts. *Transactions on Machine Learning and Artificial Intelligence*, 6(2):26–35.

Kim Schouten and Flavius Frasincar. 2016. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830.

Eloize R. M. Seno, Fábio S. I. Anno, Lucas Lazarini, and Helena M. Caseli. 2023. Classificação de polaridade orientada aos alvos de opinião em comentários sobre debate político em português. *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2023)*, pages 84–93.

Mário J. Silva, Paula Carvalho, and Luís Sarmento. 2012. Building a sentiment lexicon for social judgement mining. In *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*.

Piyush Kumar Soni and Radhakrishna Rambola. 2022. A survey on implicit aspect detection for sentiment analysis: Terminology, issues, and scope. *IEEE Access*, 10:63932–63957.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: Pretrained BERT models for Brazilian portuguese. In *Intelligent Systems*, pages 403–417, Cham. Springer International Publishing.

Marlo Souza and Renata Vieira. 2012. Sentiment analysis on twitter data for portuguese language. In *Computational Processing of the Portuguese Language*, pages 241–247, Berlin, Heidelberg. Springer Berlin Heidelberg.

Mikalai Tsytsarau and Themis Palpanas. 2012. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3):478–514.

Francielle Vargas and Thiago Pardo. 2018. *Aspect Clustering Methods for Sentiment Analysis: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings*, pages 365–374.

Francielle Vargas and Thiago Pardo. 2020. Linguistic rules for fine-grained opinion extraction: Workshop proceedings of the 14th international aaai conference on web and social media, 2020.

Haiyan Wu, Chaogeng Huang, and Shengchun Deng. 2023. Improving aspect-based sentiment analysis with knowledge-aware dependency graph network. *Information Fusion*, 92:289–299.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis : A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8.