

Evaluating Methods for Extraction of Aspect Terms in Opinion Texts in Portuguese – the Challenges of Implicit Aspects

Mateus Tarcinalli Machado, Thiago Alexandre Salgueiro Pardo

Interinstitutional Center for Computational Linguistics (NILC),
Institute of Mathematical and Computer Sciences, University of São Paulo
mateusmachado@usp.br, taspardo@icmc.usp.br

Abstract

One of the challenges of aspect-based sentiment analysis is the implicit mention of aspects. These are more difficult to identify and may require world knowledge to do so. In this work, we evaluate frequency-based, hybrid, and machine learning methods, including the use of the pre-trained BERT language model, in the task of extracting aspect terms in opinionated texts in Portuguese, emphasizing the analysis of implicit aspects. Besides the comparative evaluation of methods, the differential of this work lies in the analysis’s novelty using a typology of implicit aspects that shows the knowledge needed to identify each implicit aspect term, thus allowing a mapping of the strengths and weaknesses of each method.

Keywords: Sentiment analysis, Aspect extraction, Implicit aspects

1. Introduction

With the great expansion of social networks and e-commerce services and, consequently, the increase in the production of online reviews and comments, there was the need to develop new methods for processing such data. Sentiment analysis, or opinion mining, can be defined as an area of study at the intersection between Computer Science and Linguistics that aims to automatically determine sentiments in a text (Taboada, 2016) in order to analyze people’s opinions, feelings, assessments, attitudes and emotions concerning products, services, organizations, individuals, issues, topics, and their attributes (Liu, 2010).

Sentiment analysis can be performed at three levels: document, sentence and aspect levels. Each of them includes different tasks and challenges, and the aspect-based sentiment analysis (ABSA) can be considered the most refined level. For example, consider the sentence “The hotel room was very small”. In this case, we have a negative opinion about the “room” aspect of the entity “hotel”. So we have a triple formed by the entity (hotel), aspect (room), and sentiment (negative). In ABSA, we can still have information about the author and time when the opinion was issued, forming a quintuple (Liu, 2010).

We can divide ABSA into several tasks, as those presented in the International Workshop on Semantic Evaluation (SemEval) (Pontiki et al., 2014):

- Aspect term extraction: aims to identify terms present in a sentence that are related to an aspect. In the example, we have the word “room” related to the aspect “room”.
- Aspect term polarity: aims to find the sentiment related to an aspect term, usually positive, negative or neutral.
- Aspect category detection: aims to identify the category of the aspect mentioned in a sentence.

For example, we can find other aspect terms related to the “room” category, such as “apartment” or “dormitory”.

- Aspect category polarity: aims to detect the polarity of the mentioned aspect category.

Aspects can be mentioned in texts in two different ways: explicitly, as in the example, or implicitly, when it is not mentioned directly in the text. For example, in “The hotel was too expensive”, we have the term “expensive” implicitly referring to the “price” aspect. For this term, following the terminology of Poria et al. (2014), we adopt the name “Implicit Aspect Clue” (IAC). Liu (2010) consider any aspect terms formed by nouns or noun phrases as explicit aspects and the others as implicit aspects. In this work, we follow the definition of Cruz et al. (2014), where nouns and nominal phrases can be classified as implicit, for example, when we refer to the “room” aspect through synonyms or related terms such as “accommodation” or “apartment”. In other situations, this relationship between IAC and aspect is not so clear. In the sentences “The smartphone is too small” and “The smartphone barely fits in my pocket”, the aspect “size” can be inferred by the adjective “small” and the expression “fits in my pocket”, respectively.

Few works deal directly with the implicit aspects, a fact probably linked to the difficulty in automatically finding them and also because of the high frequency of explicit aspects present in the texts, which are often sufficient for many applications. Two surveys (Ravi and Ravi, 2015; Rana and Cheah, 2016) show that, among 68 papers dealing with aspect extraction, only 11 analyzed the implicit ones. Despite this higher frequency of explicit ones, the implicit ones are not so few, and, for some categories, the implicit aspect terms are very frequent. According to Zhang and Zhu (2013), 30% of the analyzed reviews in their corpus contained implicit aspects. In another study (Panchendrarajan et

al., 2016), the authors found that 15% of the reviews contain implicit mentions, and 92% of the aspects related to restaurant employees were mentioned implicitly. These numbers show the importance of detecting the implicit aspects.

Language is another important point in relation to studies in this area. Most of them analyze texts in English or Chinese. Among 53 articles dealing with implicit aspects analyzed in two surveys (Tubishat et al., 2018; Ganganwar and Rajalakshmi, 2019), 33 were for English and 19 were for Chinese. In Portuguese, the issue of implicit aspects is much less referenced. In another survey (Pereira, 2021) that analyzed only papers on sentiment analysis for the Portuguese language, the author found only 2 papers mentioning implicit aspects (de Freitas A and Vieira, 2015; Vargas and Pardo, 2017).

In this paper, we focus on the aspect term extraction task with emphasis on the implicit aspects. We analyze frequency-based, hybrid (frequency-based and rule-based), and machine learning methods, including the use of the Bidirectional Encoder Representations from Transformers (BERT) pre-trained language model (Devlin et al., 2019) that has got very good results in many Natural Language Processing (NLP) tasks. In the analysis of the results, besides the traditional metrics (precision, recall, and f-measure), we used the typology of implicit aspect clues proposed by Machado et al. (2022), thus allowing a better understanding of the strengths and weaknesses of each method in relation to the implicit aspect detection task. We organized the rest of the paper as follows. Section 2 presents the main initiatives related to the aspect term extraction task and implicit aspects. Section 3 presents the datasets and methods used in this paper. Section 4 describes the results achieved by the tested methods, including the analysis performed with the mentioned typology. Section 5 presents some final remarks.

2. Related work

As we mentioned earlier, there is a scarcity of studies related to implicit aspects. Therefore, we have selected for this section some works that deal with this issue for texts in Portuguese and in English.

Cruz et al. (2014) are among the few authors that deal only with implicit aspects. They performed the annotation of the indicators of implicit aspects in the corpus presented in Hu and Liu (2004), creating, as far as we know, the first corpus with such kind of annotation. The authors implemented four methods to identify the indicators of implicit aspects, being three simpler methods to serve as baseline methods and a fourth one based on supervised machine learning using the Conditional Random Fields. In their experiments, they reached an F-measure of 0.29, proving how difficult such task is. In Vargas and Pardo (2017), the authors investigated six methods for grouping explicit and implicit aspects in product reviews. Four linguistic-based methods in-

spired by the literature, a statistical method (based on *word embeddings*) and a new proposal for a linguistic-based method were tested. To evaluate the methods of aspect clustering, the authors manually annotated reviews in Portuguese on smartphones, digital cameras, and book reviews. In each review, explicit and implicit aspects were marked and grouped. The implicit aspects were shown by the clue terms that signaled them. The experiments of the proposed method achieved F-measures of 0.71 in the domain of book, 0.60 in the domain of camera, and 0.58 in the domain of smartphone.

In Cai et al. (2020), the authors developed an aspect-based sentiment analysis method with a focus on implicit aspects. The experiments were carried out on four datasets in English, containing reviews of restaurants and laptops used in the SemEval of 2015 (Pontiki et al., 2015) and 2016 (Pontiki et al., 2016). The authors formulated the task as a problem of hierarchical prediction of categories and sentiments, where first the categories of aspects of a sentence are identified, and then the sentiments related to each category are identified using a convolutional network of hierarchical graphs with the pre-trained language model BERT. In the experiments, for the domain of restaurants, they reached F-measures of 0.64 and 0.74 for the datasets from 2015 and 2016, respectively. In the domain of laptops, they achieved 0.54 and 0.62 for the same years. Although the datasets do not have the tagging of the implicit aspect terms, the approach is able to identify categories mentioned explicitly and implicitly.

In Lopes et al. (2021), the authors present an approach for aspect category detection based on multilingual and Portuguese BERT pre-trained models. In the proposed approach, the authors used BERT’s Sentence Pair Classifier to predict whether an aspect category is related to the text or not. Using hotel reviews, aspect categories, and “related” and “unrelated” labels as inputs, the experiments achieved an F-measure of 0.90, detecting both explicitly and implicitly mentioned categories.

3. Data and Methods

As mentioned before, the scarcity of studies related to implicit aspects definitely does not mean that these are not relevant. In order to improve this scenario for the Portuguese language, we implemented and tested several methods for extracting aspect terms, focusing on their effect on implicit aspects. We start by describing the datasets used, then the typology of implicit aspect clues and, finally, the methods that we test.

3.1. Datasets

In our experiments, we analyzed two datasets formed by opinion texts in Portuguese, both tagged in relation to the explicit and implicit aspect terms. The first set, presented by Vargas and Pardo (2017), comprises 180 reviews about cameras, books, and smartphones, with texts extracted from the Buscape (Hartmann et al.,

2014) and ReLi (Freitas et al., 2012) corpora. The second dataset was formed by joining the corpus of de Freitas A and Vieira (2015), which has the explicit terms tagged, with the annotation of the implicit aspect terms performed and explained by Machado et al. (2022). This set contains 194 reviews about hotels collected from the TripAdvisor website. Table 1 presents the composition of each dataset used in the experiments, separated by domains (given the interests in this paper, the last two columns show the numbers of explicit and implicit aspects in each domain).

Domains	Reviews	Words	Aspects	Expl.	Impl.
Cameras	60	3,997	352	299	53
Books	60	5,515	330	304	26
Smartphones	60	6,210	455	387	68
Hotels	194	13,940	1,417	999	415

Table 1: Statistics of used datasets

3.2. Typology of Implicit Aspect Clues

As explained, for a deeper analysis of the results, we used the typology created by Machado et al. (2022). In this work, the authors created categories and subcategories according to the type of knowledge needed to relate an IAC to its respective aspect. In summary, the categories and subcategories found were as follows:

- **Event (Action/Process/State):** the identification occurs through the identification of actions, processes, or states related to the aspect.
 - **Verb:** the IAC is identified by a verb. Example: the verb “to pay” that is related to the aspect “price”.
 - **Non-verbal form:** the IAC is identified by a term related to a verb. Example: the word “payment” that is related to the same aspect.
- **Feature:** the identification is given by terms related to the aspect or part of it.
 - **Attribute:** related to some characteristics of the aspect. Example: the IAC “material” related to the “design” aspect.
 - **Equivalence:** the IAC has a related meaning in relation to the aspect. Example: “hygiene” and the aspect “cleanliness”.
 - **Is-a:** the IAC is an item related to the aspect. Example: “breakfast” and the aspect “food”.
 - **Part-of:** the IAC is part of the aspect. Example: “bathroom” and the aspect “facilities”.
- **Qualification:** the IAC is related to a quality or sentiment about the aspect.
 - **Adjective:** the IAC is identified by an adjective. Example: “beautiful” related to the aspect “design”.

- **Equivalence:** the IAC is in an “equivalence” relation to an adjective. Example: the IAC “plain hotel” and the aspect “facilities”.
- **Nominal form:** the IAC major term is an adjective converted to another word class. Example: the IAC “beauty” related to the aspect “design”.
- **Contextual:** to identify an IACs of this category, it is necessary to have world knowledge about the product or service being analyzed, such as its operation, modes of use, or content.
 - **Location:** the IAC is related to the localization of the product. Example: the IAC “in the center of the region” is related to the “location” aspect.
 - **Related:** other contextual cases not related to location. Example: “musty smell” and the aspect “cleanliness”.

3.3. Methods for Aspect Extraction

In the experiments, we tested supervised and unsupervised methods for aspect term extraction. To enable the comparison among the methods, we divided the datasets of each domain into a training set, with 70% of the reviews, and a test set with the remaining reviews. The use of grid search to find the best values for the parameters of the methods, when necessary, was performed with data from the training set or a subset of it. Thus, all resulting metrics were calculated by analyzing only the prediction made on the test sets, allowing fair comparisons of methods.

Regarding the data, we have a difference in relation to other experiments, such as those carried out in SemEval. Usually, sentences that do not contain aspect terms are removed from the sets for this task. However, in our experiments, these sentences were kept, in order to simulate real world applications, where the user may be processing data directly collected from the web. This is a fact that can benefit or harm the methods performance depending on the algorithm.

In what follows, due to space limitations, we present an overview of the characteristics of the implemented methods. More details about them may be found in the mentioned references.

3.3.1. Freq-Baseline

This is an unsupervised frequency-based method commonly used as a baseline in aspect extraction studies. Its operation comprises selecting the most frequent nouns and noun phrases as aspect terms. In Hu and Liu (2004), the authors used a cutoff frequency of 1%, i.e., the terms found in at least 1% of the texts were considered aspects. In this work, in the same way as done by Machado et al. (2017), we varied this frequency in order to find the best cutoff value.

We also implemented a second version of this method using the Word2Vec distributional model (Mikolov et

al., 2013) to exclude candidates not related to the domain under study (Pavlopoulos and Androutsopoulos, 2014; Machado et al., 2017). Candidates were compared with Word2Vec vectors representing the general context and other vectors for the domain context. Those candidates closer to the general context than to the domain context were excluded from the aspect term list.

3.3.2. Hu & Liu

This is an unsupervised method, based on frequency and rules, created by Hu and Liu (2004). As with the Freq-Baseline, this method also uses a cutoff frequency to identify candidate aspects, but has two rule-based pruning mechanisms to exclude invalid candidates. The first eliminates from the list of candidates composite aspect terms less frequent than one of its components, and vice-versa. The second analyzes the distance between words in a composite aspect term: aspect terms with very distant words are also eliminated. Another difference in relation to Freq-Baseline is a last step that aims to identify infrequent aspects, through the analysis of the so-called opinion words, which are adjectives related to aspects already identified, and their proximity to other nouns.

In the same way that we did in the previous experiment, we also tested several cutoff values in order to find their best results, and we used the Word2Vec model to exclude unrelated candidates, with the difference that this mechanism was applied in a first experiment in the most frequent candidates (Pavlopoulos and Androutsopoulos, 2014) and in a second experiment on infrequent candidates (Machado et al., 2017).

3.3.3. Traditional Machine Learning

We selected some traditional machine learning algorithms in order to increase the basis for comparisons and test the limits of the methods. In all experiments, we performed a grid search for some parameters for each algorithm. The selected algorithms are:

- **Multinomial Naive Bayes (MNB):** classifier suitable for working with discrete features, as word counts (Manning et al., 2008).
- **Stochastic Gradient Descent (SGD):** simple classifier, but very efficient fit for linear models. It is particularly useful when the number of samples and/or features is very large. (Bottou, 2012).
- **Perceptron:** simple classifier suitable for large-scale learning (Freund and Schapire, 1999).
- **Passive Aggressive (PA):** family of algorithms for large-scale learning similar to Perceptron (Crammer et al., 2006).
- **Conditional Random Fields (CRF):** probabilistic graph models that consider characteristics of words and their surroundings, suitable for segmenting and labeling data sequences (Gandhi and Attar, 2020).

For the experiments, we used the bag-of-words model. In particular, for the Conditional Random Fields algorithm, we also used the neighboring words and the part of speech tags of the words obtained by the model `pt.core_news.lg`¹ from the spaCy (Honnibal et al., 2020) module.

3.3.4. BERT

The use of pre-trained models for NLP tasks has been increasing. These models are trained using large amounts of non-annotated text in simple tasks such as the prediction of next words or sentences. The BERT architecture was created to consider the contexts of the right and left words simultaneously, and it only requires several examples to perform a fine-tuning, and thus creates a model for other tasks such as named-entity recognition, polarity detection, or aspect extraction, which is our case.

In the experiments we carried out, we used two pre-trained models. The first is a multilingual model² pre-trained with texts from Wikipedia in 102 languages (Devlin et al., 2019), and the second is a model³ pre-trained with texts in Portuguese (Souza et al., 2020) from the brWaC corpus (Wagner Filho et al., 2018), composed by a large set of documents from the web.

In the fine-tuning process for the aspect extraction task, we used the BertForTokenClassification from the Transformers library by HuggingFace⁴. This model allows making predictions at the token level, rather than the sequence level, which is one of the ways in which the aspect extraction task can be handled. We used the training set texts, configured in IOB format, with the tags “B-ASP” for Begin of Aspect, “I-ASP” for Inside Aspect, and “O” for Outside Aspect. We separated 20% of the training sets for model validation. After testing with different numbers of epochs, we reached satisfactory results with 40 epochs, a learning rate of 0.0001, and a batch size of 13. For each pre-trained model, we performed fine-tuning for each domain individually and another one with data from all domains.

4. Results

All experiments were evaluated in the same way, by making predictions for the test set. The evaluation of the methods was carried out by comparing the list of predicted aspects with the list of aspects tagged in the corpora used. As metrics, we chose precision (abbreviated as Prec in the tables reporting the results) and recall (Rec), both calculated by the micro-average method, and F-measure (F1). Regarding the implicit aspects, we only calculated the percentage of aspect terms identified for each domain (Imp). In the implemented methods, the detection of implicit and explicit

¹<https://spacy.io/models/pt>

²<https://huggingface.co/bert-base-multilingual-uncased>

³<https://huggingface.co/neuralmind/bert-base-portuguese-cased>

⁴<https://huggingface.co/>

aspect terms was performed simultaneously, without the distinction between them, therefore, in cases of erroneous identification, it was not possible to identify to which category of aspect term, implicit or explicit, the error was related, thus making it impossible to calculate other metrics.

In Tables 2, 3, 4, and 5, we present the results for each domain separated by method category. In the first category of results, we have the frequency-based method Freq-Baseline (Freq), which, despite its simplicity, achieves reasonable results, except in the detection of implicit aspect terms, what may be explained because nouns and noun phrases are most often found as explicit aspect terms. Except for the domain of hotels, the mechanism for excluding unrelated aspect terms using Word2Vec brought an improvement in the results (FreqW2V).

Method	Prec	Rec	F1	Imp
Freq	0.65	0.48	0.55	0.00
FreqW2V	0.70	0.55	0.61	0.00
HuLiu	0.71	0.23	0.35	0.00
HuLiuW2V	0.71	0.23	0.35	0.00
HuLiuInfW2V	0.78	0.27	0.40	0.00
MNB	0.40	0.68	0.50	0.31
PA	0.78	0.44	0.56	0.08
Perceptron	0.81	0.45	0.58	0.08
SGD	0.78	0.44	0.56	0.08
CRF	0.81	0.61	0.69	0.23
BERT	0.57	0.53	0.55	0.23
BERT-cross	0.58	0.48	0.53	0.08
BERT-pt	0.56	0.61	0.58	0.15
BERT-pt-cross	0.62	0.54	0.57	0.08

Table 2: Results for the camera domain

Method	Prec	Rec	F1	Imp
Freq	0.74	0.51	0.61	0.12
FreqW2V	0.72	0.52	0.60	0.15
HuLiu	0.78	0.46	0.58	0.08
HuLiuW2V	0.79	0.47	0.59	0.09
HuLiuInfW2V	0.82	0.45	0.58	0.08
MNB	0.49	0.56	0.52	0.14
PA	0.98	0.64	0.77	0.18
Perceptron	0.94	0.64	0.76	0.18
SGD	0.97	0.58	0.73	0.13
CRF	0.94	0.71	0.81	0.23
BERT	0.89	0.72	0.79	0.26
BERT-cross	0.88	0.68	0.77	0.24
BERT-pt	0.90	0.70	0.78	0.25
BERT-pt-cross	0.90	0.72	0.80	0.25

Table 3: Results for the hotel domain

The second category of results presents the Hu & Liu (HuLiu) method. Despite being more complex and part of the method being similar to Freq-Baseline, there was

Method	Prec	Rec	F1	Imp
Freq	0.56	0.64	0.60	0.00
FreqW2V	0.69	0.59	0.63	0.00
HuLiu	0.68	0.44	0.53	0.00
HuLiuW2V	0.70	0.43	0.53	0.00
HuLiuInfW2V	0.77	0.43	0.55	0.00
MNB	0.36	0.75	0.49	0.20
PA	0.96	0.57	0.72	0.00
Perceptron	1.00	0.16	0.28	0.00
SGD	0.98	0.55	0.71	0.00
CRF	0.93	0.59	0.72	0.00
BERT	0.85	0.67	0.75	0.00
BERT-cross	0.85	0.62	0.72	0.00
BERT-pt	0.86	0.68	0.76	0.00
BERT-pt-cross	0.86	0.63	0.73	0.20

Table 4: Results for the book domain

Method	Prec	Rec	F1	Imp
Freq	0.45	0.52	0.48	0.04
FreqW2V	0.56	0.47	0.51	0.00
HuLiu	0.74	0.33	0.46	0.08
HuLiuW2V	0.64	0.30	0.41	0.08
HuLiuInfW2V	0.75	0.28	0.41	0.04
MNB	0.29	0.34	0.31	0.12
PA	0.78	0.43	0.56	0.25
Perceptron	0.78	0.46	0.58	0.29
SGD	0.78	0.41	0.54	0.25
CRF	0.83	0.46	0.59	0.25
BERT	0.62	0.47	0.54	0.17
BERT-cross	0.65	0.48	0.55	0.17
BERT-pt	0.57	0.49	0.53	0.29
BERT-pt-cross	0.67	0.53	0.59	0.12

Table 5: Results for the smartphone domain

a drop in the results. This drop was because the dataset has sentences that do not contain aspect terms. In its last step, the method searches for infrequent aspects, which caused the detection of aspects in sentences where they did not exist. The pruning mechanism with Word2Vec (HuLiuW2V) resulted in an improvement in precision, and it proved to be more efficient when applied to infrequent aspects (HuLiuInfW2V). Implicit aspect detection was not satisfactory for the same reason as Freq-Baseline.

In the third group of results, we have some traditional machine learning methods. The Multinomial Naive Bayes (MNB) classifier got the lowest results, deserving only one caveat regarding the detection of implicit aspects by finding aspects in all domains, reaching the best results in the domains of cameras and books and being the only method of the group to find implicit aspect terms in the domain of books.

The PassiveAggressive (PA) and Stochastic Gradient Descent (SGD) classifiers got reasonable results and,

in general, close to the best results in all domains. The Perceptron classifier was the only one that was not consistent in all domains, achieving a very bad result for the domain of books.

The Conditional Random Fields (CRF) classifier achieved the best results in terms of f-measure in all domains, except for books, even though it was close to the best result in this case. This was expected because the algorithm is the most appropriate for tasks that involve sequential data labeling, which is the case of the extraction of implicit aspect terms, and secondly because we used a different set of features in the training, as explained in Section 3.3.3.

The fourth and last group presents the results got with the use of the pre-trained language model – BERT. As explained in Section 3.3.3, models were fine-tuned from a multilingual model (Devlin et al., 2019) and from a Portuguese model (Souza et al., 2020), creating one model for each domain and one cross-domain model with data from all domains (referenced by “pt” and “cross”, respectively, in the tables of results). The results were good, reaching the best f-measures in the domains of books and smartphones, and being slightly lower only in the domain of cameras. The Portuguese models were a little better than the multilingual ones. Regarding the cross-domain experiments, only for the domain of smartphones, there was a more significant difference in relation to the test by domain. In the detection of implicit aspect terms, the difficulty was in the domain of books, and only the Portuguese model fine-tuned with cross-domain data could detect some aspect terms.

Tables 6, 7, 8, and 9 present the results regarding the typology of implicit aspect clues presented in Section 3.2. The number of identified aspects of each type was grouped independently of their domains, since the analysis carried out here should analyze the effectiveness of the methods regardless of the domain. Each table presents the data of a different category. In its first lines, referenced by the word “Aspects”, it presents the number of aspects present in all corpora for each subcategory and, in its last column, the total of the category (as the datasets are manually annotated according to the types of the implicit aspect clues). In the subsequent lines, the table shows the results of the methods grouped by the method category, as in the previous set of tables. Each column of the table presents the number of aspect terms identified by the method, and, in the last column, the total identified for the category.

The Freq-Baseline (Freq) and Hu & Liu (HuLiu) methods, in general, did not identify aspect terms related to different grammatical classes of nouns, which was expected, since these methods only classify as aspects nouns and noun phrases. We can verify this in the Verb subcategory of Table 6 and Attribute (Att) of Table 7. The subcategories Location and Related in Table 9 were not identified by the methods, for the same reason as before, and also for the number of compound

Method	Non-verbal	Verb	Total
Aspects	12	45	57
Freq	2	0	2 (3%)
FreqW2V	2	0	2 (3%)
HuLiu	2	0	2 (3%)
HuLiuW2V	3	0	3 (5%)
HuLiuInfW2V	2	0	2 (3%)
MNB	2	2	4 (7%)
PA	2	2	4 (7%)
Perceptron	2	2	4 (7%)
SGD	2	2	4 (7%)
CRF	3	3	6 (10%)
BERT	2	5	7 (12%)
BERT-cross	2	3	5 (8%)
BERT-pt	2	5	7 (12%)
BERT-pt-cross	2	3	5 (8%)

Table 6: Number of aspects detected from the subcategories of the Event category.

Method	Att	Eq	Is-a	Part	Total
Aspects	10	84	55	124	273
Freq	0	12	4	37	53 (19%)
FreqW2V	0	13	4	45	62 (23%)
HuLiu	0	9	5	30	44 (16%)
HuLiuW2V	0	9	6	32	47 (17%)
HuLiuInfW2V	0	9	6	31	46 (17%)
MNB	2	27	6	41	76 (28%)
PA	0	26	7	36	69 (25%)
Perceptron	0	16	7	49	72 (26%)
SGD	0	9	6	31	46 (17%)
CRF	0	29	9	54	92 (33%)
BERT	0	29	9	53	91 (33%)
BERT-cross	0	29	9	41	79 (29%)
BERT-pt	1	29	8	40	78 (29%)
BERT-pt-cross	0	28	7	51	86 (31%)

Table 7: Number of aspects detected from the subcategories of the Feature category.

terms present in these subcategories, which consists in another weakness of the methods. Strangely, the Hu & Liu method identified some aspect terms from the Adjectives (Adj) subcategory of Table 8, probably due to part of speech tagging errors. The same did not happen with the Freq-Baseline. Finally, it is worth noting the Nominal subcategory of Table 8: despite being related to nouns, the number of aspects of the subcategory is small, a fact that makes identification difficult given the frequency-based nature of both methods.

The machine learning methods Multinomial Naive Bayes (MNB), PassiveAggressive (PA), Perceptron, and Stochastic Gradient Descent (SGD) had similar results, superior to frequency-based and inferior to Conditional Random Fields (CRF) and BERT. In the Attribute (Att) subcategory of Table 7, only MNB was able to find 2 aspects, a fact that can be explained

Method	Adj	Eq	Nominal	Total
Aspects	107	10	10	127
Freq	0	1	1	2 (1%)
FreqW2V	0	1	1	2 (1%)
HuLiu	10	1	0	11 (8%)
HuLiuW2V	9	1	0	10 (8%)
HuLiuInfW2V	7	1	0	8 (6%)
MNB	38	1	1	40 (31%)
PA	44	1	1	46 (36%)
Perceptron	45	1	1	47 (37%)
SGD	41	1	1	43 (34%)
CRF	44	1	1	46 (36%)
BERT	42	2	1	45 (35%)
BERT-cross	48	2	1	51 (40%)
BERT-pt	46	2	1	49 (38%)
BERT-pt-cross	47	2	1	50 (39%)

Table 8: Number of aspects detected from the subcategories of the Qualification category.

Method	Location	Related	Total
Aspects	36	68	104
Freq	0	0	0 (0%)
FreqW2V	0	0	0 (0%)
HuLiu	0	1	1 (1%)
HuLiuW2V	0	1	1 (1%)
HuLiuInfW2V	0	1	1 (1%)
MNB	0	3	3 (3%)
PA	0	4	4 (4%)
Perceptron	0	3	3 (3%)
SGD	0	2	2 (2%)
CRF	5	13	18 (17%)
BERT	6	12	18 (17%)
BERT-cross	6	8	14 (13%)
BERT-pt	8	11	19 (18%)
BERT-pt-cross	6	9	15 (14%)

Table 9: Number of aspects detected from the subcategories of the Contextual category.

by the small number of examples in this subcategory (only 10 aspect terms). In the Location subcategory of Table 9, aspect terms were also not found. Despite having a slightly larger number of examples, there are many compound terms, which end up being difficult to be identified by methods using bag-of-word features, since they analyze each word in isolation.

Finally, the Conditional Random Fields (CRF) and BERT methods were the ones that got the best results. Only the Attribute (Att) from Table 7 had no localized aspects, probably because of the low number of examples. It is worth highlighting the Contextual category on Table 9, which theoretically includes the most difficult aspects to be identified due to the need for additional world knowledge, and even so the methods could identify a significant number of these aspect terms.

Finally, we analyze the aspect terms by checking the

number of methods that identified them, in order to discover the most difficult ones. Table 10 presents these results, where each term is accompanied by the total number of methods that found it in parentheses, except for those that were found by only one method.

Camera: “compacta” (4), “versátil” (2), “medidas” (2), “facilidade de uso”, “fácil de usar”, “facil de usar”, “volume”
Hotel: “localização” (15), “limpeza” (15), “chuveiro” (15), “elevador” (15), “banheiro” (15), “atendimento” (15), “carpete” (14), “apartamento” (14), “banheiros” (13), “sujo” (13), “estacionamento” (13), “quarto” (13), “transfer” (13), “recepcionista” (13), “café” (12), “proximo” (12), “aquecedor” (11), “box” (11), “proximidade” (11), “cama” (11), “atendentes” (10), “piscina” (9), “limpo” (9), “cozinha” (9), “perto” (9), “próximo” (9), “telefone” (9), “wi-fi” (9), “limpas” (8), “ar condicionado” (8), “apartamentos” (8), “custo” (8), “barulho” (8), “limpos” (8), “valor” (7), “localização” (7), “estrutura” (7), “localizado” (7), “sujos” (7), “coberto” (7), “receptionistas” (7), “comida” (7), “caro” (7), “cheiro de mofo” (6), “barato” (6), “localizado no centro” (6), “no centro” (5), “encontra-se estrategicamente dentro do centro financeiro” (5), “chiqueiro” (5), “sujeira” (4), “cheirando a mofo” (4), “prédio” (4), “cobertores” (3), “imunda” (3), “imundas” (3), “room service” (3), “fica bem no centro” (3), “infraestrutura” (3), “cheiro” (2), “silencio” (2), “baratinho” (2), “manchada” (2), “torneira” (2), “cortina” (2), “barata” (2), “janta” (2), “telefonar” (2), “estruturar” (2), “quartar” (2), “custar” (2), “cozinhar” (2), “bem no centro”, “roupa de cama”, “barulhentos”, “room-service”, “box do banheiro”, “barulhento”, “comido”, “academias”, “mofar”, “cheirar”, “barulhar”, “tomar”, “jantar”, “cobrir”, “lixar”, “higiene”, “cobertor”, “wifi”, “cortinas”, “lixo”, “dormir”
Livro: “mocinha”, “escreve”
Smartphone: “prático” (9), “trava” (8), “pequeno” (7), “bonito” (6), “moderno” (6), “fácil de mexer” (4), “leve” (3), “congela” (2), “restarta”, “lindo”, “operação”, “lento”, “pesado”

Table 10: Aspects and the number of methods that identified them.

Analyzing the table, we found some interesting terms that were identified as aspect terms: “facilidade de uso”, “fácil de usar”, and “fácil de mexer” (all of them meaning “easy to use”), all related to the usability aspect. If the algorithm had identified only one word of each expression, it would not be possible to identify the aspect, since both the verb “usar” (to use) and the

adjective “*fácil*” (easy) appear in other contexts not related to usability.

Some aspects of the subcategory Location of the Contextual category were also identified, with some expressions being quite complex, such as: “*encontra-se estrategicamente dentro do centro financeiro*” (it is strategically within the financial center), and “*fica bem no centro*” (it’s right in the center). Other aspect terms, even with typos, could be found, like “*facil de usar*” with missing accents (easy to use), and “*academiar*” and “*barulhar*”, probable typing errors where the correct forms should be “*academia*” (gym), and “*barulho*” (noise).

5. Final Remarks

In this work, we analyzed the results of different methods for aspect term extraction. In a first analysis, we used metrics commonly found in the literature. The results were partly as expected, with machine learning methods performing better than the unsupervised Freq-Baseline and Hu & Liu. The latter attracted attention negatively, as it was expected to achieve a result at least higher than the Freq-Baseline. As explained before, this was probably because of the dataset containing phrases without aspect terms, which would theoretically help some machine learning methods. In real world applications, therefore, we conclude that it is necessary to analyze the subjectivity of sentences, removing those that do not contain opinion, before applying the Hu & Liu method.

Machine learning methods with a bag-of-words model achieved satisfactory results. They supposedly are not the most appropriate for the task, but they have the advantage of being lightweight and easily implemented. The CRF algorithm achieved good results and, despite not being as light and easy to implement as the previous ones, it wins in both aspects when compared to BERT. This one also got good and promising results, given that there are many model options to be used together. The only problem is really the consumption of processing and memory resources that are necessary for the fine-tuning and use of the model.

In our second analysis, we used the typology of implicit aspect clues (Machado et al., 2022) to analyze in more details the results produced by each method. As mentioned before, papers that address implicit aspects are relatively rare, and, as far as we know, this was the first time that an analysis of this type was carried out. The typology easily provided a view of the strengths and weaknesses of each method.

Based on the results, we can conclude, for example, that adding rules related to verbs or adjectives in Hu & Liu method might lead to an improvement in the results. The Feature category of aspects seems promising for increasing results, given that it is not as complex as the Contextual category and has a significant number of undetected aspect terms. A suggestion would be the identification and use of relations (equivalence,

is-a, and part-of) as features for the classifiers. The Contextual category, despite having a low percentage of aspects detected, requires diverse and world knowledge to identify its aspects, which is more complicated to be computationally achieved.

Future work includes to implement these suggestions and others based on this study, and thus advance the state of the art in extracting aspect terms.

To the interested readers, more information and source codes related to the performed experiments may be found in our repository at GitHub⁵ or at the web portal of the POeTiSA project (*Portuguese processing - Towards Syntactic Analysis and parsing*)⁶.

Acknowledgements

The authors are grateful to the Center for Artificial Intelligence (C4AI), sponsored by IBM Corporation and São Paulo Research Foundation (grant #2019/07665-4).

6. Bibliographical References

- Bottou, L. (2012). Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pages 421–436. Springer Berlin Heidelberg.
- Cai, H., Tu, Y., Zhou, X., Yu, J., and Xia, R. (2020). Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 833–843, Barcelona, Spain. International Committee on Computational Linguistics.
- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. (2006). Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 7(19):551–585.
- Cruz, I., Gelbukh, A. F., and Sidorov, G. (2014). Implicit aspect indicator extraction for aspect based opinion mining. *International Journal of Computational Linguistics and Applications*, 5(2):135–152.
- Freund, Y. and Schapire, R. E. (1999). Large margin classification using the perceptron algorithm. *Machine learning*, 37(3):277–296.
- Gandhi, H. and Attar, V. (2020). Extracting aspect terms using crf and bi-lstm models. *Procedia Computer Science*, 167:2486–2495.
- Ganganwar, V. and Rajalakshmi, R. (2019). Implicit aspect extraction for sentiment analysis: A survey of recent approaches. *Procedia Computer Science*, 165:485–491.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining*, pages 168–177, New York, United

⁵<https://github.com/mtarcinalli/LREC-Extraction-of-Aspect-Terms>

⁶<https://sites.google.com/icmc.usp.br/poetisa>

- States of America. Association for Computing Machinery.
- Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*, pages 627–666. Chapman & Hall.
- Lopes, É., Correa, U., and Freitas, L. (2021). Exploring bert for aspect extraction in portuguese language. In *The International FLAIRS Conference Proceedings*, volume 34.
- Machado, M. T., Pardo, T. A. S., and Ruiz, E. E. S. (2017). Analysis of unsupervised aspect term identification methods for portuguese reviews. In *Proceedings of the 14th National Meeting on Artificial and Computational Intelligence*, pages 239–249, Uberlândia, Brazil. Brazilian Computing Society.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Panchendrarajan, R., Ahamed, N., Murugaiah, B., Sivakumar, P., Ranathunga, S., and Pemasiri, A. (2016). Implicit aspect detection in restaurant reviews using cooccurrence of words. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 128–136, San Diego, United States of America. Association for Computational Linguistics.
- Pavlopoulos, J. and Androutsopoulos, I. (2014). Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method. In *Proceedings of the 5th Workshop on Language Analysis for Social Media*, pages 44–52, Gothenburg, Sweden. Association for Computational Linguistics.
- Pereira, D. A. (2021). A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*, 54(2):1087–1115.
- Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
- Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 486–495, Denver, United States of America, June. Association for Computational Linguistics.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryiğit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, pages 19–30, San Diego, United States of America. Association for Computational Linguistics.
- Poria, S., Cambria, E., Ku, L.-W., Senticnet, C. G., and Gelbukh, A. (2014). A Rule-Based Approach to Aspect Extraction from Product Reviews. In *Proceedings of the 2nd Workshop on Natural Language Processing for Social Media*, pages 28–37, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.
- Rana, T. A. and Cheah, Y.-N. (2016). Aspect extraction in sentiment analysis: comparative analysis and survey. *Artificial Intelligence Review*, 46(4):459–483.
- Ravi, K. and Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46.
- Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2(1):325–347.
- Tubishat, M., Idris, N., and Abushariah, M. A. (2018). Implicit aspect extraction in sentiment analysis: Review, taxonomy, opportunities, and open challenges. *Information Processing & Management*, 54(4):545–563.
- Zhang, Y. and Zhu, W. (2013). Extracting implicit features in online customer reviews for opinion mining. In *Proceedings of the 22nd International World Wide Web Conference*, pages 103–104, Rio de Janeiro, Brazil. Association for Computing Machinery.

7. Language Resource References

- de Freitas A, L. A. and Vieira, R. (2015). Exploring resources for sentiment analysis in portuguese language. In *Proceedings of the 4th Brazilian Conference on Intelligent Systems*, pages 152–156, Natal, Brazil. IEEE Computer Society.
- 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 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Freitas, C., Motta, E., Milidiú, R. L., and César, J. (2012). Vampiro Que Brilha... Rá! Desafios Na Anotação De Opinião Em Um Corpus De Resenhas De Livros. In *Proceedings of the 11th Corpus Linguistics Meeting*, São Paulo, Brazil.
- Hartmann, N., Avanço, L., Balage Filho, P. P., Duran, M. S., Nunes, M. D. G. V., Pardo, T. A. S., Aluísio, S. M., et al. (2014). A large corpus of product reviews in portuguese: Tackling out-of-vocabulary words. In *Proceedings of the 9th Language Re-*

- sources and Evaluation Conference*, pages 3865–3871, Reykjavik, Iceland. European Language Resources Association.
- Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.
- Machado, M. T., Pardo, T. A. S., Ruiz, E. E. S., Fellippo, A. D., and Vargas, F. (2022). Implicit opinion aspect clues in portuguese texts: Analysis and categorization. In *Proceedings of the 15th International Conference on Computational Processing of the Portuguese*, pages 68–78, Fortaleza, Brazil. Springer.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 2, page 3111–3119. Curran Associates Inc.
- 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*, pages 403–417. Springer.
- Vargas, F. A. and Pardo, T. A. S. (2017). Clustering and hierarchical organization of opinion aspects: a corpus study. In *Proceedings of the 14th Corpus Linguistics Meeting*, pages 342–351, São Leopoldo, Brazil.
- Wagner Filho, J. A., Wilkens, R., Idiart, M., and Villavicencio, A. (2018). The brwac corpus: A new open resource for brazilian portuguese. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pages 4339–4344, Miyazaki, Japan. European Language Resources Association.