UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Methods for identifying aspects in opinion texts in Portuguese: the case of implicit aspects and their typological analysis

Mateus Tarcinalli Machado

Doctoral Thesis of the Postgraduate Program in Computer Science and Computational Mathematics (PPG-CCMC)

ICMC USP
SÃO CARLOS

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura:______________________

Mateus Tarcinalli Machado

Methods for identifying aspects in opinion texts in Portuguese: the case of implicit aspects and their typological analysis

Doctoral dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. FINAL VERSION

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Thiago Alexandre Salgueiro Pardo

USP – São Carlos
December 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

T179m

Tarcinalli Machado, Mateus
Methods for identifying aspects in opinion texts in Portuguese: the case of implicit aspects and their typological analysis / Mateus Tarcinalli Machado; orientador Thiago Alexandre Salgueiro Pardo. -- São Carlos, 2023.
143 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. Mineração de opinião. 2. Identificação de aspectos. 3. Aspectos explícitos. 4. Aspectos implícitos. I. Alexandre Salgueiro Pardo, Thiago, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Mateus Tarcinalli Machado

Métodos de identificação de aspectos em textos de opinião em português: o caso dos aspectos implícitos e sua análise tipológica

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. VERSÃO REVISADA

Área de Concentração: Ciências de Computação e Matemática Computacional
Orientador: Prof. Dr. Thiago Alexandre Salgueiro Pardo

USP – São Carlos
Dezembro de 2023

This work is dedicated to my wife Vanessa and my children Miguel, Ana Helena, and Cloe sources of my inspiration.

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).

ABSTRACT

MACHADO, M. T. **Methods for identifying aspects in opinion texts in Portuguese: the case of implicit aspects and their typological analysis**. 2023. 143p. Thesis (Doctorate in Science) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

This work has as object of study the methods for identifying aspects, which are applications derived from the aspect-based sentiment analysis, and the area of natural language processing. The aspect-based sentiment analysis is focused on analyzing opinionated texts (texts containing opinions) seeking to identify and relate feelings and aspects (characteristics) of a given entity (products, services, among others). This work is focused on the first stage, which is the identification of aspects, emphasizing the so-called implicit aspects, that is, those that are not explicitly mentioned in the texts. In the course of this work, we implemented, adapted, and improved several aspect extraction methods, including frequency-based, rule-based, hybrid, machine learning, pre-trained language models, and large language models. We also developed novel resources such as a corpus extension with annotation of implicit aspects, lexicons of nominal forms, and a typology of implicit aspects. The latter allowed a better understanding of the knowledge necessary to relate an implicit aspect clue with an aspect and its use allowed a clear view of the strengths and weaknesses of each implemented method in relation to the detection of implicit aspects.

Keywords: Opinion mining. Identification of aspects. Explicit aspects. Implicit aspects.

RESUMO

MACHADO, M. T. **Métodos de identificação de aspectos em textos de opinião em português: o caso dos aspectos implícitos e sua análise tipológica**. 2023. 143p. Tese (Doutorado em Ciências) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Este trabalho tem como objeto de estudo os métodos para identificação de aspectos, que são aplicações derivadas da análise de sentimentos baseada em aspectos e da área de processamento de linguagem natural. A análise de sentimentos baseada em aspectos é focada em analisar textos avaliativos (textos contendo opiniões) buscando identificar e relacionar sentimentos e aspectos (características) de uma determinada entidade (produtos, serviços entre outros). Este trabalho está focado na primeira etapa, identificação de aspectos, dando ênfase aos chamados aspectos implícitos, ou seja, aqueles que não são mencionados explicitamente nos textos. No decorrer desse trabalho implementamos, adaptamos e melhoramos vários métodos de extração de aspectos, entres eles métodos baseados em frequências, baseados em regras, híbridos, aprendizado de máquina, modelos de linguagem pré-treinados e grandes modelos de linguagem. Também desenvolvemos novos recursos como uma extensão de corpus com anotação de aspectos implícitos, léxicos de formas nominais e uma tipologia de aspectos implícitos. Esta última permitiu um melhor entendimento dos conhecimentos necessários para relacionar uma pista de aspecto implícito com um aspecto e seu uso possibilitou uma visão clara dos pontos fortes e fracos de cada método implementado com relação à detecção de aspectos implícitos.

Palavras-chave: Mineração de opinião. Identificação de aspectos. Aspectos explícitos. Aspectos implícitos.

LIST OF FIGURES

Figure 1 – Sentiment analysis taxonomy 24
Figure 2 – Phone aspects taxonomy example 29
Figure 3 – Visualization of the results of two sentiment analysis systems 33
Figure 4 – Convolution Example. 35
Figure 5 – OPINE system overview 41

LIST OF TABLES

Table 1 – Levels of sentiment analysis 22
Table 2 – Hu and Liu (2004) results 40
Table 3 – Popescu and Etzioni (2007) results 42
Table 4 – Pavlopoulos and Androutsopoulos (2014) results 43
Table 5 – Panchendrarajan *et al.* (2016) results 45
Table 6 – SemEval 2014 corpus 45
Table 7 – Poria *et al.* (2014) results 48
Table 8 – Costa and Pardo (2020) results 50
Table 9 – Balage Filho and Pardo (2014) results 51
Table 10 – Corpus Cruz, Gelbukh and Sidorov (2014) properties 51
Table 11 – Corpus Cruz, Gelbukh and Sidorov (2014) statistics 52
Table 12 – Distribution of PoS tags in the Cruz, Gelbukh and Sidorov (2014) corpus 52
Table 13 – Cruz, Gelbukh and Sidorov (2014) results 54
Table 14 – Implicit aspects of the SemEval 2015 and 2016 datasets. 54
Table 15 – Results of sentiment analysis based on aspect categories 56
Table 16 – Comparison with state-of-the-art approaches 58
Table 17 – Corpus details 58
Table 18 – Content covered in related works. 59

CONTENTS

1 INTRODUCTION . . . 21
1.1 Contextualization . . . 21
1.2 Objectives . . . 25
1.3 Hypothesis . . . 26
1.4 Thesis organization . . . 26

2 THEORETICAL FOUNDATION . . . 27
2.1 Aspect-based sentiment analysis . . . 27

3 RELATED WORK . . . 39
3.1 Hybrid Methods . . . 39
3.1.1 Hu and Liu (2004) . . . 39
3.1.2 Popescu and Etzioni (2007) . . . 40
3.1.3 Pavlopoulos and Androutsopoulos (2014) . . . 42
3.1.4 Panchendrarajan *et al.* (2016) . . . 44
3.2 Rule-based methods . . . 45
3.2.1 Poria *et al.* (2014) . . . 45
3.2.2 Costa and Pardo (2020) . . . 48
3.3 Machine learning methods . . . 49
3.3.1 Balage Filho and Pardo (2014) . . . 49
3.3.2 Cruz, Gelbukh and Sidorov (2014) . . . 51
3.4 Deep learning methods . . . 53
3.4.1 Cai *et al.* (2020) . . . 53
3.4.2 Wang *et al.* (2021) . . . 56
3.5 Other related works . . . 57
3.5.1 Vargas and Pardo (2018) . . . 57
3.6 Literature synthesis . . . 58

4 LEARNING RULES FOR AUTOMATIC IDENTIFICATION OF IMPLICIT ASPECTS IN PORTUGUESE . . . 61

5 IMPLICIT OPINION ASPECT CLUES IN PORTUGUESE TEXTS: ANALYSIS AND CATEGORIZATION . . . 73

6 EVALUATING METHODS FOR EXTRACTION OF ASPECT TERMS IN OPINION TEXTS IN PORTUGUESE – THE CHALLENGES OF IMPLICIT ASPECTS . . . 85

7 EVALUATING FROM THE CLASSICAL FREQUENCY TO THE LARGE LANGUAGE MODELS-BASED METHODS FOR ASPECT EXTRACTION IN PORTUGUESE AND THEIR PERFORMANCE ON DETECTING TYPOLOGICALLY DIFFERENT IMPLICIT ASPECTS . . . 97

8 CONCLUSION . . . 117

REFERENCES . . . 121

APPENDIX 131

APPENDIX A – NILC AT ABSAPT 2022: ASPECT EXTRACTION FOR PORTUGUESE . . . 133

1 INTRODUCTION

1.1 Contextualization

Sentiment analysis is an area of research in applied computing related to natural language processing that aims to analyze people's opinions, sentiments, evaluations, attitudes, and emotions in relation to entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It features a wide variety of studies and is often referred to in the literature by slightly different names and tasks, such as opinion mining, opinion extraction, sentiment mining, and subjectivity analysis, among others. In industry, the term sentiment analysis is more commonly used, while in academia, the area is also often referred to as opinion mining (LIU, 2012). According to Taboada (2016), sentiment analysis is a growing field located at the intersection of linguistics and computer science that aims to automatically determine sentiments contained in a text.

This curiosity about other people's opinions is probably as old as verbal communication itself (MÄNTYLÄ; GRAZIOTIN; KUUTILA, 2018). We have observed several times in history that leaders question their subordinates about their opinions to prepare them to face opposition or increase their popularity. Examples of attempts to detect internal dissent can be found as far back as ancient Greek times (RICHMOND, 1998). Ancient works from the East and West also address these issues. In "The Art of War", for example, we have an entire chapter that deals with espionage, recruiting and betraying spies. At the beginning of the "Iliad", the leader of the Greeks, Agamemnon, tries to assess the fighting spirit of his men. In the fifth century BC in Athens, voting appeared as a method of measuring opinion on public policy (THORLEY, 2004). The use of questionnaires emerged in the first decades of the 20th century (DROBA, 1932) and in 1937 the journal Public Opinion Quarterly [1] was created.

Natural language processing research has a long history in computer science. As for sentiment analysis, few studies were conducted before the year 2000. In recent years, there has been a large increase in papers produced focusing on this area, as we can see in the work of Mäntylä, Graziotin and Kuutila (2018) who performed an analysis of the evolution of the area and found around 7,000 published papers, 99% of which appeared after 2004, demonstrating how fast and recent the growth of the area is.

The first academic studies aimed at identifying public opinion emerged during the Second World War and were politically motivated (STAGNER, 1940; KNUTSON, 1945). The emergence of modern sentiment analysis occurred in the mid-2000s and initially focused on product reviews available on the Web (DAVE; LAWRENCE; PENNOCK,

[1] http://poq.oxfordjournals.org/

2003). Since then, the use of sentiment analysis has expanded into numerous other areas, such as forecasting financial markets (NASSIRTOUSSI *et al.*, 2014) and reactions to terrorist attacks (BURNAP *et al.*, 2014). Regarding emotions, research is advancing from simple polarity detection to more complex nuances, for example, differentiating negative emotions such as anger and grief (CAMBRIA *et al.*, 2015).

Some authors (MEDHAT; HASSAN; KORASHY, 2014; RANA; CHEAH, 2016; YADOLLAHI; SHAHRAKI; ZAIANE, 2017) divide sentiment analysis into three levels: document, sentence, and aspect. Each of these levels aims to address distinct challenges, with aspect-based analysis being the most refined. Table 1 presents these levels and examples of analysis.

Table 1 – Levels of sentiment analysis

Sentiment analysis at document level		
Text		Sentiment
Canon T3 is a great camera. It has excellent image quality despite the slightly expensive price.		positive
Sentiment analysis at sentence level		
Text		Sentiment
Canon T3 is a great camera.		positive
Sentiment analysis at aspect level		
Text	Aspect	Sentiment
It has excellent image quality despite the slightly expensive price.	image quality	positive
	price	negative

At the document level, the aim is to identify opinion words in a text to determine the polarity of the document in its entirety and whether its overall opinion is positive or negative. In the example presented in Table 1, we can see that in the document we have two positive opinions (about the camera and image quality) and one negative opinion (about the price), but overall, the document is positive. This level of analysis is important in social and psychological studies conducted on social media, consumer satisfaction, patient analysis in medical settings, and numerous others (YADOLLAHI; SHAHRAKI; ZAIANE, 2017).

Sentiment analysis performed at the sentence level aims to find the polarity of a sentence. At this level of analysis, it is important to identify whether the target sentence is subjective or objective, whether it contains some form of opinion, and only after this process, analyze the sentences in search of their polarity. In the example presented in Table 1, the positivity of the sentence is clear, on the other hand, if we analyze the second sentence of the document, this identification would not be so simple given that it contains a positive and a negative feeling. This level of sentiment analysis is typically influenced by the context surrounding the sentence and is considered especially important for applications dealing with tweets, Facebook posts, and short messages, among others.

Finally, aspect level is a refined sentiment analysis model that deals with determining people's opinions about a specific feature (aspect) of a product, service, or any entity (MEDHAT; HASSAN; KORASHY, 2014; RANA; CHEAH, 2016; YADOLLAHI; SHAHRAKI; ZAIANE, 2017). To perform this level of analysis, it is essential to identify the entities (products, services, etc.) mentioned and their respective aspects (characteristics of that entity). From there, the polarity of opinions directed towards each identified aspect is determined, as we can see in the example presented in Table 1, where we have two aspects related to two sentiments. The results of the aspect-level sentiment analysis can then finally be summarized and visualized, for example, in a table containing the aspects found and their respective polarities (RANA; CHEAH, 2016). This level of processing can also be called aspect-based sentiment analysis (LIU, 2012).

Some authors who study aspect-based sentiment analysis divide the problem into three tasks (PAVLOPOULOS; ANDROUTSOPOULOS, 2014):

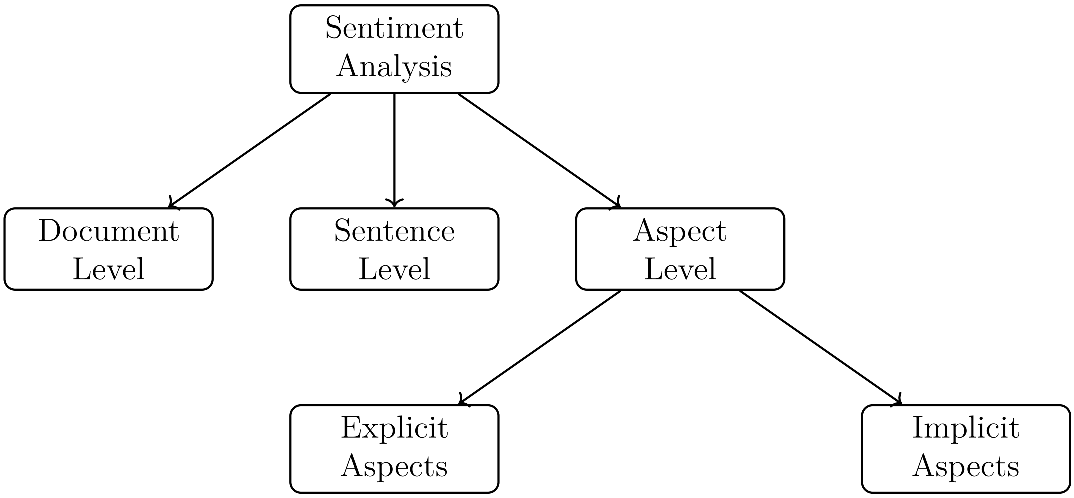
- **Extraction of aspect terms**: from texts about a specific entity or entities of the same type (analyses of opinion texts about a specific phone model or any phone model). This stage extracts and possibly sorts the found aspect terms ("battery", "screen") from the target entity, including multi-word terms ("hard disk") (LIU, 2012; LONG; ZHANG; ZHUT, 2010; SNYDER; BARZILAY, 2007; YU *et al.*, 2011). At the end of this stage, each aspect term is considered the name of a different aspect. The found aspect terms can be subsequently grouped together.

- **Aspect term sentiment estimation**: This step estimates the polarity and possibly also the intensity (strongly negative, moderately positive) of opinions for each aspect term of the target entity, usually calculated across multiple texts. Sorting texts by sentiment polarity is a popular research topic (LIU, 2012; PANG; LEE, 2005; TSYTSARAU; PALPANAS, 2012). The aim of this step of aspect-based sentiment analysis is to estimate the polarity and intensity of opinions about each aspect term of the target entity.

- **Aspect aggregation**: some systems group aspect terms that are synonymous or close ("price", "cost") or, group aspect terms to obtain aspects of a more general granularity ("chicken", "steak" and "fish" can be replaced with "food") (LIU, 2012; LONG; ZHANG; ZHUT, 2010; ZHAI *et al.*, 2010; ZHAI *et al.*, 2011). A polarity (and intensity) score can then be calculated for each more general aspect ("food") by combining (averaging) the polarity scores of the aspect terms that belong to the most general aspect.

Two categories of opinions may appear in texts: regular opinions and comparative opinions. A regular opinion expresses a sentiment about an entity or aspect of that entity,

Figure 1 – Sentiment analysis taxonomy



for example, “Coca-Cola tastes good” presents a positive feeling about the flavor aspect of the entity “Coca-Cola”. A comparative opinion, on the other hand, presents several entities and compares some of their aspects, for example, the sentence “Coke tastes better than Pepsi” compares the entities “Coca-Cola” and “Pepsi” based on their flavor and expresses a preference for “Coca-Cola” (LIU, 2012). This category of opinion presents an even greater challenge for performing aspect-based sentiment analysis, as the polarity of the aspect may not be clearly specified and there is more than one entity involved.

Another characteristic that makes identification difficult is the way the aspect appears in a sentence. An aspect is considered explicit if it is mentioned directly in a sentence, otherwise, it is considered implicit (LIU, 2012). In the sentence “The image quality of this camera is great”, the expression “image quality” presents an explicit aspect. On the other hand, in “This camera is expensive”, the word “expensive” implicitly refers to the “price” aspect. Many expressions related to implicit aspects are formed by adjectives and adverbs that are used to describe or qualify the aspects, for example, “expensive” (“price”) and “reliable” (“reliability”). They can also be verbs and verb phrases, for example, in the sentence “I can install the software easily”, the verb “install” refers to the “installation” aspect. They can also appear in more complex forms, as in the sentence “This camera doesn’t easily fit in a coat pocket”, where “fits in a coat pocket” indicates the “size” aspect.

According to Zhang and Zhu (2013), around 30% of Chinese reviews contain implicit aspects. In another work (PANCHENDRARAJAN *et al.*, 2016), the authors analyze that more than 15% of the sentences in their corpus contained one or more implicit aspects and that 92% of aspects related to restaurant employees were mentioned implicitly. Most methods for extracting aspects focus only on those explicitly mentioned. For example, in Ravi and Ravi (2015) the authors evaluated 160 papers related to the area of sentiment analysis, of which only 23 were related to aspect extraction and only 4 analyzed the implicit ones. In other research (RANA; CHEAH, 2016), 45 surveys related to aspect extraction were evaluated. In this case, only 7 papers analyzed the issue of implicit aspects.

This minority of works that deal with the problem of implicit aspects is related to the greater frequency of explicit aspects, and sometimes, depending on the application, some loss of content related to the non-identification of implicit aspects may be acceptable. Another reason is related to the greater difficulty in implementing methods for extracting implicit aspects.

Another principal issue to be addressed is language. Most works analyze texts only in English or Chinese. In the works of Tubishat, Idris and Abushariah (2018) and Ganganwar and Rajalakshmi (2019) 53 articles dealing with implicit aspects were analyzed, of which 33 were applied to texts in English, 19 in Chinese, and 1 in French. In the Portuguese language, therefore, the number of papers is very restricted. According to Pereira (2020) that analyzed only papers on sentiment analysis for the Portuguese language, we have only 2 datasets that present tagged implicit aspects Freitas and Vieira (2015) and Vargas and Pardo (2018) and only the second work presents methods related to implicit aspects.

1.2 Objectives

Despite the evident lack of studies that deal with implicit aspects and the Portuguese language, these facts do not mean that the problem is not relevant. As we mentioned, in several sets of data, the authors point out that between 15% and 30% of the aspects are mentioned implicitly, and for some categories, these percentages increase significantly. These are considerable amounts, and the non-detection of these aspects negatively impacts the results and often leads to the loss of valuable information. The identification of implicit aspects and the treatment of texts in Portuguese, despite being complex problems, should enrich sentiment analysis applications, thus advancing the state of the art in the area. Therefore, the objectives of this work are:

- Evaluate existing aspect extraction methods for other languages and verify their applicability to Portuguese;
- Adapt existing methods to improve the extraction of implicit aspects;
- To develop new methods for extracting aspects and, in particular, to advance in the identification of implicit aspects;
- Develop new linguistic resources that enable a better understanding of implicit aspects.

In addition to methods already considered classic, one should focus on the development and use of pre-trained models and large language models for the task since these models have allowed for better results in several natural language processing tasks.

1.3 Hypothesis

From the presented gaps and established objectives, we raise three hypotheses that govern this work. They are:

- **Hypothesis 1:** the use of pre-trained models and large language models for feature detection can advance the results obtained for Portuguese with traditional methods;
- **Hypothesis 2:** it is possible to extend aspect detection methods to improve the detection of implicit aspects;
- **Hypothesis 3:** The satisfactory identification of implicit aspects requires linguistic modeling of domain-specific knowledge.

1.4 Thesis organization

This document is structured as follows: in Chapter 2, we present the theoretical foundation on sentiment analysis and identification of explicit and implicit aspects. In Chapter 3 related works are presented and organized according to the approach used. Next, we present the papers published during this doctorate:

- Chapter 4: Learning rules for automatic identification of implicit aspects in Portuguese (MACHADO *et al.*, 2021);
- Chapter 5: Implicit opinion aspect clues in Portuguese texts: analysis and categorization (MACHADO *et al.*, 2022);
- Chapter 6: Evaluating Methods for Extraction of Aspect Terms in Opinion Texts in Portuguese – the Challenges of Implicit Aspects (MACHADO; PARDO, 2022a);
- Chapter 7: Evaluating from the Classical Frequency to the Large Language Models-based Methods for Aspect Extraction in Portuguese and their Performance on Detecting Typologically Different Implicit Aspects(MACHADO; PARDO, 2023).

Finally, in Chapter 8 we present our general conclusions about the studies carried out.

2 THEORETICAL FOUNDATION

In this chapter, we will present the main fundamentals related to sentiment analysis. We will discuss concepts, problems, examples, and methods of aspect extraction, emphasizing the implicit aspects, which are the focus of this work.

2.1 Aspect-based sentiment analysis

As we saw in the previous chapter, sentiment analysis or opinion mining refers to the computational treatment of opinions, feelings, and emotions expressed in texts. One of the main application areas of sentiment analysis is product evaluation, a task that can be quite complex. For example, in a single sentence mentioning more than one aspect of a product, each one relates to different sentiments. A word can represent different sentiments depending on the context in which it appears. For example, a product considered cheap is usually something positive, on the other hand, a product made with cheap material refers to a negative sentiment. This analysis of product characteristics and their respective sentiments is what we call aspect-based sentiment analysis, which performs deeper analysis on evaluative texts, identifying and relating aspects to their respective sentiments (THET; NA; KHOO, 2010).

To present the problem, we use the following segment from a review of a Canon camera (adapted from (LIU, 2012)) (for ease of reference, a number has been associated with each sentence):

> Posted by: John Smith Date: September 10, 2011
> "(1) I bought a Canon G12 camera six months ago. (2) I simply loved it. (3) The quality of the photos is incredible. (4) Battery life is also long. (5) However, my wife found the camera too heavy."

The first point that we can observe regarding this review is that it is formed by several opinions, positive and negative. Sentences (2), (3), and (4) refer to positive opinions about the camera, while (5) refers to a negative opinion. We can also observe the targets related to these opinions, which can be the product itself or some aspect of the product. In the example, sentence (2) considers the product in its entirety, and sentences (3), (4), and (5) are about specific aspects: "photo quality", "battery life", and " weight", respectively. Another feature of the review is its author, "John Smith", except for the opinion contained in (5), which was issued by the author's wife. Finally, we can see that the opinion was published on "September 10, 2011".

From this example, we can formally define an opinion. An opinion comprises two main components: a target g and a sentiment s about that target. Therefore, g can be an entity or aspect of an entity about which an opinion was issued, and s is a sentiment that can be positive, negative, neutral, or even a numerical scale representing the intensity of the sentiment. In the example, the opinion target of sentence (2) is “Canon G12” and the target of sentence (3) is “the quality of the photos”.

Other components that appear in the example are the issuer (h) and the time (t) at which the opinion was issued. Finally, we have the object or entity itself (e) as a component: a product, person, service, event, organization, or topic. It is associated with a tuple $e : (T, W)$, where T is a hierarchy of components (or parts) and subcomponents, and W is a set of attributes of e. Each component has its own set of subcomponents and attributes.

Example 1: A particular cell phone model is an entity. It has a set of components (battery, screen, etc.) and a set of attributes (voice quality, size, weight, etc.). The battery component also has its own set of attributes, for example, battery life and battery size.

Based on this definition, an object can be represented as a tree, hierarchy, or taxonomy. The root of the tree is the object itself. Each non-root node is a component or subcomponent of the object. Nodes can also be associated with sets of attributes or properties. An opinion can be expressed about any node or any attribute of the node. Figure 2 presents an example of a tree with the aspects of a phone.

Example 2: Following Example 1, one can express an opinion about the phone itself (the root node), for example, “I don’t like this phone”, or about one of its attributes, for example, “The photo quality on this phone is terrible”. Likewise, an opinion may also be expressed about any of the phone’s components or attributes of the component.

It is often useful to simplify this definition for two reasons. First, natural language processing is an arduous task. It is extremely challenging to study the text at a more refined level. Second, for an average user, it is too complex to use a hierarchical representation of an object and opinions about the object. Thus, we flatten the tree to omit the hierarchy and use the term aspect to represent components and attributes. In this simplification, the object itself also comes to be seen as an aspect (but a special aspect), which is the root of the original tree. An opinionated comment about the object itself is called a general opinion about the object (e.g., “I like the phone”). An opinionated comment about any specific aspect is called a specific opinion about an aspect of the object, for example, “The phone’s camera is really cool”. In this case, “camera” is an aspect of the phone.

Given these components, following (LIU, 2012), we can define opinion as a quintuple:

$(e_{\mathrm{i}},\ a_{\mathrm{ij}},\ s_{\mathrm{ijkl}},\ h_{\mathrm{k}},\ t_{\mathrm{l}})$

Where: e_{i} is the name of an entity, a_{ij} is an aspect of e_{i}, s_{ijkl} is a sentiment about

Figure 2 – Phone aspects taxonomy example

```
Phone
├─ Battery
│  ├─ Capacity
│  ├─ Life
│  └─ Removable
├─ Camera
│  ├─ Focus
│  ├─ Photo Quality
│  │  ├─ Day Shots
│  │  └─ Night Shots
│  ├─ Video Quality
│  └─ Zoom
│     ├─ Digital
│     └─ Optical
├─ Communication
├─ Connectivity
├─ Durability
├─ General
├─ Memory
├─ Operating System
├─ Processors
└─ Screen
```

the a_{ij} aspect of the entity e_i, h_k is the issuer of the opinion, and t_l is the time when this opinion was expressed by h_k. In this definition, e_i and a_{ij} together represent the target of the opinion.

Some important points related to this definition are:

1. The subscripts are purposeful, being a way of emphasizing that each information component of the quintuple must correspond to the others, that is, a sentiment s_{ijkl} must have been emitted by an emitter h_k over an aspect a_{ij} of an entity e_i at a time t_l;

2. The five components are essential. For example, if we do not have the time component, it will not be possible to analyze the period in which the opinion was issued (years ago, a cell phone being small was an advantage; today, the larger the device, the better). Authorship is also relevant, for example, for a company, it is important to know how its product is evaluated according to the user profile;

3. This definition covers most but not all possible forms of opinion. For example, it does not cover the situation presented in “The viewfinder and lens are too close”, which expresses an opinion about the distance between the two parts. It also does not cover the context of the opinion, for example, in “This car is too small for a tall person”, the “car” is not necessarily too small for everyone;

4. This definition provides a structure for transforming unstructured text into structured data, thus enabling, for example, storage in databases;

5. This setting is only valid for regular opinions. It is not valid for comparative opinions, for example.

From this quintuple, we can list the main tasks related to sentiment analysis, but for that, we need a few more definitions.

- **Opinionated Document:** This can be a product review, forum post, or blog that reviews products or services. In the most general case, the opinionated document d consists of a sequence of sentences $d = (s1, s2, ..., sm)$;

- **Entity category and entity expression:** An entity category represents a single entity, that is, a specific product or service. An entity expression is a word or text expression referring to an entity category. Each entity must have a unique name in each application. The process of grouping entity expressions into entity categories is called entity categorization. For example, a “phone” entity can be referenced by “smartphone”, “cell phone”, “iPhone”, among others;

- **Aspect category and aspect expression:** Similar to entities. An entity aspect category represents a unique aspect of the entity, that is, a characteristic or attribute of the entity. An aspect expression is a word or phrase from a text that indicates an aspect category. For example, “image”, “photo”, and “picture” are related to the same aspect “photo quality” of a camera. So, it is necessary to extract the expressions of aspects and categorize them. Aspect expressions are usually nouns and noun phrases, but they can also be verbs, verb phrases, adjectives, and adverbs;

- **Implicit and explicit aspects:** aspects mentioned literally in a text are called explicit aspects, while implicit aspects can be understood by the sentence but are not mentioned literally. With the explicit aspects being much more numerous, most research focuses on this category of aspects (SCHOUTEN *et al.*, 2015).

Example 1: Consider the sentences “The food at that restaurant tastes great” and “The food is delicious at that restaurant”. In the first sentence, the taste aspect of the food entity is explicitly mentioned; in the second, we can infer that the sentence refers to the same aspect, although it is not explicitly mentioned.

Example 2: In the sentence “This product costs a lot”, the “price” aspect is not mentioned literally, although it is implied by the word “cost”. This links the word “cost” to the implied “price” aspect. However, this correlation is not always so obvious.

Example 3: The sentence “I couldn’t sleep because of the noise” was taken from a hotel review. This example is a negative comment about the rooms or the location of the hotel, but there is no correlation of any word in the text to any aspect that could be used. The same happens in the sentences “I would recommend this restaurant” and “I will definitely come back”, in which the customer’s overall experience is the aspect, but this aspect is not associated with any word in the sentences.

The identification of implicit aspects is a much more complicated task than the identification of explicit aspects. The problems increase with the possibility that the same aspect term can be associated with several entities.

Example 4: In the sentences “The pizza is too small” and “The size of the pizza is too small”, the term “size” is related to the entity pizza. In the sentences “Restaurant is small” and “The size of the restaurant is small”, the same term “size” is associated with the entity restaurant. Likewise, the opinion word “small” that is attached to the aspect size refers to the size of the pizza in the first two sentences and the size of the restaurant in the next two sentences, leading to ambiguity. This problem increases when we deal with many aspects.

Example 5: People also tend to mention multiple aspects in a single sentence. For example, consider the sentences “Although the food is expensive, it was delicious” and

"The food was delicious at that little restaurant". In the first sentence, the price and taste aspects of the food entity are implicitly mentioned. In the second example, the flavor aspect of the food entity and the size of the restaurant entity are implicitly mentioned.

Based on these definitions (LIU, 2012) and a set of opinion documents D, we arrive at the following six main tasks of sentiment analysis.

1. **Entity extraction and categorization:** aims to identify all entity expressions in D and categorize synonymous expressions into the same entity category. Each entity expression category indicates a single entity e_{i};
2. **Aspect extraction and categorization:** aims to identify all aspect expressions of entities and categorize these expressions into aspect categories. Each aspect category of the entity e_{i} represents an aspect a_{ij};
3. **Extraction and categorization of opinion makers:** aims to identify opinion makers and categorize them;
4. **Time extraction and standardization:** aims to identify when opinions were issued and standardize this information;
5. **Aspect sentiment rating:** is intended to determine whether the opinion about an aspect a_{ij} is positive, negative, or neutral, or to assign a numerical sentiment rating to the aspect;
6. **Opinion quintuple generation:** aims to produce opinion quintuplets (e_{i}, a_{ij}, s_{ijkl}, h_{k}, t_{l}) expressed in a document d based on the results of the tasks above.

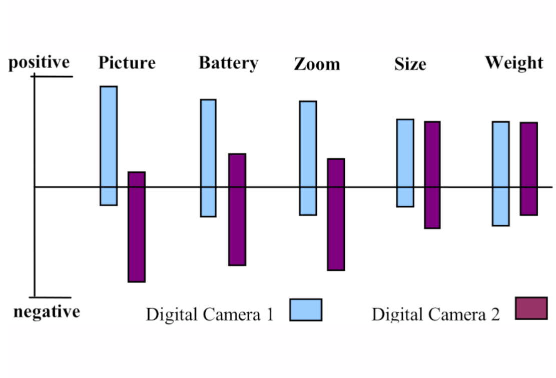
Sentiment analysis performed based on these tasks is often referred to as aspect-based sentiment analysis (HU; LIU, 2004). To illustrate these tasks, here is an example adapted from (LIU, 2012):

> Posted by: bigJohn Date: September 15, 2011
> "(1) I bought a Samsung camera, and my friends bought a Canon camera yesterday. (2) Last week, we both used the cameras a lot. (3) My Samy's photos are not that good, and the battery life is also short. (4) My friend was very happy with his camera and loved the image quality. (5) I want a camera that can take good photos. (6) I will return it tomorrow."

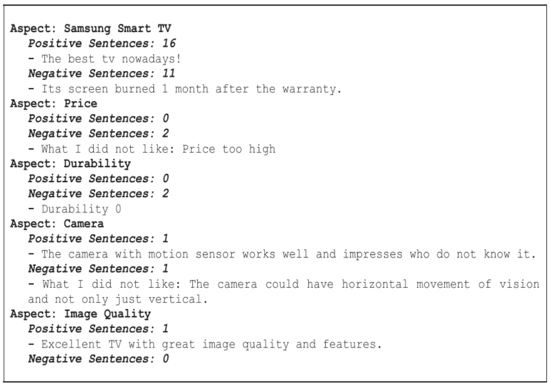
Task 1 should extract the entity expressions "Samsung", "Samy", and "Canon". After that, it should group "Samsung" and "Samy" as they represent the same entity. Task 2 should extract the aspect expressions "image", "photo", and "battery life", and then

Figure 3 – Visualization of the results of two sentiment analysis systems

(a) (LIU; HU; CHENG, 2005)



(b) (CONDORI; PARDO, 2017)



```
Aspect: Samsung Smart TV
   Positive Sentences: 16
   - The best tv nowadays!
   Negative Sentences: 11
   - Its screen burned 1 month after the warranty.
Aspect: Price
   Positive Sentences: 0
   Negative Sentences: 2
   - What I did not like: Price too high
Aspect: Durability
   Positive Sentences: 0
   Negative Sentences: 2
   - Durability 0
Aspect: Camera
   Positive Sentences: 1
   - The camera with motion sensor works well and impresses who do not know it.
   Negative Sentences: 1
   - What I did not like: The camera could have horizontal movement of vision
   and not only just vertical.
Aspect: Image Quality
   Positive Sentences: 1
   - Excellent TV with great image quality and features.
   Negative Sentences: 0
```

group “image” and “photo” into the same aspect category. Task 3 must determine that the issuer of opinions in sentence (3) is bigJohn and the holder of opinions in sentence (4) is a friend of bigJohn. Task 4 must find the date the text was posted (September 15, 2011). Task 5 should discover that sentence (3) carries a negative opinion about the Samsung camera’s image quality and a negative opinion about the battery life. Sentence (4) brings a positive opinion of the Canon camera in general and for the image quality. Sentence (5) apparently expresses an opinion that could be positive if there were an associated entity, but it does not. To generate the opinion quintuplets for sentence (4), we need to know that the expression “his camera” refers to the Canon camera. Task 6 should generate the following opinion quintuplets:

- (Samsung, quality_of_image, negative, bigJohn, September 15, 2011)
- (Samsung, battery_life, negative, bigJohn, September 15, 2011)
- (Canon, GENERAL, positive, friend_of_bigJohn, September 15, 2011)
- (Canon, quality_of_image, positive, friend_of_bigJohn, September 15, 2011)

Finally, after processing opinion texts, the resulting information must be converted into a format that can be easily understood. For this purpose, charts and/or summarization techniques can be used. In Liu, Hu and Cheng (2005), the authors present a system called Opinion Observer that generates visualizations like those in Figure 3a that allow the user to understand the weaknesses and strengths of a product. In the example, it can be seen that camera 1 has better image quality, battery, and zoom than camera 2 and the size and weight aspects are close.

Condori and Pardo (2017) developed Opizer-E which summarizes the results in text form. Figure 3b shows the sentiments related to the aspects accounted for by the number of positive and negative sentences, in addition to presenting examples of the sentences found. Therefore, we can conclude that as important as the extraction of data from the quintuple is how this data is presented.

According to Cambria (2016), existing approaches to sentiment analysis can be classified into knowledge-based techniques and statistical methods.

Knowledge-based techniques aim to identify words or expressions indicative of sentiment in the input text that are contained in knowledge bases built from linguistic corpora, such as lexicons and WordNets. In addition to knowledge bases, these techniques also explore language rules to identify the context of words. Thus, these techniques depend heavily on the quality of linguistic resources to achieve reliable results.

Statistical methods use machine learning algorithms, which are trained from linguistic features extracted from texts. They can identify sentiment-related keywords as well as the frequencies of co-occurrences with other keywords. However, in general, they are semantically weak and need a lot of data for training.

Deep learning refers to a branch of machine learning that uses multi-layer artificial neural networks, also known as deep neural networks. We can understand this type of network as a regular neural network with many hidden layers organized in a hierarchy with the objective of extracting attributes from the input data. In this hierarchy, the first layers extract more abstract characteristics, while the last layers extract more specific characteristics. The reason for using a greater number of layers is that as the layers become deeper, richer information can be explored, which helps in classification (LECUN; BENGIO; HINTON, 2015).

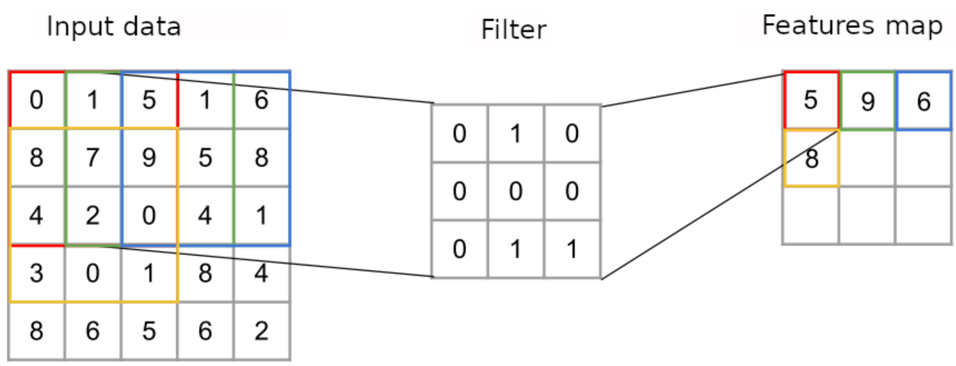
Deep learning techniques are capable of automatically extracting useful features in natural language processing tasks, in contrast to the manual feature engineering required in traditional machine learning. These models have become popular in sentiment analysis (ZHANG; WANG; LIU, 2018) and natural language processing tasks (YOUNG *et al.*, 2018). The most current approaches used to solve sentiment analysis problems involve the use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These networks are capable of extracting meaningful information from the input data, thus improving text classification.

- **Convolutional Neural Networks (CNN):** networks that have one or more layers that perform convolutional operations. This type of network is widely used in image and video processing, as it was designed to receive multidimensional matrices as input (two-dimensional in the case of images and three-dimensional for videos). In the first layers, the raw data is processed by convolutional layers, see Figure 4, which apply a series of filters to the input data, generating attribute maps. Each filter is applied to all input data, multiplying its weights by the input values. The result of this process is passed to a non-linear activation function. The feature maps generated by the filters highlight distinct parts of the input.

- **Recurrent Neural Networks (RNN):** deep neural network that deals with sequences of inputs, generally used to deal with speech and language problems. This type of network processes one element at a time from a sequence of data, maintaining a history of previous entries in its hidden layers. Its architecture can vary from fully interconnected neural networks, where the input of each node is the output of all the others and the network input can be any node, to partially connected networks, where the nodes have a structure like the previous one but some of them have more than one input and output (JAIN; MEDSKER, 1999). As information is preserved, networks can use information from previous nodes to predict the next input while respecting the context of previous inputs. However, when dealing with long sequences, recurrent neural networks have problems predicting labels that require information from the distant past.

- **Long Short-Term Memory (LSTM):** to fix the problem with long sequences of recurrent neural networks, (SCHMIDHUBER; HOCHREITER, 1997) proposed the Long Short-Term Memory (LSTM) network, a type of RNN that manages internal information with a series of memory mechanisms. The main difference between an RNN and an LSTM is that the latter has a cell state that passes through the entire network. In this cell, LSTM adds and deletes new information according to its importance. This mechanism ensures that information necessary for classification is available throughout the network, allowing the definition of a context for future decisions.

Figure 4 – Convolution Example.



To work with texts, many deep learning models represent words in documents as feature vectors, which can capture semantic or syntactic dependencies. One of the ways to carry out this type of representation is distributional models. These models are inspired by the idea that words that occur in the same context tend to have similar meanings. This idea is called the distributional hypothesis and is addressed in Karlgren and Sahlgren (2001) where the philosophical and linguistic motivations for using this type of technique are presented. For example, if we think of the expression "travel to", the words that will appear next will most likely be the names of places, cities, and countries, among others.

Based on this distributional hypothesis, we can think of a representation of words in the form of vectors created from their co-occurrences. In this model, each word is represented by a vector, and each position of this vector indicates the frequency of occurrence of this word with another word in each corpus. In this vector space formed by the vectors of all the words in this corpus, we will have semantically similar words with close vector representations, following the idea of the distributional hypothesis.

The technique called word embedding is used to perform language modeling and resource learning, which transforms words contained in a vocabulary into vectors of continuous real numbers. Some recent models for representing words in vectors are:

- **Word2Vec:** introduced by Google in 2013, Word2Vec uses two predictive models based on neural networks. The first model, continuous bag-of-words (MIKOLOV *et al.*, 2013a), predicts the target word from the context words, while the second model, skip-gram, does the opposite, predicting the context words given the target word.
- **Global vectors for word representation (GloVe):** presented by Stanford University in 2014, uses an unsupervised algorithm to obtain vector representations of words. Training is conducted considering the joint frequency distribution of terms in a corpus (PENNINGTON; SOCHER; MANNING, 2014).
- **Fasttext:** presented by Facebook in 2016, presents similar characteristics to the Word2Vec model but considers the construction of subwords in the formation of vocabulary (BOJANOWSKI *et al.*, 2017).
- **Embeddings from language models (ELMo):** presented by the Allen Institute for Artificial Intelligence in 2018, is different from traditional models as each token is assigned a function related to the entire input sentence. The vectors are derived from a bidirectional LSTM that is trained from a large corpus with a coupled language model (PETERS *et al.*, 2018).
- **Bidirectional Encoder Representations from Transformers (BERT):** presented by Google in 2018. The BERT model creates pre-trained models with deep bidirectional representations of unlabeled text, jointly conditioning left and right contexts across all layers. As a result, the obtained model can be adjusted by adding a new output layer, with the aim of creating models for a wide range of tasks (DEVLIN *et al.*, 2019).

As presented in Chapter 1, Pavlopoulos and Androutsopoulos (2014) divides aspect-based sentiment analysis into three subtasks: aspect term extraction, aspect term sentiment estimation, and aspect aggregation. The objectives of this work are related to the subtask of extracting aspect terms, with a special focus on implicit aspects and the challenges they bring to processing. According to Tubishat, Idris and Abushariah (2018), methods for extracting aspects can be divided into three categories: supervised, semi-supervised, and unsupervised.

- **Unsupervised methods:** do not require labeled data for their operation, which is an advantage, given the difficulty of manual annotation required for the task of identifying aspects. Some examples of approaches in this category are dependency analysis, association rule mining, rule-based extraction, and clustering.
- **Supervised methods:** require labeled data for their operation. Some examples of approaches include association rule mining, fuzzy, rule-based, and traditional machine learning algorithms.
- **Semi-supervised methods:** use both labeled and unlabeled data. Some more commonly used methods can be based on mining association rules, clustering, and modeling topics.

Identifying implicit aspects is a challenging task in itself: people can express opinions in vastly diverse ways, with grammatically incorrect sentences, abbreviations, and language habits that differ from person to person. Below, we present some questions specifically related to the extraction of implicit aspects (GANGANWAR; RAJALAKSHMI, 2019):

- **Multiple implicit aspects in a sentence**: when a single implicit aspect is present, we can easily associate the sentiment of an entire sentence with the sentiment of that aspect. When multiple implicit aspects are present, finding the sentiment of the aspect becomes more challenging. This is because aspects can have different polarities than the overall polarity of the sentence. For example, in the statement "The photos taken may have been blurry due to the lack of image stabilizer, but overall, it is a great option for a low budget", two different aspects (camera quality and price) are implicitly mentioned with different sentiments related to each one.
- **Ambiguous opinion words:** the same word can be used to describe distinct aspects depending on its context. For example, in "The pizza is very small", the word "small" is related to the "size" aspect of the pizza entity. On the other hand, in "The pizzeria is very small", the same word "small" is related to the "size" aspect, but this time of the pizzeria entity.
- **Nouns or noun phrases can be implicit aspects:** most articles consider adjectives and verbs as opinion words, but in some cases, nouns or noun phrases can act as

implicit aspect clues. For example, in the sentence "Overall, it's a good deal", the noun phrase "good deal" implies a positive sentiment and not an aspect, as nouns and noun phrases usually represent.

- **Double implicit problem:** opinion words act as clues in detecting implicit aspects. There are situations where both appearance and opinion are implicit. In these cases, finding the aspect/sentiment pair becomes even more challenging. For example, in the sentence "I received calls even on the banks of the São Francisco", the expression "I received calls" is indicative of the signal aspect of the telephone, and the fact that I received calls close to the banks of the river indicates that the telephone has good reception, leading to a positive feeling about the signal.

3 RELATED WORK

In this chapter, we will discuss some works related to aspect extraction. We selected those that present the methods and corpus that we will use, in addition to other important works that represent the state-of-the-art in the area.

3.1 Hybrid Methods

3.1.1 Hu and Liu (2004)

In this work, the authors presented a new corpus that is still widely used in tasks related to opinion mining. The authors also implemented unsupervised methods for aspect identification and polarity analysis.

Corpus

To carry out the experiments, the authors created a corpus with texts in English using customer reviews of five electronic products: digital cameras (Canon and Nikon), DVD player, mp3 player, and cell phone. Reviews were collected from *Amazon.com* and *cnet.com* websites. For each product, the authors collected the first 100 reviews. These texts were manually analyzed, with their aspects identified and their respective opinions marked as positive or negative.

Method description

After collecting the data, performing the PoS tagging, and identifying the noun phrases, the authors extracted the aspects of the analysis by implementing the following steps:

1. **Identification of frequent aspects:** selection of the most frequent aspects identified in the text. Nouns and noun phrases that appeared in more than 1% of sentences were selected;

2. **Identification of compound aspects:** bigrams and trigrams of nouns were extracted from the sentences in the same order in which they appeared. These aspects were included in the list of candidate aspects;

3. **Calculation of p-support:** the p-support of a candidate aspect t refers to the number of sentences containing t. This calculation does not consider sentences that contain a candidate t', where the candidate t is a subset of t';

4. **Calculation of the compactness index:** in a sentence that contains compound candidate terms, the distance between each pair of candidate aspects is checked. If this distance is greater than three words, this index is incremented;

5. **Compactness pruning:** terms that have a compactness index greater than one, that is, that have appeared more than once in a non-compact form, are removed;

6. **P-support pruning:** candidates that have a p-support index less than three, which is an indication that another form that contains the candidate is the most frequent, are also removed;

7. **Opinion word extraction:** adjectives found close to a candidate aspect are added to a list of opinion words;

8. **Detection of infrequent aspects:** sentences, where no candidate terms were found, are re-analyzed for opinion words. If any are found, the noun closest to that opinion word is added to the list of candidate aspects.

Results

The authors evaluated the results of each step of the algorithm, presented in Table 2. Columns 3 and 4 present the precision and recall of frequent aspects identification. The results indicate that the most frequent aspects contain many errors, leading to low precision. Columns 5 and 6 show the results after compactness pruning, which provides a significant improvement in accuracy, whereas recall remained stable. Columns 7 and 8 bring the results after p-support pruning, further improving accuracy, with recall continuing at the same previous level. The last two columns, 9 and 10, show the results of the detection of infrequent aspects. Recall is greatly improved, but there is a drop in accuracy. The authors conclude that this ends up not being a big problem due to the low frequency of these terms, which does not affect most users.

Table 2 – Hu and Liu (2004) results

Domain	Aspects	Frequent		Compactness		P-support		Infrequent	
		Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Camera 1	79	0.552	0.671	0,634	0.658	0.825	0.658	0.747	0.822
Camera 2	96	0.594	0.594	0,679	0.594	0.781	0.594	0.710	0.792
Cell phone	67	0.563	0.731	0,676	0.716	0.828	0.716	0.718	0.761
Mp3 player	57	0.573	0.652	0,683	0.652	0.754	0.652	0.692	0.818
DVD	49	0.531	0.754	0,634	0.754	0.765	0.754	0.743	0.797

3.1.2 Popescu and Etzioni (2007)

The authors divided the opinion mining problem into the following subtasks: identifying product aspects, identifying opinions about product aspects, identifying the polarity of opinions, and classifying opinions. In this work, the authors present OPINE, an unsupervised information extraction system that incorporates a solution for each of these subtasks.

Corpus

In the experiments, the authors used a set of 1,621 reviews in English on seven classes of products, namely:

- datasets on five product classes from Hu and Liu (2004);
- two sets of reviews taken from *tripadvisor.com* (hotels) and *amazon.com* (scanners).

Method description

The steps of the system developed by the authors are described in Figure 5. Although the system also identifies implicit aspects and performs the grouping of opinions, the authors omitted the description of these algorithms in this work.

Figure 5 – OPINE system overview

Input: product class C, reviews R.
Output: set of [feature, ranked opinion list] tuples
R' ← parseReviews(R);
E ← findExplicitFeatures(R', C);
O ← findOpinions(R', E);
CO ← clusterOpinions(O);
I ← findImplicitFeatures(CO, E);
RO ← rankOpinions(CO);
$\{(f, o_i, ...o_j)...\}$←outputTuples(RO, I ∪ E);

OPINE extracts the explicit aspects of the given product class from opinion texts. The system recursively identifies parts and properties of a given product class, repeating the process until no candidates are found. Initially, the system extracts noun phrases from reviews and retains those with a frequency greater than a pre-defined limit. The system then evaluates each noun phrase by calculating scores (pointwise mutual information) between the sentence and the meronymy discriminators associated with the product class. To distinguish parts of properties, we used the WordNet (MILLER, 1995; MILLER, 1998) hierarchy of hypernyms (which enumerates different classes of properties) and morphological clues (for example, suffixes "-iness", " -ity").

Results

Two annotators labeled a set of 450 extractions performed by OPINE as correct or incorrect, and those that the annotators agreed on were used to evaluate the system. The agreement between the annotators was 86%, the precision calculated for the system was 89% and the recall was 73%.

Table 3 presents the results compared with the work of Hu & Liu (HU; LIU, 2004). OPINE precision was 22% higher, while recall was down 3%. The authors point out two significant differences between OPINE and Hu & Liu's system: the OPINE aspect evaluator uses the pointwise mutual information assessment to evaluate each aspect candidate and OPINE incorporates statistics from Web PMI as well as reviewing the data in its assessments.

Table 3 – Popescu and Etzioni (2007) results

	Hu & Liu		OPINE	
Domain	Precision	Recall	Precision	Recall
Camera 1	0.75	0.82	0.94	0.80
Camera 2	0.71	0.79	0.93	0.73
Cell phone	0.72	0.76	0.95	0.73
Mp3 player	0.69	0.82	0.94	0.79
DVD	0.74	0.80	0.95	0.78

3.1.3 Pavlopoulos and Androutsopoulos (2014)

In this paper, the authors implemented methods for extracting aspect terms. They contributed in three diverse ways: they presented a new corpus, developed new specific metrics for this type of task, and implemented improvements to the method of Hu and Liu (HU; LIU, 2004).

Corpus

The authors constructed and made available three new English-language datasets from three domains (restaurants, laptops, and hotels), with manual annotations of all aspect term occurrences.

The restaurant dataset contains 3,710 sentences from (GANU; ELHADAD; MARIAN, 2009) reviews. The annotators marked the aspect terms of each sentence, but the implicit aspects were not considered. Of these sentences, 1,248 contain exactly one aspect term, 872 have more than one, and 1,590 have no aspect term. 593 aspect terms formed by multiple words and 452 aspect terms formed by a single word were found.

The hotel dataset consisted of 3,600 English sentences from online customer reviews of 30 hotels. Of these sentences, 1,326 contain exactly one aspect term, 652 more than one, and 1,622 none. There are 199 multi-word aspect terms and 262 single-word aspect terms.

The laptop dataset contains 3,085 sentences obtained from 394 online customer reviews. Of these sentences, 909 contain exactly one aspect term, 416 have more than one, and 1,760 have none. We found 350 distinct terms of multiple words and 289 terms of aspects formed by a single word.

These datasets were used at the 2014 *SemEval* conference.

Method description

The authors implemented and evaluated four methods for extracting aspect terms:

- **Freq Baseline:** returns the most frequent nouns and noun phrases from reviews in each dataset, ordered by decreasing frequency with respect to sentences. The most frequent candidates are selected, which have appeared in at least 1% of the sentences (HU; LIU, 2004);
- **Hu & Liu:** statistical and rule-based hybrid method detailed in Section 3.1.1 (HU; LIU, 2004);
- **Hu & Liu + *Word2Vec*:** the authors added a pruning step using the distributional model *Word2Vec* trained from *Wikipedia*. This pruning is carried out right after the detection of the most frequent aspects. To do this, the ten most frequent terms were selected, and from their vectors, a centroid vector called the domain centroid was calculated. A similar process was carried out with the twenty most frequent terms in the Brown Corpus (news category); this vector was named the common language centroid. Candidates are compared with the two centroids using the cosine similarity measure. Candidates closest to the common language centroid are discarded;
- ***Freq Baseline* + *Word2Vec*:** the same pruning mechanism applied in Hu & Liu's method was used in conjunction with *Freq Baseline*.

The authors also proposed new metrics to analyze the results of aspect term extraction: weighted precision, weighted recall, and average weighted precision. As it is not within the scope of this work, we will omit the discussion of these measures.

Results

Table 4 shows the results calculated with the average weighted precision metric. All four methods achieved the best results on the restaurant dataset. At the other extreme, the laptop dataset appears to have been the most difficult. This is probably because it contains frequent nouns and noun phrases that are not aspect terms. This dataset also has more aspect terms made up of multiple words.

Table 4 – Pavlopoulos and Androutsopoulos (2014) results

Domain	Freq	Freq+Word2Vev	Hu&Liu	Hu&Liu+Word2Vec
Restaurants	43.40	45.17	52.23	66.80
Hotels	30.11	30.54	49.73	53.37
Laptops	9.09	7.18	34.34	38.93

3.1.4 Panchendrarajan *et al.* (2016)

The authors presented a method for detecting aspects that overcame limitations observed in previous works: the limitation of the number of implicit aspects per sentence and the ambiguity when an opinion word is used to describe distinct aspects.

Corpus

To form the corpus, the authors collected a set of 1,000 reviews in English about restaurants from the website *Yelp* and manually marked the implicit and explicit aspects. Many distinct aspects were found in the domain of restaurants, but not all reviews presented implicit aspects. About 15.6% of the sentences contain one or more implicit aspects. However, for some categories of aspects, the implicit ones had a greater importance. For example, more than 92% of employee-related aspects were mentioned implicitly.

Method description

To identify the aspects, the authors created two models. The first model uses a maximum entropy classifier to identify explicit aspects. The second model identifies opinion words in a sentence and predicts the implicit aspect related to that opinion word using word co-occurrence.

Given a review, explicit aspects and entities are identified using the first trained model and added to a list of candidate aspects. The sentences are analyzed, seeking to extract opinion words. The second model is then used to identify the aspects related to these opinion words and add them to the list of candidate aspects. Selecting aspects from this list is a two-step process:

- **Step 1:** an implicit aspect from the list of candidate aspects is chosen as a potential aspect candidate, using co-occurrence between the opinion word and other words in the sentence. If there is only one candidate aspect, it will be chosen as a potential candidate. Otherwise, for each aspect candidate, a score is calculated.
- **Step 2:** from a potential candidate, its opinion target is extracted to verify that the prediction is correct. These targets are extracted using the double propagation approach proposed by Qiu *et al.* (2011), which propagates information between opinion words and targets using grammar rules based on the dependency relationships between the words. If the opinion target identified is related to the potential candidate, it will be chosen as the winning candidate, otherwise, it will be discarded.

Results

The evaluations were performed using cross-validation with 10-folds on a dataset composed of 1,000 reviews. In the experiments, two modifications were made to the presented approach: not executing step 2 and disregarding the occurrence of words in the calculation of the candidate aspects' scores.

The results are presented in Table 5. Methods 1, 2, and 3 assume that the first model can identify explicit features and entities with an accuracy of 100%. Method 4 shows the results of applying the first model.

Table 5 – Panchendrarajan *et al.* (2016) results

Method	Precision	Recall	F-measure
1. Initial Solution	0.947	0.758	0.842
2. Modification 1	0.916	0.752	0.826
3. Modification 2	0.931	0.754	0.834
4. Initial Solution with first model	0.886	0.694	0.779

3.2 Rule-based methods

3.2.1 Poria *et al.* (2014)

The authors consider that the extraction of explicit aspects has been widely researched and that there are several approaches to this task in contrast to the limited number of works related to the extraction of implicit aspects. They consider the task difficult, but important given the presence of the phenomenon of implicit aspects being present in most opinion documents. In this work, the authors propose a novel approach for detecting explicit aspects and clues of implicit aspects in opinion documents and divide this task into two steps: identifying clues of implicit aspects and mapping them to the corresponding aspects.

Corpus

To evaluate the explicit aspects extraction algorithm, the authors used the English language corpus provided by Hu and Liu (2004) and the dataset from the 2014 congress *SemEval*, presented in Table 6. For the implicit aspects extraction algorithm, they used the corpus developed by Cruz, Gelbukh and Sidorov (2014), who manually labeled each implicit aspect clue and its corresponding aspects found in the Hu and Liu (2004) corpus.

Table 6 – SemEval 2014 corpus

	Sentences containing n aspects			
Domain	$n = 0$	$n \geq 1$	$n \geq 2$	total($n \geq 0$)
Restaurants	1,732	2,212	881	3,944
Laptops	1,883	2,065	456	3,948

Method description

For pre-processing, the authors carried out two tasks: construction of sentence dependency trees and lemmatization of the elements of the obtained trees. As auxiliary tools, they used the lexicon of implicit aspects developed in Cruz, Gelbukh and Sidorov (2014) and the opinion lexicon *SenticNet 3* (CAMBRIA; OLSHER; RAJAGOPAL, 2014). A dependency relation is a binary relation characterized by the following:

- **Relationship type:** specifies the nature of the (syntactic) connection between the two elements;
- **Head of the relationship:** pivot element of the relationship;
- **Dependent:** element that depends on the pivot and that usually inherits some of its characteristics (for example, number and gender in case of concordance).

Manually created dependency rules were employed in the parse trees to extract the aspects. In processing these rules, a token is considered active if it acts as the head of a relationship, although there are exceptions. Once the active token has been identified as the trigger for a rule, there are several ways to calculate its contribution, depending on its dependency relationship with other tokens. The rules process each token in the dependency tree found. The proposed aspect analyzer was based on two categories of general rules:

- **Rules for sentences that contain a subject and noun relationship in their dependency tree:** if an active token h is in a syntactic subject relationship with a token t, then:
  1. If t has any adverbial or adjective modifier and this modifier exists in *SenticNet*, then t is extracted as aspect;
  2. If the sentence does not have an auxiliary verb then:
     - If a verb t is modified by an adjective or adverb or is in an adverbial clause-modifying relationship with another token, then h and t are extracted as aspects. In (1), "battery" is in a subject relation with "lasts", and "lasts" is modified by the adjective "little", so both "lasts" and "battery" are extracted as aspects.
       (1) The battery lasts little.
     - If t has an indirect object relationship with a token n and that is a noun not found on *SenticNet*, then n is extracted as aspect. In (2), "like" is in indirect object relation with "lens", so that "lens" is extracted as aspect.
       (2) I like the lens of this camera.
     - If t has an indirect object relationship with a token n and this is a noun that has been found on *SenticNet*, then n is extracted as aspect. If another token n_1, which is a noun, has a relationship with n, then n_1 is extracted as an aspect. In (3), "like" is in indirect object relation to "beauty", which

is connected to “screen” via a prepositional relation. Thus, “screen” and “beauty” are extracted as aspects.

(3) I like the beauty of the screen.

- If t is in an open clause complement relation with a token t_1, then the aspect t-t_1 is extracted if t-t_1 exists in the opinion lexicon. If t_1 is connected to a noun t_2, then t_2 is extracted as an aspect. In (4), “like” and “comment” are in clause-complement relation and “comment” is connected to “camera” using a prepositional relation. Here, the morphosyntactic classification of “camera” is a noun, and therefore “camera” is extracted as an aspect.

  (4) I would like to comment on the camera of this phone.

3. If t is in a copular relationship with a copular verb and this verb exists in the lexicon of implicit aspects, then t is extracted as an aspect. In (5), “expensive” is extracted as an aspect.

   (5) The car is expensive.

4. If t is in a copular relationship with a copular verb and the token h is a noun, then h is extracted as an aspect. In (6), “camera” is extracted as an aspect.

   (6) The camera is good.

5. If t is in a copular relationship with a copular verb and this verb is connected to another verb t_1 with any dependency relationship and both verbs exist in the implicit aspects lexicon, then both t_1 and t are extracted as aspects. In (7), “lightweight” is in copula relation with “is” and “lightweight” is connected to “carry” through an open clause complement relation. Both “lightweight” and “carry” are extracted as aspects.

   (7) The phone is very lightweight to carry.

- **Rules for sentences that have no subject and noun relationship in their dependency tree:**

  1. If an adjective or adverb h is in an infinitive complement or open clause relationship with a token t and h exists in the lexicon of implied aspects, then h is extracted as an aspect. In (8), “big” is extracted as an aspect.

     (8) Too big to hold.

  2. If h is connected to a noun t using a prepositional relation, then both h and t are extracted as aspects. In (9) “smoothness” is extracted.

     (9) Love the smoothness of the mp3 player.

  3. If h is in a direct object relationship with a t, t is extracted as an aspect. In (10), “mention” is in a direct object relation with “price”, therefore “price” is extracted as an aspect.

     (10) Not to mention the price of the phone.

- **Additional rules:**

  - For each aspect term detected by the previous rules, if an aspect term h is in coordination or joint relationship with another t, then t is also extracted as an aspect. In (11), “awesome” is first extracted as an aspect. Since “awesome” is related to “easy”, then “use” is also extracted as an aspect.

    (11) The camera is awesome and easy to use.

  - If t is extracted as an aspect and t has a compound noun modifier h, then the aspect h-t is extracted and t is removed from the list of aspects. In (12), “chicken” and “casserole” are in a noun compound modifier relationship, so “chicken casserole” is extracted as an aspect.

    (12) We ordered the chicken casserole, but what we got were a few small pieces of chicken, all dark meat and on the bone.

Results

The authors presented an unsupervised method for extracting explicit and implicit aspects using lexicons and sentence dependency structure. The results are shown in Table 7. The proposed method outperformed Hu and Liu (2004), Popescu and Etzioni (2007), and Qiu *et al.* (2011) methods.

Table 7 – Poria *et al.* (2014) results

Domain	Precision	Recall
Camera 1	0.9015	0.9225
Camera 2	0.8215	0.8615
Cell phone	0.9325	0.9332
Mp3 player	0.9225	0.9415
DVD	0.8925	0.9125
Laptop	0.8215	0.8432
Restaurant	0.8521	0.8815

3.2.2 Costa and Pardo (2020)

In this paper, the authors analyzed lexicon-based methods for extracting aspects of opinions written in Portuguese. In the first experiment, the authors used aspect ontologies to identify explicit and implicit aspects. As defined by Chandrasekaran, Josephson and Benjamins (1999), ontologies are conceptual organizations about types of objects and their properties and relationships in a specific knowledge domain. In a second experiment, the authors extended a frequency-based method of feature detection using the distributional models Word2Vec (MIKOLOV *et al.*, 2013a) to enrich the process.

Corpus

To evaluate the experiments, the authors manually annotated the aspects and polarities of a set of 105 reviews of electronic products (digital cameras and cell phones), with a total of 1,980 words. These reviews were extracted from the Portuguese corpus of Hartmann *et al.* (2014).

Method description

The first method implemented, based on ontologies, is quite simple, if a term of the revision under analysis is found in the lexical resource in use, this term is extracted as an aspect. These lexical resources are usually considered expensive, and the results of the method are linked to the quality of these resources.

The second method implemented used a lexicon created with nouns extracted from the corpus of revisions by Hartmann *et al.* (2014). For the formation of this lexicon, nouns that had a relative frequency were extracted, that is, in relation to all recovered nouns, only the 3% most frequent were selected.

For the experiments using the Word2Vec model, the authors trained a word embedding model with the corpus of Hartmann *et al.* (2014) using the skip-gram model from Word2Vec (MIKOLOV *et al.*, 2013a) with 600 dimensions. To enrich the lexical resources of the previous methods, the authors added the three words with vectors closest to each existing word in the ontologies and the lexicon of nouns. The distance between the vectors was calculated using the cosine measure for lexical similarity (JURAFSKY; MARTIN, 2000).

Results

To evaluate the results, the authors used two different tests, one pessimistic and one optimistic. In the pessimistic evaluation, an aspect is only considered correctly identified if the extracted aspect was the same as the manually annotated aspect. In the optimistic evaluation, if part of the extracted aspect is part of the annotated aspect, the system considers it correct.

The results of the experiments are presented in Table 8. Overall, the ontology-based method performed better. This result can be explained by the quality of the ontologies used, produced by Vargas and Pardo (2018).

3.3 Machine learning methods

3.3.1 Balage Filho and Pardo (2014)

In this work, the authors implemented a supervised method for aspect extraction using Conditional Random Fields (LAFFERTY; MCCALLUM; PEREIRA, 2001). The

Table 8 – Costa and Pardo (2020) results

Measures	Methods			
	Ontologies	Ontologies + *Word2Vec*	Nouns	Nouns + *Word2Vec*
Optimistic accuracy	**0.530**	0.346	0.150	0.115
Pessimistic accuracy	**0.522**	0.343	0.150	0.113
Optimistic recall	0.448	0.433	0.401	**0.458**
Pessimistic recall	**0.414**	0.399	0.360	0.405
Optimistic F-meausre	**0.485**	0.384	0.218	0.183
Pessimistic F-measure	**0.461**	0.368	0.211	0.176

authors added semantic labels to the set of attributes selected for model training and tested their efficiency.

Corpus

To validate the experiments, the authors used the corpus in English for the analysis of restaurants and laptops from the SemEval-2014 (PONTIKI *et al.*, 2014).

Method description

The authors used the sequential labeling algorithm Conditional Random Fields. To model the system, the authors built a set with six attributes:

1. word;
2. part of speech (POS);
3. chunk;
4. named entity category (NE);
5. semantic function label (SRL);
6. frame more generic on FrameNet.

According to the authors, the use of the first four attributes is consistent with the best approaches in aspect-based sentiment analysis. The last two attributes were chosen to be validated with the work. To extract the first five attributes, the Senna (COLLOBERT, 2011) system was used, which uses neural networks and deep learning to make predictions for natural language processing. To extract the frames the tool ARK SEMAFOR (DAS *et al.*, 2010) was used, which uses probabilistic analysis to find the resources of FrameNet (BAKER; FILLMORE; LOWE, 1998). As the chosen set of attributes is all related to the words of the texts, the authors selected the highest-level structure of the identified frame.

Results

The Table 9 presents the results of the performed experiments. The results are broken down by the sets of attributes used, e.g., "+ Frame" indicates all the resources broken down above (Word, POS, Chunk, NE, SRL) plus the frame attribute. In general, the SRL and Frame attributes did not improve the results. There was an increase in accuracy and a decrease in recall. Nevertheless, if the user is only interested in accuracy, attributes can be useful. This may be the case in scenarios where a lot of information is available, such as on the web, and we want to be certain about the information retrieved.

Table 9 – Balage Filho and Pardo (2014) results

Atributos	Restaurantes			Laptops		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Word + POS	83.76	68.69	75.48	80.87	39.44	53.03
+ Chunk	83.38	68.16	75.01	78.83	39.29	52.44
+ NE	83.45	68.07	74.98	79.93	39.60	52.96
+ SRL	82.79	67.46	74.34	78.22	38.99	52.04
+ Frame	87.72	34.03	49.04	83.62	14.83	25.19

3.3.2 Cruz, Gelbukh and Sidorov (2014)

In this paper, the authors consider that although there are many works related to the extraction of aspects in opinion texts, few deal with the issue of implicit aspects. They define an implicit aspect as an opinion target that is not explicitly specified in the text, and name words related to the aspect as indicators of implicit aspects. The authors also mention the lack of a corpus with identified implicit aspects. To solve this gap, they manually marked the indicators of implicit aspects of an existing corpus.

Corpus

The authors used the English language corpus developed by Hu and Liu (2004) to manually mark the implicit aspect indicators relating them to their respective aspects. Table 10 presents the properties of the corpus formed by 314 reviews of electronic products sold on the Amazon website: DVD player, Canon camera, Nikon camera, MP3 player, and Nokia cell phone.

Table 10 – Corpus Cruz, Gelbukh and Sidorov (2014) properties

	DVD	Canon	MP3	Nikon	Cell Phone
Reviews	99	45	95	34	41
Words per Review	572.3	1,236.4	1,575.5	924	1,085.3
Sentences per Review	7.47	13.26	18.90	3.64	13.31

Other corpus statistics are presented in Table 11. The first column of the "sentence level" section shows how many sentences each product has. The second column shows the number of sentences from documents that have at least one implicit aspect indicator (line labeled "IAI #") and, in the last column, the percentage of sentences that have at least one indicator ("IAI%"). The "token level" and "type level" sections present the same characteristics for these levels of granularity.

Table 11 – Corpus Cruz, Gelbukh and Sidorov (2014) statistics

	DVD	Canon	MP3	Nikon	Cell Phone
Sentence Level					
Sentences	740	597	1,796	346	546
IAI#	147	63	155	36	44
IAI%	19.86%	10.55%	9.03%	10.40%	8.05%
Token Level					
Tokens	56,661	55,638	149,676	31,416	44,497
IAI#	164	79	214	50	66
IAI%	0.289%	0.141%	0.142%	0.159%	0.148%
Type Level					
Types	1,767	1,881	3,143	1,285	1,619
IAI#	72	63	136	40	42
IAI%	4.07%	3.34%	4.32%	3.11%	2.59%

The distribution of PoS tags found in implicit aspect indicators is presented in Table 12. Each row represents a tag category found in Penn Treebank. The first line represents adjectives (JJ, JJR, JJS), the second the nouns (NN, NNS, NNP, NNPS), and the third the verbs (VB, VBD, VBG, VBN, VBP, VBZ). The IAI column shows how many vocabulary words were seen labeled with each label group. The third column describes how many words with the tag were found in sentences with implicit aspect indicators. The last column presents the distribution of labels observed as indicators of implicit aspects.

Table 12 – Distribution of PoS tags in the Cruz, Gelbukh and Sidorov (2014) corpus

POS	IAI	Tags/Sentence	P(IAI)
JJ	157	527	0.2818
NN	167	1,692	0.3000
VB	220	1,112	0.3836
Others	19	3,900	0.0346

Method description

The authors implemented four methods to identify indicators of implicit aspects, three of which are simpler methods to serve as a basis for comparison and a fourth based on supervised machine learning using the Conditional Random Fields (LAFFERTY; MCCALLUM; PEREIRA, 2001) algorithm.

- *Baseline 1 (BSLN1)*: the first method checks whether words appear in a sentiment lexicon (DING; LIU; YU, 2008). All words found in the text that have positive or negative tags in the lexicon are classified as indicators of implicit aspects.
- *Baseline 2 (BSLN2)*: the authors implemented a text classifier Naive Bayes in order to identify whether a sentence has at least an implicit aspect indicator. If so, opinion words, that is, those found in the sentiment lexicon, were labeled as indicators of implicit aspects. The classifier was trained with the words in the corpus (stop words excluded) and the 500 best bigrams obtained by an association measure Pointwise Mutual Information.
- *Baseline 3 (BSLN3)*: as a third comparative basis method, the authors implemented a second-order hidden Markov model as a classifier. To do this, they used their bigrams and trigrams as characteristics of the texts.

The fourth implementation using Conditional Random Fields tested several combinations of text attributes for model training:

- *Word characteristics (WT):* word combination and morphosyntactic tagging;
- *Characteristics of n-gram characters (CNG):* n-gram characters with a maximum of 6 characters;
- *Context characteristics (CNTX):* formed by the current word, before, and after the word with their respective morphosyntactic tags;
- *Class Sequential Characteristics (CLS):* combination of the current word and two previous words with their respective morphosyntactic tags;

Results

The authors observed that the characteristics of the words (WT) provided high precision, but low recall, according to the results presented in Table 13. Character n-grams (CNG) features increased recall but decreased accuracy. Context and class sequential features increased both accuracy and recall, and the best performance was achieved by combining all feature sets.

3.4 Deep learning methods

3.4.1 Cai *et al.* (2020)

The authors developed an aspect-based sentiment analysis method, focusing on implicit aspects. For this purpose, they used the so-called sentiment analysis based on categories of aspects. The authors differentiate some subtasks related to sentiment analysis.

Table 13 – Cruz, Gelbukh and Sidorov (2014) results

	Precision	Recall	F-measure
BSLN1	0.0381	**0.3158**	0.0681
BSLN2	0.1016	0.1379	0.1170
BSLN3	0.5307	0.1439	0.2264
CRF+WT	**0.6271**	0.0575	0.1053
CRF+CNG	0.4765	0.1925	0.2742
CRF+CNTX	0.5030	0.1148	0.1869
CRF+WT,CNG	0.4697	0.1992	0.2795
CRF+WT,CNG,CNTX	0.5209	0.2031	0.2932
CRF+WT,CNG,CNTX,CLS	0.5458	0.2064	**0.2970**

- **Aspect-based sentiment classification:** aims to detect the sentiment polarities of aspect terms mentioned in an opinionated sentence. The main limitation of this type of analysis lies in the fact that aspect terms need to be noted before classifying the aspect sentiment;

- **Sentiment analysis based on aspect terms:** performs aspect term extraction and aspect sentiment classification together;

- **Aspect category-based sentiment analysis:** involves joint detection of aspect categories and category-oriented sentiment classification.

Corpus

The experiments were performed on four English-language datasets on restaurant analytics and laptops used at the 2015 (PONTIKI *et al.*, 2015) and 2016 (PONTIKI *et al.*, 2016) conferences. The authors observed that the datasets have several analyses containing implicit aspects. Table 14 shows that almost 25% of the aspects in the dataset are implicit.

Table 14 – Implicit aspects of the SemEval 2015 and 2016 datasets.

	Implicit	Total	Percentage
Restaurant-2015	623	2,499	24.93%
Restaurant-2016	836	3,368	24.82%

Method description

The authors formulated the task as a hierarchical prediction problem of categories and sentiments, where first the categories of aspects of a sentence are identified, and then the sentiments related to each identified category. To do this, they used a hierarchical graph convolutional network (*Hier-GCN*). The hierarchical prediction problem has been studied in some previous works:

- (SILLA; FREITAS, 2011) defined the hierarchical classification task and presented new perspectives on the analyzed approaches;

- (ALY; REMUS; BIEMANN, 2019) investigated the use of hierarchical networks in multi-label text classification;

- (BRUNA *et al.*, 2014) proposed spectral graph convolutional neural networks;

- (KIPF; WELLING, 2017) designed a convolutional architecture via localized first-order approximation of spectral graph convolutions.

Inspired by these works, the authors proposed a Hier-GCN model that decomposes the internal relationships between categories and the interrelationships between sentiment polarities into two hierarchically related subgraphs and obtains the final predictions in a hierarchical manner.

To encode the sentences and their attributes, the authors used bidirectional encoder representations of transformers (BERT) (DEVLIN *et al.*, 2019), which is a pre-trained model with numerous texts and has achieved superior results in many natural language processing tasks. Let $H \in \mathbb{R}^{d\times(n+2)}$ denote the final hidden states generated by BERT, with two special tokens ([CLS] and [SEP]) inserted at the beginning and end of each entry r. To represent the categories, the authors used m separate self-attention sublayers on top of H to obtain the representations of the m categories denoted by $C \in \mathbb{R}^{d\times m}$. Furthermore, they employed the hidden state of the first token [CLS] as a shared representation of the sentiment of each category, denoted by $S \in \mathbb{R}^{d}$.

The model formulated by the authors contains three modules: in a lower module, a language model BERT was used to obtain hidden representations. In an intermediate module, they used a hierarchical graph convolutional network, where a lower-level graph convolutional network (GCN) was used to model the internal relationships between several categories, and the higher-level GCN to capture the interrelationships between categories of aspects and feelings. Based on the representations generated from Hier-GCN, the upper module performs hierarchical category and sentiment prediction to generate the final output.

Results

The authors implemented and compared the results of seven systems:

- *Cartesian-BERT*: performs the Cartesian product between aspect and sentiment categories, thus considering all possible pair combinations, and uses BERT as a sentence encoder;

- *Pipeline-BERT*: pipeline of aspect category detection and sentiment classification, both using BERT as the encoder;
- *AddOneDim-LSTM*: adds a label next to the sentiment to predict the presence or absence of each aspect category (SCHMITT *et al.*, 2018). In this way, the occurrence and sentiment prediction of each category are unified as a simple multi-class classification problem;
- *AddOneDim-BERT*: similar to the previous method but using BERT as sentence encoder;
- *Hier-BERT*: hierarchical method using BERT as a sentence encoder;
- *Hier-Transformer-BERT*: using Transformers to model the internal relationships between category and the interrelationships between category and sentiment based on Hier-BERT;
- *Hier-GCN-BERT*: proposed hierarchical prediction model with Hier-GCN as the relation learning module based on Hier-BERT.

All systems except AddOneDim-LSTM were proposed by the authors. Table 15 presents the results obtained. It is possible to notice that the AddOneDim-LSTM system found it difficult to deal with many categories. In the laptop data set, the algorithm failed to converge. When the method was used in conjunction with BERT the results were positive. The Cartesian BERT system also proved to be less effective, probably due to the substantial number of categories. The pipeline method achieved low precision but high recall, probably due to ignoring the relationships between the two subtasks. The hierarchical methods achieved significantly better F-measure performance than the other systems. This indicates that this type of modeling is best suited for aspect category detection in conjunction with sentiment classification.

Table 15 – Results of sentiment analysis based on aspect categories

Method	Restaurant-2016			Laptop-2016			Restaurant-2015			Laptop-2015		
	Prec	Rec	Med F	Prec	Rev	Med F	Prec	Rec	Med F	Prec	Rec	Med F
Cartesian-BERT	74.96	63.84	68.94	**64.99**	27.40	39.54	**72.02**	49.15	58.42	**73.06**	21.18	32.83
Pipeline-BERT	43.62	**79.06**	56.21	31.92	**51.56**	39.42	38.12	**70.00**	49.35	36.91	51.62	43.02
AddOneDim-LSTM	61.56	42.82	50.50	-	-	-	54.33	28.44	37.32	-	-	-
AddOneDim-BERT	71.75	67.95	69.79	58.83	39.49	47.23	68.84	55.86	61.67	64.17	39.57	48.94
Hier-BERT	70.97	69.65	70.30	59.51	41.93	49.19	67.46	57.98	62.36	65.47	41.26	50.61
Hier-Transformer-BERT	73.72	73.21	73.45	58.06	48.29	52.72	70.22	59.96	**64.67**	65.63	51.95	57.79
Hier-GCN-BERT	**76.37**	72.83	**74.55**	61.43	48.42	**54.15**	71.93	58.03	64.23	71.90	**54.73**	**62.13**

3.4.2 Wang *et al.* (2021)

In this work, the authors present a method to automate the search for the best way to concatenate different embeddings. To this end, they formulated the method inspired by recent progress in neural architecture research. The authors tested the method on different tasks, including aspect extraction.

Corpus

For the experiments, the authors used the English datasets in the laptop and restaurant domains of SemEval 2014 (PONTIKI *et al.*, 2014), the restaurant domain of SemEval 2015 (PONTIKI *et al.*, 2015), and the restaurant domain of SemEval 2016 (PONTIKI *et al.*, 2016). They also used 4 other languages in the restaurant domain of SemEval 16 to test a multilingual approach.

Method description

To find the best embedding set, the task model and a controller interact continuously. The task model predicts the output of the executed task, while the controller searches for the best combination of embedding concatenation for the test task to achieve the highest accuracy.

Given an embedded combination generated by the controller, the task model is thrilled and evaluated and returns a reward to the controller. The controller receives this reward, updates its parameters, and chooses a new combination of embeddings. The process was repeated 30 times and the embedding combination with the greatest accuracy was chosen to calculate the final results.

The authors used the following candidate embeddings: ELMo (PETERS *et al.*, 2018), Flair, Flair multilingual embeddings (M-Flair) (AKBIK; BLYTHE; VOLLGRAF, 2018), BERT base, Multilingual BERT (M-BERT) (DEVLIN *et al.*, 2019), GloVe word embeddings (PENNINGTON; SOCHER; MANNING, 2014), fastText word embeddings (BOJANOWSKI *et al.*, 2017), non-contextual character embeddings (LAMPLE *et al.*, 2016), and XLM-R (CONNEAU *et al.*, 2020).

Results

The method showed very good results, as shown in Table 16, reaching an f-measure of 87.4% in the domain of laptops and, in the domain of restaurants, 92.0%, 80.3%, and 81.3%. These results were obtained in the datasets of SemEval for the years 2014 (PONTIKI *et al.*, 2014), 2015 (PONTIKI *et al.*, 2015), and 2016 (PONTIKI *et al.*, 2016) respectively.

3.5 Other related works

3.5.1 Vargas and Pardo (2018)

In this work, the authors investigated six methods for grouping explicit and implicit aspects in product analyses. Four linguistic-based methods inspired by literature were

Table 16 – Comparison with state-of-the-art approaches

Method	14Lap	14Res	15Res	16Res	es	nl	ru	tr
Xu *et al.* (2018)	84.2	84.6	72.0	75.4	–	–	–	–
Xu *et al.* (2019)	84.3	–	–	78.0	–	–	–	–
Wang *et al.* (2020)	–	–	–	72.8	74.3	72.9	71.8	59.3
Wei *et al.* (2020)	82.7	87.1	72.7	77.7	–	–	–	–
XLM-R+Fine-tune	85.9	90.5	76.4	78.9	77.0	77.6	77.7	74.1
ACE+Fine-tune	**87.4**	**92.0**	**80.3**	**81.3**	**79.9**	**80.5**	**79.4**	**81.9**

tested, a statistical method (based on embeddings of words), and a new linguistic-based method proposal. The methods were tested in three different domains: cell phones, digital cameras, and books.

Corpus

To evaluate the aspect grouping methods, the authors manually annotated a corpus in Portuguese about cell phones, digital cameras, and book reviews. Three commonly used domains were selected to test how robust and generalizable the methods are. Each domain had 60 reviews. In each review, explicit and implicit aspects were tagged and grouped. Implicit aspects were indicated by the clue terms that signaled them. Table 17 presents more details of the corpus.

Table 17 – Corpus details

Domain	Aspects	Explicit	Implicit	Clusters
Books	103	91	12	21
Cameras	132	109	23	36
Cell Phones	180	142	38	48

As in this work, we will not implement aspect grouping methods, we chose to omit the description of these methods. On the other hand, we will use this corpus for our experiments.

3.6 Literature synthesis

Analyzing the works chronologically, we can observe the evolution of aspect extraction methods, starting from simpler algorithms based on word frequencies and rules until we reach recent implementations using deep learning and distributional models. Regarding the results, it is more difficult to make this comparison due to the different methodologies and corpus used. We hope that in our studies we can compare some of the methods presented appropriately, that is, using the same data sets and evaluation metrics. Table 18 presents a summary of the works analyzed, indicating whether:

- presented a new corpus;

- worked with explicit aspects;
- worked with implicit aspects;
- what type of approach is used by the authors;
- which data language is used.

Most of the works presented do not deal directly with implicit detection. Another point to note is the dependence on the domain, with methods closely linked to the context under study and vocabulary used. For the Portuguese language, there seems to be a delay in the methods used for the English language. This fact is probably due to the smaller number of resources available for the language. For example, in 2014 the first corpus was created in Portuguese with annotated aspects (FREITAS *et al.*, 2012), and only in 2018 was a corpus created with annotation of implicit aspects (VARGAS; PARDO, 2018).

We also noticed the scarcity of work using deep learning, in our searches, we only found the work of Aires *et al.* (2018). In our work, we intend to use deep learning methods in conjunction with more recent distributional models such as BERT and apply techniques such as fine-tuning and transfer learning. In the next chapters, we will present our published works.

Table 18 – Content covered in related works.

Papers	Topics Covered			Approach	Language
	Corpus	Explicit	Implicit		
Hu and Liu (2004)	X	X		Hybrid	English
Popescu and Etzioni (2007)		X	X	Hybrid	English
Cruz, Gelbukh and Sidorov (2014)	X	X	X	Machine Learning	English
Poria *et al.* (2014)		X	X	Rule-Based	English
Pavlopoulos and Androutsopoulos (2014)	X	X		Hybrid	English
Balage Filho and Pardo (2014)		X		Machine Learning	English
Panchendrarajan *et al.* (2016)		X	X	Hybrid	English
Vargas and Pardo (2018)	X	X	X		Portuguese
Cai *et al.* (2020)		X	X	Deep Learning	English
Costa and Pardo (2020)	X	X	X	Hybrid	Portuguese
Wang *et al.* (2021)		X		Deep Learning	English

In the next chapters we will continue presenting the papers resulting from the work of this doctorate, presented and published in national and international conferences and journals.

4 LEARNING RULES FOR AUTOMATIC IDENTIFICATION OF IMPLICIT ASPECTS IN PORTUGUESE

Our first work, entitled “Learning rules for automatic identification of implicit aspects in Portuguese” (MACHADO *et al.*, 2021) was presented on November 30th at the Symposium in Information and Human Language Technology (STIL 2021) (RUIZ; TORRENT, 2021), which took place online from November 29th to December 3rd, 2021. The work was published in the proceedings of the event[1].

STIL is a biannual language technology event supported by the Brazilian Computing Society (SBC) and the Brazilian Special Interest Group in Natural Language Processing (CE-PLN). It is a multidisciplinary conference covering a wide spectrum of disciplines related to human language technology, such as linguistics, computer science, psycholinguistics, and information science. It aims to bring together academic and industrial participants working in these areas. The event was held concurrently with BRACIS 2021 (10th Brazilian Conference on Intelligent Systems).

The work presented a rule-based method for identifying aspects. From a training set, the algorithm generates a set of rules based on PoS tagging and dependency analysis of the words present in the sentences. It was compared with two frequency-based (HU; LIU, 2004; MACHADO; PARDO; RUIZ, 2017) methods and another rule-based one (PORIA *et al.*, 2014). For the experiments, we used a dataset (VARGAS; PARDO, 2018) with opinionated texts about cameras, books, and smartphones. Below is the complete work, inserted here with the permission of the editor.

[1] https://doi.org/10.5753/stil.2021.17787

Learning rules for automatic identification of implicit aspects in Portuguese

Mateus Tarcinalli Machado[1], Thiago Alexandre Salgueiro Pardo[1], Evandro Eduardo Seron Ruiz[2], Ariani Di Felippo[3]

[1]Núcleo Interinstitucional de Linguística Computacional (NILC)
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo

mateusmachado@usp.br, taspardo@icmc.usp.br

[2]Faculdade de Filosofia, Ciência e Letras de Ribeirão Preto, Universidade de São Paulo

evandro@usp.br

[3]Núcleo Interinstitucional de Linguística Computacional (NILC)
Departamento de Letras, Universidade Federal de São Carlos

ariani@ufscar.br

Abstract. *This sentiment analysis work is focused on the task of identifying aspects, emphasizing the so-called implicit aspects, i.e., those that are not explicitly mentioned in the texts. For this, we analyzed frequency-based methods, adapted rules from the English language to Portuguese, and developed a method that learns new rules through corpus analysis.*

Resumo. *Este trabalho de análise de sentimentos está focado na tarefa de identificação de aspectos, dando ênfase aos chamados aspectos implícitos, ou seja, aqueles que não são mencionados explicitamente nos textos. Para isso, analisamos métodos baseados em frequência, adaptamos regras da língua inglesa para o português e desenvolvemos um método que aprende novas regras por meio de análise de corpus.*

1. Introduction

Sentiment analysis is an area of applied computing research related to natural language processing, which aims to analyze people's opinions, feelings, assessments, attitudes, and emotions concerning entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [Liu 2012]. It presents a wide range of studies and is often referred to in the literature by slightly different names and tasks, such as opinion mining, opinion extraction and sentiment mining, among others. According to [Taboada 2016], sentiment analysis is a growing field at the intersection between linguistics and computer science that aims to determine automatic sentiments in a text.

Many authors [Medhat et al. 2014, Rana and Cheah 2016, Yadollahi et al. 2017] divide sentiment analysis into three levels: document, sentence, and aspect. Each level aims to address different challenges, with aspect-based analysis being the most refined. At the document level, the goal is usually related to the identification of the polarity of the whole document, to determine whether its overall opinion is positive or negative. This level of analysis is important in social and psychological studies carried out on social networks, consumer satisfaction, analysis of patients in medical settings, and many others [Yadollahi et al. 2017]. Sentiment analysis performed at the sentence level aims to find the polarity of a sentence. At this level of analysis, it is also important to identify whether the target sentence is subjective or objective, i.e., whether it contains an opinion. This level of sentiment analysis is usually influenced by the context around the sentence and is considered very important for applications dealing with texts from social networks, comments and short messages, among others. Finally, the aspect level deals with more refined sentiment analysis strategies, trying to determine people's opinion about a specific characteristic (aspect) of a product, a service, or an entity [Medhat et al. 2014, Rana and Cheah 2016, Yadollahi et al. 2017]. To carry out this level of analysis, it is essential to identify the entities mentioned and their respective aspects and the so-called words of opinion related to the aspects. From there, the polarity of opinions directed to each identified aspect is determined. The results of the aspect-level sentiment analysis can then be finally summarized and visualized, for example, in a table containing the found aspects and their respective polarities [Rana and Cheah 2016]. This level of processing can also be called aspect-based sentiment analysis [Liu 2012] and is very useful for language processing applications, as text summarization [López Condori and Pardo 2017].

In this paper, we focus on the aspect detection task, emphasizing the treatment of implicit aspects. An aspect is considered explicit if it is mentioned directly in a sentence, otherwise, it is considered implicit [Liu 2012]. For example, in the sentence "The image quality of this camera is great", the expression "image quality" is an explicit aspect term. On the other hand, in "This camera is expensive", the word "expensive" implicitly refers to the "price" aspect.

It is a fact that we have a smaller amount of works that deal with implicit aspects [Ravi and Ravi 2015, Rana and Cheah 2016]. For Portuguese language, in particular, this is still a challenge [Pereira 2020]. Several authors [Zhang and Zhu 2013, Panchendrarajan et al. 2016] point out that between 15% and 30% of the aspects of opinionated texts are mentioned implicitly, and, for some categories, these percentages increase significantly. The non-detection of these aspects leads to the loss of important information and, consequently, negatively affects the results of related applications. The identification of implicit aspects and the treatment of texts in the Portuguese language, despite being complex problems, are relevant tasks and have the potential to advance the state of the art in the area.

This paper presents an analysis of frequency and rule-based methods for aspect detection. We also develop a method that learns extraction rules from a corpus. The rest of this work is organized as follows: Section 2 presents the main related work, especially the ones for Portuguese language. Section 3 presents our methods and datasets. Section 4 describes the results obtained by the implemented methods. Finally, Section 5 discusses the strengths and weaknesses of these methods.

2. Related work

There are few works that explicitly try to characterize, model and identify implicit aspects in the sentiment analysis area, which may be explained by the difficulties of dealing with them. We briefly synthesize the main related work in what follows.

The authors of [Cai et al. 2020] formulated the task as a problem of hierarchical prediction of categories and sentiments, where first the algorithm identify categories of aspects in a sentence and then the sentiments related to each detected category. To do this, they used a convolutional network of hierarchical graphs (Hier-GCN). To codify the sentences and their attributes, the authors used representations of bidirectional transformer encoders (BERT) [Devlin et al. 2018], which is a pre-trained model and has achieved excellent results in many natural language processing tasks. Their results show that this type of modeling is suitable for the detection of the aspect category together with the classification of sentiments, achieving in its highest result 0.76 precision, and 0.73 recall.

In [Marcacini et al. 2018], the authors developed a method based on heterogeneous networks that combines features as labeled aspects, unlabeled aspects, and linguistic features. To perform the classification, they developed an algorithm that propagates the labels using linguistic features as a bridge for this propagation. As a result, the method obtained f-measure between 0.56 and 0.68, depending on the analyzed domain.

The work of [Balage Filho 2017] analyzed the following aspect extraction methods: frequency-based, relation-based, and machine learning-based ones. The frequency-based method is the simplest. It selects the most frequent nouns and noun phrases as aspects. The relation-based one analyzes the relationships between aspects and the opinion related to them. In machine learning, an annotated training set is used to create a model that identifies the aspects. In their experiments, the frequency-based method showed the best result with 0.49 f-measure.

In the work of [Costa and Pardo 2020], the authors analyzed lexical-based methods for extracting aspects from opinions written in Portuguese. In a first experiment, the authors made use of aspect ontologies, in order to identify explicit and implicit aspects. In a second experiment, the authors extended a frequency detection method of aspects using the distributional models Word2Vec [Mikolov et al. 2013a] in order to enrich the process. The experiment performed with ontologies achieved the best results, with 0.53 precision, and 0.44 recall.

3. Data and Methods

3.1. Methods

In this work, we applied four methods for aspect detection: two frequency-based and two rule-based methods. For the frequency-based ones, we implemented the Freq Baseline and Freq Baseline with Word2Vec methods to serve as comparative bases for other methods. Regarding the rule-based methods, we tried to translate and apply the rules discovered in [Poria et al. 2014] to texts in Portuguese, as well as to use an automatic method for learning new detection rules. Before detailing the methods, we introduce the dataset that we use.

3.2. Dataset

For the execution of the experiments, we worked with a corpus formed by opinionated texts in Portuguese, resulting from the work of [Vargas and Pardo 2018]. The corpus is one of the few existent corpora with identified implicit aspects, even considering works in English. It has 60 product reviews from 3 different domains: cameras, books, and smartphones. Table 1 shows the characteristics of the corpus. It is important to note that the percentages of implicit aspects found, with values close to 15% in the domains of cameras and smartphones, demonstrate the relevance of the specific analysis of this category of aspects. We evaluated all experiments in this corpus.

Table 1. Corpus composition.

Domains	Reviews	Aspects	Explicits	% Explicits	Implicits	% Implicits
Cameras	60	352	299	84.94%	53	15.06%
Books	60	330	304	92.12%	26	7.88%
Smartphones	60	455	387	85.05%	68	14.95%

In the corpus we found different cases of implicit aspects. For example, in "Very compact and very beautiful", the size and design aspects can be identified by their qualifiers. In "Camera of small measurement", the size aspect appears as a semantically close word. Finally, in "Works anywhere", only with specific knowledge about the context, in this case the functioning of the smartphone, we can understand that the phrase refers to the signal aspect. These are just a few examples, which demonstrate how complicated the task of identifying implicit aspects can be.

In a pre-processing stage, we prepared the texts for the execution of the methods. We selected these processes according to the needs of the algorithms. In our case, we performed the following processes:

- **Dataset division into training and test sets**: with the aim of consistently comparing the implemented methods, we performed all experiments using the same training (for searching parameters or rules, as we will see later) and testing (to validate the methods) datasets;
- **Sentences tokenization:** split of texts into sentences;
- **Words tokenization:** split of sentences into tokens, which can be words, symbols and punctuation, among others;
- **PoS tagging:** identification of morpho-syntactic categories of the tokens;
- **Dependency analysis:** construction of sentence dependency trees, to understand the relationship between tokens.

To perform these tasks, except for the division in training and testing sets, we used spaCy [Honnibal et al. 2020] module with pt_core_news_lg model from the Python programming language.

3.3. Method 1: Freq-Baseline

This method selects nouns and noun phrases as aspect candidates and analyzes their frequencies in relation to the number of sentences in the analyzed dataset. The algorithm selects the most frequent candidates as aspects. It is a method with an easy implementation that achieves good results, which is why it is often used as a basis for comparison to evaluate other algorithms.

The definition of the most frequent aspects is performed by comparing the frequency of the candidates being analyzed with a pre-defined cutoff frequency. In [Hu and Liu 2004], the authors used a cutoff frequency of 1%, a value that ended up being used in other studies as well. As in [Machado et al. 2017], in this work, we varied this cutoff frequency and analyze what would be the most appropriate value. We worked with cutoff frequencies in an interval of 0.01% and 10% with an increment of 0.01% at each execution of the algorithm. Thus, at each iteration, we performed the calculations of precision, recall, f-measure, and percentages of explicit and implicit aspects correctly detected.

3.4. Method 2: Freq-Baseline + Word2Vec

As in [Pavlopoulos and Androutsopoulos 2014] and [Machado et al. 2017], in this work we used the distributional model Word2Vec [Mikolov et al. 2013b] to exclude candidates for aspects that are not related to the domain under analysis. The algorithm compared each candidate aspect with two vectors formed by the centroid of the Word2Vec vectors of words related to the context under study and of the Word2Vec vectors of general domain words. The candidate aspects closest to this general centroid vector were discarded, thus remaining only words more related to the context of the dataset. As it is a method used only as a baseline, we will not go into further details about it. More information can be found in the mentioned references. As in the previous experiment, we searched for the best cutoff frequencies, which remained the same.

3.5. Method 3: Adapted Rules of [Poria et al. 2014]

In our first experiment with rules, we adapted the rules from the work of [Poria et al. 2014]. The method comprises handcrafted rules that analyze the sentence dependency trees together with two lexicons: one with a list of implicit aspects and another called SenticNet [Cambria et al. 2014] that is a concept-level knowledge base containing a set of semantics, sentics, and polarities associated to natural language concepts.

As we did not find an implicit lexicon for the Portuguese language, we created one based on the existent corpus of [Vargas and Pardo 2018]. To maintain consistency with the other experiments and do not introduce bias in the results, we used only the terms found in the training set.

We translated the sentences of the presented examples in [Poria et al. 2014] and adapted the rules to Portuguese, changing the order or type of some components. Finally, we executed them in the adopted dataset.

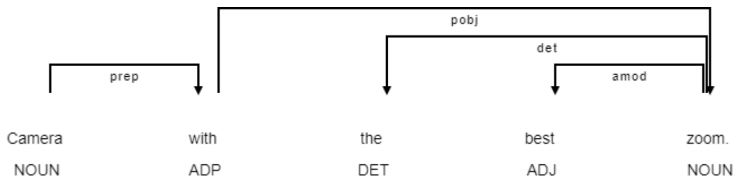
3.6. Method 4: Automatically Learned Rules

The results of the adaptations of the rules developed in [Poria et al. 2014] for Portuguese were below expectations. This motivated us to look for more adequate rules for the Portuguese language and with better detection of implicit aspects. For this, we used the dependency trees of the sentences of the training set.

We analyzed the elements that had some relationship of dependency on the aspects identified in the datasets and, together with the morpho-syntactic analysis performed, we created a set of rules. As an example, Figure 1 presents the dependency tree, obtained from spaCy, of the sentence “Camera with the best zoom.”. In this sentence, the term “zoom” is marked as an aspect, therefore, for the formation of the rules, the terms that have some relationship with it were used.

In the first stage, the elements that have a direct relationship with the marked aspects were selected. The relations and morpho-syntactic characteristics of the elements

Figure 1. Example of dependency tree.



were combined, forming aspect extraction rules. For example, see the rules presented in Table 2. We can interpret the first rule of this table as follows: if a noun (NOUN) is in an adjectival modifier relationship (amod) with an adjective (ADJ), then the first token is extracted as aspect. This same process was repeated, but this time analyzing combinations of three, four, and five tokens, thus generating 1,200 rules for camera domain, 1,470 for books, and 1,564 for smartphones.

Table 2. Example of learned rules.

Tokens		Rules			
Token 1	Token 2	Class 1	Relation	Class 2	Aspect
zoom	best	NOUN	amod	ADJ	Token 1
		*	amod	ADJ	Token 1
		NOUN	*	ADJ	Token 1
		NOUN	amod	*	Token 1
		NOUN	*	*	Token 1
		*	amod	*	Token 1
		*	*	ADJ	Token 1
with	zoom	ADP	pobj	NOUN	Token 2
			⋮		

After the execution of the rule generation process with all the sentences in the corpus, the next step was the selection of the best rules. For that, the rules were created using the class DependencyMatcher of the module spaCy in the language Python, and each one was executed in the training set.

We calculated, for each rule, the Laplace precision and support metrics. We tested rules with a support value between 1 and 60 combined with a Laplace precision ranging between 0.1 and 1, with an increment of 0.1. For each combination of Laplace precision and support, we selected the rules that met these requirements, ran them on the test suite, and calculated the precision, recall, and f-measure metrics. To select the best set of rules, we choose the rules that got the highest result for f-measure.

4. Results

Our first experiments should serve as baseline results. Although simple, we already have a first contribution with the analysis of cutoff frequencies. As mentioned, several authors use the value of 1% as a cutoff. We find a cutoff frequency between 0.40% and 0.53% as a more appropriate value. Using the pruning mechanism with Word2Vec, as in other works, led to an increase in accuracy and a decrease in recall. The identification of implicit aspects, as expected because of the characteristics of the method, was practically non-existent. The results are summarized in Table 3.

Table 3. Results for Frequency-based methods.

Domain	Method	% Cutoff	Precision	Recall	F-measure	% Explicits	% Implicits
Camera	Freq	0.53	0.57	0.57	0.57	35.16	0.00
Camera	Freq+W2V	0.53	0.72	0.56	0.63	34.07	0.00
Book	Freq	0.42	0.47	0.67	0.55	23.17	0.00
Book	Freq+W2V	0.42	0.70	0.59	0.64	15.85	0.00
Smartphone	Freq	0.40	0.46	0.53	0.49	27.61	4.17
Smartphone	Freq+W2V	0.40	0.56	0.47	0.51	23.88	0.00

As mentioned, the results of the experiment with the rules adapted from [Poria et al. 2014] did not reach satisfactory results. Therefore, we focused on finding new rules that would be more appropriate for the Portuguese language. The results are summarized in Table 4. For the camera domain, we got the best f-measure result with 26 rules selected by a minimum Laplace precision of 0.30 and support of 6. Regarding the book domain, we only found 1 rule selected by the Laplace precision of 0.40 and support of 21. Finally, in the smartphone domain, we found 3 rules selected by the Laplace precision of 0.10 and support of 18.

Table 4. Results for learned rules.

Domain	Laplace[1]	Support	Rules	Precision	Recall	F-measure	% Explicits	% Implicits
Camera	0.30	6	26	0.50	0.58	0.54	36.26	15.38
Book	0.40	21	1	0.81	0.54	0.65	13.41	20.00
Smartphone	0.10	18	3	0.61	0.42	0.50	20.90	4.17

Analyzing the learned rules, we could observe that most of the rules show nouns as aspects, changing only the classes and relationships with other terms.

We also selected the common rules between the domains, to find rules that were theoretically less susceptible to the context. Table 5 shows the rules that were found, and, as we can see, there was a greater number of rules found between the domains of cameras and smartphones because of the greater similarity between them. The common rules between smartphone and book were the same as found between camera, smartphone, and book. The first rule can be interpreted as: any token in a nominal subject relation with a noun that is in a determiner relation with a determiner.

Table 5. Common rules between domains.

Sets	Rule	Aspect
camera ∩ smartphone	['NOUN', 'nsubj'], '>', ['DET', 'det']	Token 1
	['NOUN', 'obj'], '>', ['DET', 'det']	Token 1
	['NOUN', 'conj'], '>', ['PUNCT', 'punct']	Token 1
camera ∩ smartphone ∩ book	['NOUN', 'nsubj'], '>', ['DET', 'det']	Token 1

The results were inferior to the previous experiment, as we can see in Table 6. There was no improvement, even in terms of accuracy, that could have increased with a smaller number of rules. This result was probably due to the simplicity of the found rules, which, as previously mentioned, always ended up selecting only nouns as aspects, similar to what occurs in the frequency-based methods.

Table 6. Results for rules common to all the domains.

Domain	Rules	Precision	Recall	F-measure	% Explicits	% Implicits
Camera	Camera ∩ Smartphone	0.49	0.31	0.38	15.38	0.00
Camera	Camera ∩ Smartphone ∩ Book	0.62	0.12	0.21	5.49	0.00
Book	Camera ∩ Smartphone	0.53	0.57	0.55	17.07	20.00
Book	Camera ∩ Smartphone ∩ Book	0.81	0.54	0.65	13.41	20.00
Smartphone	Camera ∩ Smartphone	0.61	0.42	0.50	20.09	4.17
Smartphone	Camera ∩ Smartphone ∩ Book	0.73	0.24	0.36	10.45	0.00

Finally, Table 7 presents a summary of the results got by each method. We can observe that only the rules we found could find implicit aspects in all domains. Considering f-measure, in the camera domain, the frequency-based method with Word2Vec was superior and, in the other domains, it was practically equivalent to the results of the rules, although it was not able to find implicit aspects.

Table 7. Results by implemented method.

	Camera		Book		Smartphone	
Method	F-measure	% Implicits	F-measure	% Implicits	F-measure	% Implicits
Freq	0.57	0.00	0.55	0.00	0.49	4.17
Freq + Word2Vec	**0.63**	0.00	0.64	0.00	**0.51**	0.00
Poria et al. rule-based	0.10	0.00	0.14	0.00	0.14	**7.35**
Corpus-based rules	0.54	**15.38**	**0.65**	**20.00**	0.50	4.17

5. Conclusion

Regarding the frequency-based methods, we could observe that, despite being simple, they achieved interesting results. With the study of the cutoff frequency, we could still get a significant improvement in these results. Using Word2Vec also proved to be effective in eliminating some aspects that were wrongly identified, a fact observed by the increase in precision. On the other hand, it ended up eliminating the few implicit aspects that had been identified.

The adaptation of the rules from English to Portuguese proved inefficient. Although at first the analysis of the examples of the rules proved promising, in the experiments, the results were below expectations. The method for finding rules in our last experiment has shown promise. Although we did not get a vast improvement in the f-measure, the improvement in implicit aspect detection was substantial. In future work, we intend to consider additional elements for analysis and creation of rules, such as sentiment lexicons, and to separately search for rules for explicit and implicit aspects.

The interested reader may find more information at the web portal of the POeTiSA (*POrtuguese processing - Towards Syntactic Analysis and parsing*) project[2].

Acknowledgements

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

[2]https://sites.google.com/icmc.usp.br/poetisa

References

Balage Filho, P. P. (2017). *Aspect extraction in sentiment analysis for portuguese language*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brazil.

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 (Online). International Committee on Computational Linguistics.

Cambria, E., Olsher, D., and Rajagopal, D. (2014). Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

Costa, R. W. M. and Pardo, T. A. S. (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. SBC.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

López Condori, R. E. and Pardo, T. A. S. (2017). Opinion summarization methods: Comparing and extending extractive and abstractive approaches. *Expert Systems with Applications*, 78:124–134.

Machado, M. T., Pardo, T. A. S., and Ruiz, E. E. S. (2017). Analysis of unsupervised aspect term identification methods for portuguese reviews. *Anais do XIV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC*, pages 239–249.

Marcacini, R. M., Rossi, R. G., Matsuno, I. P., and Rezende, S. O. (2018). Cross-domain aspect extraction for sentiment analysis: A transductive learning approach. *Decision Support Systems*, 114:70–80.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.

Panchendrarajan, R., Ahamed, N., Murugaiah, B., Sivakumar, P., Ranathunga, S., and Pemasiri, A. (2016). Implicit aspect detection in restaurant reviews using cooccurence of words. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 128–136.

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 (LASM)@ EACL*, pages 44–52.

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

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 second workshop on natural language processing for social media (SocialNLP)*, pages 28–37.

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*, 2:325–347.

Vargas, F. A. and Pardo, T. A. S. (2018). Aspect Clustering Methods for Sentiment Analysis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11122 LNAI:365–374.

Yadollahi, A., Shahraki, A. G., and Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):1–33.

Zhang, Y. and Zhu, W. (2013). Extracting implicit features in online customer reviews for opinion mining. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 103–104.

5 IMPLICIT OPINION ASPECT CLUES IN PORTUGUESE TEXTS: ANALYSIS AND CATEGORIZATION

Our second work was the article “Implicit opinion aspect clues in Portuguese texts: analysis and categorization” (MACHADO *et al.*, 2022) which was presented at the International Conference on Computational Processing of Portuguese Language (PROPOR 2022) (PINHEIRO *et al.*, 2022), which took place between May 21 to May 23, 2022. The paper was published in the proceedings of the event[1].

PROPOR has been held since 1993 and is the main conference in the area of Portuguese Language Technology, aimed at presenting the results of academic and technological research and the cooperation of research groups in the area. It is currently held every two years, sometimes in Brazil, sometimes in Portugal.

In this work, we aimed to understand how implicit aspects are formed in texts and their characteristics. To do this, we analyzed texts from two corpora Vargas and Pardo (VARGAS; PARDO, 2017) and Freitas and Vieira (FREITAS; VIEIRA, 2015). For this last, we manually annotated the Implicit Aspect Clues (IACs), which are words or expressions that refer to the implicit aspects, thus creating a new set of data for implicit analysis.

The Implicit Aspect Clues (IACs) were analyzed to understand the type of knowledge necessary to identify the corresponding aspect. The result was the creation of a typology of implicit aspects with these classified into categories. As we will see in the next sections, this typology will allow us to understand the weaknesses and strengths of aspect extraction methods in relation to implicit ones. Below is the complete work, inserted here with the permission of the editor.

[1] https://link.springer.com/chapter/10.1007/978-3-030-98305-5_7

Implicit opinion aspect clues in Portuguese texts: analysis and categorization

Mateus Tarcinalli Machado*[1] [0000−0003−1848−3082],
Thiago Alexandre Salgueiro Pardo[1] [0000−0003−2111−1319],
Evandro Eduardo Seron Ruiz[2] [0000−0002−7434−897X],
Ariani Di Felippo[3] [0000−0002−4566−9352] and
Francielle Vargas[1] [0000−0003−3164−2682]

[1] Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo,
São Carlos, Brazil
mateusmachado@usp.br, taspardo@icmc.usp.br, francielleavargas@usp.br
[2] Department of Computing and Mathematics, University of São Paulo,
Ribeirão Preto, Brazil
evandro@usp.br
[3] Interinstitutional Center for Computational Linguistics (NILC)
Department of Language and Literature, Federal University of São Carlos,
São Carlos, Brazil
arianidf@gmail.com

Abstract. Although very useful, aspect-based sentiment analysis shows several challenges. The occurrence of implicit aspects is one of them. To improve our understanding of implicit aspects, this work analyzes their composition and categorizes the implicit aspect clues found in two corpora of opinionated texts of different domains in the Portuguese language. As results of this work, in addition to the implicit aspect typology, we present four lexicons of implicit aspect clues for different domains and the annotation of implicit aspect clues in an existing corpus.

Keywords: Sentiment analysis · Implicit aspects · Lexicon · Corpus annotation.

1 Introduction

Sentiment analysis is a growing research area of natural language processing, at the intersection between Linguistics and Computer Science. It aims to automatically determine sentiments in texts [19], analyze people's opinions, feelings, assessments, attitudes, and emotions concerning entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [9].

The most refined sentiment analysis level is the aspect-based one, which must determine people's opinion about specific features (the aspects) of an entity [11,

* Author to whom correspondence should be addressed.

16, 25]. To do this, it is necessary to identify the entities, their respective aspects, and the so-called opinion-words related to each aspect. Most systems that perform aspect-based sentiment analysis divide the problem into three steps [14]:

– **Aspect term extraction**: this step identifies the aspect terms. For example, analyses of opinionated texts about a smartphone might reveal "measurements", "volume", and "response time" aspects [9, 10, 18, 26].
– **Aspect Term Sentiment Estimation**: this step estimates the polarity and/or the intensity of opinion for each aspect term of the entity [9, 13, 20].
– **Aspect Aggregation**: some systems group aspect terms that are synonymous or near synonymous (e.g., "price" and "cost") or group aspect terms ("chicken", "steak" and "fish" can be replaced by "food") [9, 10, 27, 28].

A challenging issue related to the aspect term extraction step is how the aspects appear in the texts. They may be explicit or implicit aspects. Explicit aspects are those directly mentioned in the text, such as the aspect "size" in "The size of the smartphone is adequate". On the other hand, implicit aspects are not directly referenced, as in the examples "The smartphone is too small" and "The smartphone barely fits in my pocket", where "size" is signaled by the adjective "small" and the expression "fits in my pocket". These terms or expressions that refer to implicit aspects are called Implicit Aspect Clues (IACs). We use a definition similar to that established by Cruz, Gelbukh, and Sidorov [1], where, from a predefined set of aspects, we find the related IACs, differing from Liu [9] that defined explicit aspects as nouns and noun phrases and other expressions as implicit aspects. According to Zhang and Zhu [29], about 30% of the reviews in their corpus contain implicit aspects. Panchendrarajan et al. [12] says that over 15% of the sentences in their corpus had one or more implicit aspects and that 92% of the aspects related to restaurant employees were implicitly mentioned. Such numbers show the relevance of the phenomenon.

Given the difficulty of dealing with implicit aspects, most of the researches in the area addresses only explicit aspects. A survey done by K. Ravi and V. Ravi [17] evaluated 160 papers on sentiment analysis and reported that 23 performed aspect extraction and only 4 analyzed implicit aspects. Rana and Cheah [16] evaluated 45 researches related to aspect extraction and found only 7 studies that analyzed the issue of the implicit aspects. Besides the difficulty, one may also explain the lack of researches that deal with implicit aspects by the high frequency of the explicit ones, which are enough for several applications.

Another important issue to be addressed is the language. Most works analyze texts written only in English or Chinese. In the papers of Tubishat, Idris, and Abushariah [21] and Ganganwar and Rajalakshmi [5], 53 articles dealing with implicit aspects were analyzed, of which 33 applied to texts in English, 19 in Chinese, and 1 in French. In Portuguese, the number of works is scarce. In the survey carried out by Pereira [15], the author found only two papers dealing with implicit aspects, namely, Freitas and Vieira [3] and Vargas and Pardo [22].

In order to better understand the issue, this paper presents a deep analysis and categorization of the IACs found in two corpora of opinionated texts of different domains in the Portuguese language, analyzing the composition of IACs and the necessary inferences to identify the corresponding aspects. As far as we know, there is no other initiative that does a similar work. We also introduce four lexicons of IACs and annotate a corpus with the related aspects. We hope that such effort and the resources that it brings may contribute to improve sentiment analysis applications for Portuguese.

In what follows, we briefly introduce the main related work, especially the ones that present corpus analyses for Portuguese. In Section 3, we present the analyzed data and the methods that we used, introducing our lexicons of IACs. In Section 4, we report a summary of everything that we found in our corpus analysis. Finally, in Section 5, we present some final remarks.

2 Related work

As commented before, few works deal with implicit aspects. Fundamental to any work is the existence of a corpus containing annotated implicit aspects. We selected some of these few works that present this type of dataset.

In Cruz, Gelbukh, and Sidorov [1], the authors argue about the scarcity of resources and, to solve this gap, they manually annotated IACs in Hu and Liu [8] corpus for English. The authors implemented four methods to identify IACs, three simpler methods to serve as baselines, and a fourth based on supervised machine learning.

Vargas and Pardo [22] organized and clustered aspects in opinions in Portuguese. As a result, the authors presented one of the few existent corpora with identified implicit aspects for this language. The authors [23] also investigated methods for grouping explicit and implicit aspects in opinion texts, using the previous corpus. They compared six clustering methods: four linguistic methods, one statistical, and a new one that analyzes the relations between detected aspects, such as synonyms, hypernyms and diminutives, among others. Finally, the authors [24] describe a set of rules for extracting explicit and implicit aspects using psychological verbs and semantic relations.

Freitas and Vieira [3] evaluated available resources for the analysis of feelings in the Portuguese language. For this, they formed a dataset with reviews about hotels, manually marking the explicit aspects, the existence of implicit ones, and the related feeling. The authors compare different part of speech taggers, sentiment lexicons, the impact of linguistic rules, and the position of adjectives. The authors [4] also present a method for aspect-based sentiment analysis, in which the identification of explicit and implicit aspects is based on a domain ontology.

3 Data and Methods

3.1 Methods

As mentioned before, the resources that incorporate the identification of implicit aspects are scarce. To improve this scenario, we analyzed the dataset presented by Freitas and Vieira [3] and performed the identification of the IACs in this corpus, thus creating another dataset suitable for this type of task. We also categorized all IACs and those present in the research of Vargas and Pardo [22], creating a typology related to the composition of each term and the necessary inferences to identify the aspects.

3.2 Datasets

We analyzed the datasets of Vargas and Pardo [22] and Freitas and Vieira [3]. The first one has 180 reviews of cameras, books, and smartphones extracted from Buscapé [6] and ReLi [2] corpora. The authors manually marked the explicit and implicit aspects and grouped the similar ones. The second corpus has 194 reviews about hotels collected from the TripAdvisor website. The explicit aspects and the presence of implicit aspects in the texts were manually annotated, but IACs were not tagged.

3.3 Identification of IACs

We initially analyzed the 194 reviews from Freitas and Vieira [3], and located the respective IACs in the texts. To do this, we aimed to get the smallest number of terms or words that might identify the implicit aspect. For this purpose, a human annotator marked all the terms that implicitly referred to the aspects in the corpus.

For example, the review that contains the sentence "*O preço indica o que receberá como serviço: um hotel simples*" (The price indicates what you will receive as a service: a simple hotel.) was already marked with the presence of an implicit aspect related to "instalation" and, in this case, the annotator marked the expression "*hotel simples*" (simple hotel) as the related IAC. In a second step, the IACs were checked for cases of dubiousness (when they appeared related to other aspects) and minimality (if they were in their shortest viable form). For example, it was checked whether "*hotel simples*" appeared related to some other aspect and also whether the expression could be reduced to "hotel" or "simples" in a way that it would indicate only one aspect. For example, if we consider only the word "*hotel*" as IAC, we can find it related to other aspects, such as "cleanliness" in "*o hotel é limpo*" (the hotel is clean), or not related to any aspect, as in "*tive que ficar neste hotel*" (I had to stay at this hotel). The same happens with the word "*simples*" (simple) in "*café da manhã simples*" (simple breakfast), where the word is related to the aspect "meal", and in "*uma variedade de locais simples para comer*" (a variety of simple places to eat), where it is not related to any aspect. Therefore, only the complete expression ("hotel simples") is related to the aspect "installation". After these identification steps, the list of IACs was once more reviewed by other humans (also from the Computational Linguistics area), in order to possibly refine it.

3.4 Lexicons of IACs

After the identification of IACs, we analyzed the datasets to create lexicons of IACs. These resources may be very useful for developing tools for identifying implicit aspects and improving the results of sentiment analysis tasks.

We analyzed the IACs and grouped them by domain, removing and correcting any eventual spelling errors. We could then create four lexicons, one for each domain. IACs were grouped according to the aspects they refer to. A summary of the composition of the corpora and lexicons of IACs is shown in Table 1.

Table 1. Statistics for corpora and lexicons

Domains	Reviews	Words	Aspects	Explicit aspects		Implicit aspects		Lexicon
Cameras	60	3,997	352	299	84.94%	53	15.06%	39
Books	60	5,515	330	304	92.12%	26	7.88%	24
Smartphones	60	6,210	455	387	85.05%	68	14.95%	48
Hotels	194	13,940	1,417	999	70.50%	415	29.50%	213

As illustration, Figure 1 shows part of our lexicon of IACs for camera domain. In the example, the IACs "*filma em hd*" (shots in hd) and "megapixels" are related to the aspect "resolution" of the camera.

```
<Resolution>
  <item>filma em hd (shoots in hd)</item>
  <item>megapixels</item>
</Resolution>
<Weight>
  <item>leve (weightless)</item>
  <item>leveza (weightlessness)</item>
</Weight>
```

Fig. 1. Part of the lexicon of IACs related to cameras

In a second analysis, we verified the part of speech tags of the terms identified as aspects. For this, we used the tagger in spaCy [7]. Table 2 presents these results, and we can see a dominance of nouns for the explicit aspects. For the implicit ones, this distribution is more diversified, but we can observe a higher frequency of NOUNs, ADJs, and VERBs. We also noticed that there are terms of different part of speech tags that refer to the same aspects, such as the nouns "*pagamento*" (payment) and "*custo*" (cost), the adjectives "*caro*" (expensive) and "*barato*" (cheap) and the verb "*pagar*" (to pay), all related to "price".

Table 2. The part of speech tags of the aspect terms

Explicit aspects											
Camera			**Book**			**Smartphone**			**Hotel**		
TAG	**QTY**	**%**	**TAG**	**QTY**	**%**	**TAG**	**QTY**	**%**	**TAG**	**QTY**	**%**
NOUN	337	73.4%	NOUN	228	64.4%	NOUN	358	70.1%	NOUN	835	81.1%
ADP	56	12.2%	PROPN	99	28.0%	PROPN	78	15.3%	ADP	91	8.8%
PROPN	42	9.2%	ADP	10	2.8%	ADP	34	6.7%	PROPN	71	6.9%
ADJ	17	3.7%	VERB	7	2.0%	ADJ	33	6.5%	ADJ	30	2.9%
X	3	0.7%	DET	6	1.7%	X	2	0.4%	VERB	2	0.2%
Implicit aspects											
NOUN	28	32.2%	VERB	18	28.6%	ADJ	35	31.8%	NOUN	494	49.1%
ADJ	24	27.6%	NOUN	16	25.4%	VERB	25	22.7%	ADJ	134	13.3%
VERB	13	14.9%	ADP	8	12.7%	NOUN	23	20.9%	ADP	111	11.0%
SCONJ	9	10.3%	DET	5	7.9%	ADP	9	8.2%	VERB	95	9.4%
PROPN	6	6.9%	PRON	4	6.3%	SCONJ	8	7.3%	ADV	69	6.9%

3.5 Categorization of IACs

To better understand the composition of implicit aspects and the necessary inference to identify them, we analyzed each of the IACs. In particular, we aimed to identify what type of information, or characteristics of the text, led to the identification of the aspects. For example, in the sentence "*A camera é muito fácil de usar*" (The camera is very easy to use), we have the IAC "*fácil de usar*" (easy to use) related to the "usability" aspect. Here, we can see that the IAC makes use of an action that refers to the implicit aspect. In other situations, the identification of this relationship can be more complicated, as in the IAC "*auto explicativa*" (self-explanatory), also related to the "usability" aspect.

A human annotator made an initial categorization of the IACs. Other annotators (computational linguists) then reviewed the identified categories. After this step, the annotator made a second round of annotation, refining the categories into subcategories, and the discussion and review process was repeated. With this analysis, we came with four main categories of IACs: event, features, qualification and contextual-related IACs. Table 3 presents these categories, possible subcategories that we also identified, and examples. We explain the categories in what follows.

Event (Action/Process/State): the identification of IACs takes place through the identification of actions, processes, or states related to the aspect. It is not necessarily a verb; it can be a term that refers to the verb tha it is related to the aspect. Here, the IAC may be sub-classified as "non-verbal form", or, when this type of IAC is directly identified by the verb, we classified them as "verb". For this category, it is necessary that the person use, in addition to knowledge about the context, some knowledge about the language. For example, in"*pagamento*" (payment), it is necessary to identify that the noun is related to the verb "to pay", and that both are actions related to the "price" aspect. In other situations, this relationship is not so clear, for example, in "*recebi chamadas até na beira do* *rio*" (received calls even on the riverbank), the expression refers to the aspect "signal" of a smartphone and its main part is the expression "to receive calls".

Feature: the IACs of this type are directly related to the aspect or part of the aspect. They can be found as "attributes", or related to the aspects in the following relations: "equivalence", "is-a", or "part-of". For the attribute subcategory, the person needs to know the characteristics of the aspect, for example, the IACs "*material*" (related to the "design" aspect) and "*a recepção é muito boa*" (the reception is very good) ("signal" aspect). In the equivalence relationship, the IAC has a very similar meaning to the aspect, as in "*higiene*" (hygiene) ("cleanliness" aspect) and in "*cada centavo investido*" (every penny invested) ("price" aspect). In the is-a subcategory, we have items related to the aspect, such as the IAC "*café da manhã*" (breakfast), which is a type of "food", or "*pessoal que atende na portaria*" (people who work at the concierge), where the person refers to the concierge staff who is a type of "employee". Lastly, in the part-of case, the IAC is part of the aspect, as in "*banheiro*" (bathroom) or "*local para guardar roupas*" (place to store clothes), both related to hotel "facilities".

Qualification: the IAC is presented as a quality or sentiment about the aspect, without directly mentioning it. In most of the cases, the IAC contains an "adjective" as its major component. In other cases, the IAC appears in an "equivalence" relation to an adjective related to the aspect. We also found IACs represented by adjectives converted to "nominal forms". To differentiate adjectives and nominal forms, it is necessary to make use of knowledge about the language. We have, for example, the IAC "*bonita*" (beautiful), that is an adjective, and "*beleza*" (beauty), that is a nominal form, both related to "design". In the equivalence subcategory, we find IACs as "*hotel simples*" (simple hotel), where the combination of the adjective and the noun refers to the aspect (in this case, "facilities").

Contextual: to relate an IAC of this category to its respective aspect, it is necessary to access world knowledge about the product or service being analyzed, such as its operation, modes of use, or content. We have subdivided this category into "location" and "related". In the "location" subcategory, we classified the IACs related to the place where the product is located. This reference to location can be direct as in "*fica na região central*" (is in the center of the region) or indirect as in "*consegue ir a pé até o mercado público*" (can walk to the public market). In any case, the reader must access knowledge of expressions related to places in order to correctly identify the aspect. The "related" subcategory includes other difficult cases that do not refer to location. For example, in "*sociedade do big brother*" (big brother society), some knowledge about the content of a book is required. In the case of "*cheiro de mofo*" (musty smell), the person needs to be aware that the lack of cleaning causes such smell in order to be able to relate the IAC to the cleanliness aspect.

Table 3. Categories, subcategories, aspects and examples of IACs

Category	Subcategory	Aspect	Corresponding IAC
Event	Non-verbal form	Usability	*fácil de usar* (ease of use)
		Price	*pagamento* (payment)
	Verb	Subject	*fala de* (talks about)
		Battery	*descarrega* (discharge)
Feature	Attribute	Design	*material* (material)
		Image	*modo noite* (night mode)
	Equivalence	Employee	*equipe* (team)
		Cleanliness	*higiene* (hygiene)
	Is-a	Food	*pratos* (dishes)
		Employee	*recepcionista* (receptionist)
	Part-of	Facilities	*banheiro* (bathroom)
		Attendance	*atendentes* (attendants)
Qualification	Adjective	Design	*elegante* (elegant)
		Installation	*mobiliado* (furnished)
	Equivalence	Installation	*hotel simples* (simple hotel)
		Price	*diárias acessíveis* (affordable rates)
	Nominal form	Design	*beleza* (beauty)
		Speed	*rapidez* (rapidity)
Contextual	Location	Location	*fácil acesso ao centro* (easy access to downtown)
		Location	*consegue ir a pé até o mercado público* (can walk to the public market)
	Related	Installation	*infiltrações* (infiltrations)
		Cleanliness	*cheira a mofo* (musty smell)

4 Results

Analyzing the data from the lexicons of IACs, we noticed that the studied domains are very different, resulting in difficulties for building domain-independent tools for extracting aspects. However, we did find some common terms, mainly for the camera and smartphone domains, as they are both electronic products. Table 4 presents these common IACs. It is worth mentioning the IAC "*barata*", which is a homograph case that may be translated to "cheap" (not expensive) or "cockroach" (the insect) in English, depending on the context where it is used. For instance, in the Table 4, when it is used for camera domain, it refers to the "price" aspect; when it is used to hotel domain, it refers to "cleanliness".

Table 5 shows the percentage of categories and subcategories in each domain. Cameras and smartphones showed some similar tendencies, in particular, the qualification category prevailed, with 47% and 51%, respectively. For books, the most frequent one was "event", with 52%. For hotels, "feature" was the most frequent one.

Table 4. Common identified IACs

IAC	Domain	Aspect	Domain	Aspect
barata (cheap/cockroach)	camera	price	hotel	cleanliness
barato (cheap)	hotel	price	smartphone	price
beleza (beauty)	camera	design	smartphone	design
demora a responder (delay to respond)	camera	velocity	smartphone	velocity
fácil de usar (easy to use)	camera	usability	smartphone	usability
leve (lightweight)	camera	weight	smartphone	weight
leveza (lightness)	camera	weight	smartphone	weight
versátil (versatile)	camera	design	smartphone	design

Table 5. Distribution of categories and subcategories of IACs in the domains

Category	Subcategory	Books	Cameras	Smartphones	Hotels
Event	Non-verbal form	0%	11%	4%	1%
	Verb	**52%**	17%	26%	1%
Feature	Attribute	0%	15%	3%	0%
	Equivalence	20%	2%	0%	19%
	Is-a	4%	0%	1%	13%
	Part-of	0%	0%	0%	**30%**
Qualification	Adjective	0%	**47%**	**51%**	11%
	Equivalence	0%	0%	0%	2%
	Nominal form	0%	6%	9%	0%
Contextual	Local	0%	0%	0%	9%
	Related	24%	2%	4%	14%

5 Final remarks

This paper presented a deep study of implicit aspects and how to categorize them according to the necessary inferences for their identification. We hope that such characterization, together with the IAC lexicons and the annotated corpora, may foster the development of better sentiment analysis applications for Portuguese.

It is important to notice that the categories that we propose have some overlapping characteristics. For instance, it is not only the "contextual" category that makes use of "world knowledge", but some predominant characteristic may be found in order to categorize the cases. However, we do acknowledge the fact that, depending on the analysis style, different categorizations might be proposed. The categorization in this paper is one possible proposal among other possibilities.

The interested reader may find more information at the web portal of the POeTiSA project (*POrtuguese processing - Towards Syntactic Analysis and parsing*) or in our repository at GitHub.

https://sites.google.com/icmc.usp.br/poetisa
https://github.com/mtarcinalli/Implicit-opinion-aspect-clues-in-Portuguese-texts

Acknowledgement

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

References

1. Cruz, I., Gelbukh, A.F., Sidorov, G.: Implicit aspect indicator extraction for aspect based opinion mining. International Journal of Computational Linguistics and Applications **5**(2), 135–152 (2014)
2. Freitas, C., Motta, E., Milidiú, R.L., César, J.: Vampiro Que Brilha... Rá! Desafios Na Anotação De Opinião Em Um Corpus De Resenhas De Livros. In: XI Encontro de Linguística de Corpus (ELC 2012). São Paulo (2012)
3. d. Freitas, L.A., Vieira, R.: Exploring resources for sentiment analysis in portuguese language. In: 2015 Brazilian Conference on Intelligent Systems (BRACIS). pp. 152–156 (Nov 2015)
4. Freitas, L.A.D., Vieira, R.: Ontology-based feature-level sentiment analysis in portuguese reviews. International Journal of Business Information Systems **32**(1), 30–55 (2019)
5. Ganganwar, V., Rajalakshmi, R.: Implicit aspect extraction for sentiment analysis: A survey of recent approaches. Procedia Computer Science **165**, 485–491 (2019)
6. 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.: A large corpus of product reviews in portuguese: Tackling out-of-vocabulary words. In: LREC. pp. 3865–3871 (2014)
7. Honnibal, M., Montani, I., Van Landeghem, S., Boyd, A.: spaCy: Industrial-strength Natural Language Processing in Python (2020)
8. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 168–177. ACM (2004)
9. Liu, B.: Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies **5**(1), 1–167 (2012)
10. Long, C., Zhang, J., Zhu, X.: A review selection approach for accurate feature rating estimation. In: Coling 2010: Posters. pp. 766–774 (2010)
11. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal **5**(4), 1093–1113 (2014)
12. Panchendrarajan, R., Ahamed, N., Murugaiah, B., Sivakumar, P., Ranathunga, S., Pemasiri, A.: Implicit aspect detection in restaurant reviews using cooccurence of words. In: Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 128–136 (2016)
13. Pang, B., Lee, L.: Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. pp. 115–124 (2005)
14. Pavlopoulos, J., Androutsopoulos, I.: 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 (LASM)@ EACL. pp. 44–52 (2014)
15. Pereira, D.A.: A survey of sentiment analysis in the portuguese language. Artificial Intelligence Review pp. 1–29 (2020)

16. Rana, T.A., Cheah, Y.N.: Aspect extraction in sentiment analysis: comparative analysis and survey. Artificial Intelligence Review **46**(4), 459–483 (2016)
17. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems **89**, 14 – 46 (2015)
18. Snyder, B., Barzilay, R.: Multiple aspect ranking using the good grief algorithm. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference. pp. 300–307 (2007)
19. Taboada, M.: Sentiment analysis: An overview from linguistics. Annual Review of Linguistics **2**, 325–347 (2016)
20. Tsytsarau, M., Palpanas, T.: Survey on mining subjective data on the web. Data Mining and Knowledge Discovery **24**(3), 478–514 (2012)
21. Tubishat, M., Idris, N., Abushariah, M.A.: Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges. Information Processing & Management **54**(4), 545–563 (2018)
22. Vargas, F.A., Pardo, T.A.S.: Clustering and hierarchical organization of opinion aspects: a corpus study. In: Proceedings of the 14th Corpus Linguistics Meeting. pp. 342–351 (2017)
23. Vargas, F.A., Pardo, T.A.S.: Aspect clustering methods for sentiment analysis. In: 13th International Conference on the Computational Processing of Portuguese. pp. 365–374. Canela, RS, Brazil (2018)
24. Vargas, F.A., Pardo, T.A.S.: Linguistic rules for fine-grained opinion extraction. In: the Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media. pp. 1–6 (2020)
25. Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: Current state of text sentiment analysis from opinion to emotion mining. ACM Computing Surveys (CSUR) **50**(2), 1–33 (2017)
26. Yu, J., Zha, Z.J., Wang, M., Chua, T.S.: Aspect ranking: identifying important product aspects from online consumer reviews. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies. pp. 1496–1505 (2011)
27. Zhai, Z., Liu, B., Xu, H., Jia, P.: Grouping product features using semi-supervised learning with soft-constraints. In: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). pp. 1272–1280 (2010)
28. Zhai, Z., Liu, B., Xu, H., Jia, P.: Clustering product features for opinion mining. In: Proceedings of the fourth ACM international conference on Web search and data mining. pp. 347–354 (2011)
29. Zhang, Y., Zhu, W.: Extracting implicit features in online customer reviews for opinion mining. In: Proceedings of the 22nd International Conference on World Wide Web. pp. 103–104 (2013)

6 EVALUATING METHODS FOR EXTRACTION OF ASPECT TERMS IN OPINION TEXTS IN PORTUGUESE – THE CHALLENGES OF IMPLICIT ASPECTS

Our third work was the technical report entitled “NILC at ABSAPT 2022: Aspect Extraction for Portuguese” (MACHADO; PARDO, 2022b) included in Appendix A, since it is not an article.

Here we present our fourth paper, “Evaluating Methods for Extraction of Aspect Terms in Opinion Texts in Portuguese – the Challenges of Implicit Aspects” (MACHADO; PARDO, 2022a) which was presented at the 13th Edition of the Language Resources and Evaluation Conference (LREC 2022) (CALZOLARI *et al.*, 2022), which took place between 20 and 25 June 2022. The paper was published in the proceedings of the event[1].

The conference is organized by the European Language Resources Association, and it is the major event on language resources and evaluation for human language technologies. LREC aims to provide an overview of the state-of-the-art, explore new research and development directions and emerging trends, exchange information regarding language resources and their applications, evaluation methodologies and tools, ongoing and planned activities, industrial uses and needs, requirements coming from e-science and e-society, with respect to both scientific and technological issues as well as policy and organizational ones.

In this work, we tested the following methods for aspects extraction: Freq-Baseline (HU; LIU, 2004), Hu & Liu (HU; LIU, 2004), both extended with a Word2Vec (MIKOLOV *et al.*, 2013b) pruning step (PAVLOPOULOS; ANDROUTSOPOULOS, 2014; MACHADO; PARDO; RUIZ, 2017), Multinomial Naive Bayes (MANNING; RAGHAVAN; SCHÜTZE, 2008), Passive Aggressive (CRAMMER *et al.*, 2006), Perceptron (FREUND; SCHAPIRE, 1999), Stochastic Gradient Descent (BOTTOU, 2012), Conditional Random Fields (CRF) (GANDHI; ATTAR, 2020), and BERT pre-trained models (DEVLIN *et al.*, 2019; SOUZA; NOGUEIRA; LOTUFO, 2020). The results were presented, and we carried out a deeper error analysis of it using the presented implicit aspects typology. Below is the complete work, inserted here with the permission of the editor.

[1] http://www.lrec-conf.org/proceedings/lrec2022/index.html

Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3819–3828
Marseille, 20-25 June 2022

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 inspired 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 model[2] pre-trained with texts from Wikipedia in 102 languages (Devlin et al., 2019), and the second is a model[3] 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 HuggingFace[4]. 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), "*localizaçao*" (7), "*estrutura*" (7), "*localizado*" (7), "*sujos*" (7), "*coberto*" (7), "*recepcionistas*" (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*", "*academiar*", "*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 finetuning 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 GitHub[5] 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 cooccurence 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, oppportunities, 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., Felippo, 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.

7 EVALUATING FROM THE CLASSICAL FREQUENCY TO THE LARGE LANGUAGE MODELS-BASED METHODS FOR ASPECT EXTRACTION IN PORTUGUESE AND THEIR PERFORMANCE ON DETECTING TYPOLOGICALLY DIFFERENT IMPLICIT ASPECTS

Our latest work (MACHADO; PARDO, 2023) was submitted to the Natural Language Engineering journal. This is a renowned open access journal that meets the needs of professionals and researchers working in all areas of automatic language processing, whether from the perspective of theoretical or corpus linguistics, translation, lexicography, computer science or engineering. Its goal is to bridge the gap between traditional research in computational linguistics and the implementation of practical applications with potential real-world use.

In this paper, we corrected a limitation of the experiments by improving the evaluation of results through 5-fold validation. In addition to retesting previous methods with some improvements, we included the Panchendrarajan *et al.* (2016) method and the use of LLMs. Once again, in addition to analyzing the results using traditional metrics, we made use of the implicit aspects typology, which this time can present a complete view of the implicit aspects found, due to the use of the 5-fold. Below is the complete work, inserted here with the permission of the editor.

Natural Language Engineering (2019), 1–00
doi:10.1017/xxxxx

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Evaluating from the classical frequency to the large language model-based methods for aspect extraction in Portuguese and their performance on detecting typologically different 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,
São Carlos, Brazil
mateusmachado@usp.br

(Received xx xxx xxx; revised xx xxx xxx; accepted xx xxx xxx)

Abstract
Contributing to the sentiment analysis research field, this paper presents a comprehensive evaluation of aspect extraction methods, comparing methods from different paradigms, as those based on frequency, rules, traditional machine learning, BERT, and some recent large language models (LLMs). In particular, we emphasize the detection of implicit aspects, which have proven to be a challenging issue in the area. Using a typology that codifies the types of implicit aspects and the necessary knowledge to detect them, we carry out an analysis that allows a better understanding of the results, going beyond the traditional evaluation metrics.

Keywords: Opinion Mining; Identification of aspects; Explicit aspects; Implicit aspects

1. Introduction

Sentiment analysis (also named opinion mining) is a research area that acquired great importance with the increase in data production caused by the expansion of the internet. It is located at the intersection between computing and linguistics (Taboada 2016), which aims to analyze texts and to identify opinions, sentiments, assessments, attitudes, and emotions concerning products, services, organizations, individuals, issues, topics, and their attributes (Liu et al. 2010).

At a more refined level of sentiment analysis, we find the aspect-based sentiment analysis (ABSA), which seeks to identify the sentiment (generally positive, negative, or neutral) and the aspect to which this sentiment is related. For example, in the sentence "The phone gets the signal well", we have a positive opinion about the aspect signal of a telephone entity. It is unnecessary to stress that this kind of information is very attractive for several natural language processing (NLP) applications, which include commercial interests in assessing user satisfaction with new products to political interests on possible election results, and people's opinions about specific public investments.

There are also more challenging tasks. For instance, in the sentence "I was able to make calls even on the riverbank", there is the same positive opinion about the signal, but mentioned implicitly, which increases the difficulty in detecting it, as it is necessary to have the background

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

knowledge that it is usual that cell phones may not have signal in some far locations. One study (Zhang and Zhu 2013) indicates that 30% of the aspects of a corpus are mentioned implicitly. Another study (Panchendrarajan et al. 2016) emphasizes that, in the domain of restaurants, 15.6% of the aspects are implicit, and, for some categories, this percentage is higher, for example, more than 92% of employee-related aspects are mentioned implicitly. Therefore, properly dealing with implicit aspects is a relevant task if one aims to more systematically and confidently identify and extract opinion aspects.

This work focuses on detecting aspects in opinionated texts, with special interest in implicit aspects. For this, we test a wide range of methods, including more traditional and simple methods, classical machine learning-based methods, hybrid methods, pre-trained models, and recent large language models. These last two approaches gave a great boost in the results of several NLP tasks, and it is interesting to contrast their performance with those of other methods, showing their limitations and potentialities. An additional innovation of this work is the analysis of the results using a typology of implicit aspects, which relates the aspect to the knowledge necessary for the interpretation of the expression as an aspect. This analysis allows a deeper understanding of the results and the difficulties in detecting implicit aspects. In particular, we run our experiments for the Brazilian Portuguese language (for which the typology of implicit aspects was built), and use benchmark datasets available for this language.

The rest of the work is organized as follows. Section 2 briefly introduces the main related work. Section 3 presents the data and methods used in this paper. Section 4 shows and discuss the obtained results, while Section 5 brings some final remarks.

2. Related work

We can classify methods for extracting aspects into four categories (Tubishat et al. 2018): supervised, semi-supervised, unsupervised, and hybrid. Unsupervised methods have the advantage of not requiring manually annotated data. We can cite as examples the frequency-based methods, dependency analysis, association rule mining, rule-based extraction, and clustering. On the other hand, supervised methods are those that require labeled data to extract aspects. Examples include association rule mining, fuzzy mining, rule-based mining, and traditional machine learning algorithms.

The semi-supervised methods used both types of data labeled and unlabeled. Some methods of this type are those based on mining association rules, clustering, and topic modeling. Finally, hybrid approaches combine one or more methods of different types, such as a pipeline using rules and then applying some machine learning algorithm. We briefly bring below some works that deal directly with the issue of implicit aspects and others that represent the state-of-the-art in the task of aspect extraction.

Panchendrarajan et al. (2016) presented a method for detecting aspects that overcame some issues usually observed in the area: the limitation on the number of implicit aspects per sentence and the ambiguity when an opinion word is used to describe different aspects. To identify aspects, the authors created two models. The first model uses the maximum entropy classification technique to identify explicit aspects. The second model identifies opinion words (adjectives that refer to aspects) in a sentence and predicts the implicit aspect related to that opinion word using word co-occurrence. The approach yields an F-measure of 84.2% when applied to a set of 1,000 restaurant reviews collected from Yelp.

Xu et al. (2019) post-trained a BERT model for a question-answering task using a dataset composed of opinionated texts. Inspired by the concept of machine reading comprehension in formal documents, they named the problem review reading comprehensions (RRC). The main objective of the trained model is to answer questions, but, to demonstrate the generality of the method, the authors performed tests on other tasks such as aspect extraction. In the 2014 SemEval (Pontiki et al. 2014) dataset, in the domain of laptops, the method achieved an F-measure of 84.26%, while for the 2016 (Pontiki et al. 2016) restaurants domain, it had an f-measure of 77.97%.

Cai et al. (2020) introduced a new task called Aspect-Category-Opinion-Sentiment Quadruple Extraction, which aims to extract these four pieces of information from a sentence, allowing the identification of implicit aspects and sentiments. The task resolution was formulated as a problem of hierarchical prediction of categories and sentiments, where initially the categories of aspects are identified, and then the related sentiments are detected using a convolutional network of hierarchical graphs with the pre-trained BERT language model. In the domain of restaurants, the authors achieved F-measures of 64% and 74%, and for laptops 54% and 62%, for data from 2014 (Pontiki et al. 2014) and 2015 evaluations (Pontiki et al. 2015), respectively.

Wang et al. (2021) presented a method to automatically identify the best combination of concatenated embeddings to be used in prediction tasks. For this, they developed a controller that selects the embeddings and uses reinforcement learning concepts in order to optimize its parameters based on the accuracy of the model. The method showed very good results, reaching an F-measure of 83.9% and 74.9% in the domain of laptops and, in the domain of restaurants, 74.9% and 75.6%. These results were obtained in the datasets of SemEval for the years 2014 (Pontiki et al. 2014) and 2015 (Pontiki et al. 2015), respectively.

Costa and Pardo (2020) presented two unsupervised methods. The first uses an ontology containing terms related to the analyzed domain (reviews in Portuguese of electronic devices (Hartmann et al. 2014)) and the second through the creation of a lexicon from the corpus, analyzing the most frequent nouns and searching for related ones using the Word2Vec model. Finally, the experiment carried out with the ontology obtained the best result, reaching 0.48 in the f-measure.

Aires et al. (2018) used deep learning to identify and classify aspects in small datasets. In a slightly different approach, the authors trained a classifier for each of the ten aspects found in the dataset. In their first experiment, they trained the models using support vector machines and in the second they trained an LSTM network with a single layer containing 100 nodes for 40 epochs. The LSTM network performed better, achieving an accuracy of 99% for most aspects.

In what follows, we introduce the datasets and methods that we explore in this paper.

3. Data and Methods

Given the relevance of the implicit aspects and relative scarcity of works in Portuguese, we implemented and tested several methods with the aim of improving this scenario. We continue describing the data and resources used and the methods implemented.

3.1 Datasets and Resources

All experiments were conducted on data from two datasets with opinion texts in Portuguese and both labeled with explicit and implicit aspect terms. The first collection (Vargas and Pardo 2017) consists of 180 reviews of cameras, books, and smartphones, extracted from the Buscape (Hartmann et al. 2014) and ReLi (Freitas et al. 2012) corpora. The second dataset was formed by combining the corpus of De Freitas and Vieira (2015), which has the explicit terms labeled, with the annotation of the implicit aspect terms performed by Machado et al. (2022). This dataset contains 194 reviews about hotels that were collected from the TripAdvisor website. For the experiments, we selected only sentences containing aspects. Table 3.1 displays the composition of each collection we used, grouped by domains. The last two columns show each domain's explicit and implicit aspects.

Table 1. Statistics of the used datasets

Domain	Reviews	Words	Aspects	Expl.	Impl.
book	60	5,515	330	304	26
camera	60	3,997	352	299	53
smartphone	60	6,210	455	387	68
hotel	194	13,940	1,417	999	415

3.1.1 Typology of Implicit Aspect Clues

To analyze the results more deeply, we used a typology created by Machado et al. (2022). The authors created categories and subcategories according to the type of knowledge needed to relate an Implicit Aspect Clue (IAC) to its respective aspect. The following categories and subcategories were mapped through the analysis of the datasets by the authors:

- **Event (Action/Process/State):** some action, process, or state related to the aspect that allows its identification. It can appear in the form of verbs (subcategory verbs) or as a term that refers to a verb (subcategory non-verbal form), such as a nominal form of a verb;
- **Feature:** IACs of this category are identified by having some relationship with the aspect. It can be an attribute of the aspect or some term or expression that is in an equivalence, is-a, or part-of relationship with the aspect;
- **Qualification:** the IAC is presented as a quality or a sentiment about the aspect, without mentioning it directly. It can be an adjective, a nominal form of an adjective, or a term that is in an equivalence relationship with an adjective;
- **Contextual:** these are the most complicated IACs to identify, as they require knowledge about the context in which the aspect is inserted, such as its operation, mode of use, or content. It was subdivided into related and location, the latter pertaining to the hotel domain.

To facilitate understanding of the typology, Table 3.1.1 summarizes categories and subcategories, presenting examples of each one. Table 3.1.1 presents the number of aspects found in each domain. We can notice that similar products such as cameras and smartphones have similar distributions of implicit aspects for each category. In the domain of books, which has a smaller number of implicit aspects, these are concentrated in verbs. For hotels, we have many related to their location (contextual) and structure (feature).

3.1.2 Nominal forms of verbs and adjectives

Based on other studies (Machado and Pardo 2022), we verified that nominal forms of verbs and adjectives are very often related to the implicit aspects. To improve the detection of this type of aspect, we created two content lexicons derived from verbs or adjectives. The words present in the lexicons were taken from the dictionaries of quality names [a] and the dictionary of deverbal names [b] of the Portuguese Language Portal [c]. Table 3.1.2 presents examples and quantities of each category found in the lexicons of nominal forms.

As demonstrated in the studies that resulted in the typology of implicit aspects, adjectives, verbs, and their respective nominal forms, are often found as indicators of implicit aspects. The aim is to use this implicit aspect formation characteristic to improve its detection.

[a] http://www.portaldalinguaportuguesa.org/recursos.html?action=derdict&type=s0a
[b] http://www.portaldalinguaportuguesa.org/main.html?action=derdict
[c] http://www.portaldalinguaportuguesa.org/

Table 2. Typology of implicit aspects

Category	Subcategory	Aspect	Example
Event	Non-verbal form (NV)	Price	*pagamento* (payment)
	Verb	Meal	*comer* (to eat)
Feature	Attribute (Att)	Design	*material* (material)
	Equivalence (Eq)	Employee	*equipe* (team)
	Is-a	Employee	*recepcionista* (receptionist)
	Part-of (Part)	Facilities	*banheiro* (bathroom)
Qualification	Adjective (Adj)	Installation	*mobiliado* (furnished)
	Equivalence (Eq)	Installation	*hotel simples* (simple hotel)
	Nominal form (NF)	Location	*proximidade* (proximity)
Contextual	Location (Loc)	Location	*fácil acesso ao centro* (easy access to downtown)
	Related (Rel)	Cleanliness	*cheira a mofo* (musty smell)

Table 3. Statistics of typology of implicit aspect clues

	Event		Feature				Qualification			Contextual	
Domain	NV	Verb	Att	Eq	Is-a	Part	Adj	Eq	NF	Loc	Rel
book	0	13	0	5	1	0	0	0	0	0	6
camera	6	9	8	1	0	0	25	1	3	0	1
hotel	3	5	0	78	53	124	47	10	1	36	58
smartphone	3	18	2	0	1	0	35	0	6	0	3

Table 4. Nominal forms of verbs and adjectives.

Derived	Quantity	Examples	Quantity	Nominal Form
Verb	2,902	facilitar (to facilitate) usar (to use) ler (to read)	3,926	facilitação (facilitation) uso (use), usagem (usage) leitura (reading)
Adjective	1,560	acessível (accessible) confiável (confiável) útil (useful)	1,606	acessibilidade (accessibility) confiabilidade (reliability) utilidade (utility)

3.2 Methods for Aspect Extraction

Using the previous data, we may implement, improve, test, validate, and analyze the results of methods for aspect extraction.

We implemented and tested both supervised and unsupervised methods. To allow a fair comparison between the methods, we used the same datasets in all experiments, which were converted to the traditional IOB format (short for inside, outside, beginning), with the "B-ASP" for begin of aspect, "I-ASP" for inside of aspect, and "O" for outside of aspect. In the search for the best parameters of the evaluated methods, we divided the data into a training set (with 70% of the texts) and a test set (with the remaining 30%). For the calculation of the evaluation metrics, we divided the data into 5 folds, thus allowing the use of cross-validation.

The main features of the implemented methods are presented in what follows. Due to space limitations, more details can be obtained from the mentioned references or from our code repository.

3.2.1 Frequency-based Baseline

The Frequency-based baseline (for simplicity, Freq-Baseline, from now on) is a frequency method that is based on the fact that most aspects are found in texts in the form of nouns or noun phrases. The method identifies the most frequent nouns and noun phrases and classifies them as aspects using a cutoff frequency. In (Hu and Liu 2004), the authors used a cutoff frequency of 1%. In our work, as well as in the work of Machado et al. (2017), we tested a range of frequencies in search of the best one. This search was performed by applying the method to the total set of texts and validating it on the test set. The best cutoff frequency found was used with the 5 folds cross-validation sets.

In a second experiment, we used the Word2Vec (Mikolov et al. 2013) distributional model[d] to verify and exclude candidate terms for aspects that were not related to the context of the data set (Pavlopoulos and Androutsopoulos 2014; Machado et al. 2017). For this, the Word2Vec vector of each candidate was compared with a common context vector, formed by the most frequent words of a literary corpus, and with another formed by the most frequent words of the analyzed dataset. Terms closer to the common vector than to the contextualized one were excluded from the list of aspects.

Studies (Machado et al. 2022) have indicated that implicit aspects often appear in the form of verbs, adjectives, or words derived from them. In order to improve the detection of implicit aspects, in addition to searching for the most frequent nouns and noun phrases, we also classified adjectives, verbs, and derivatives, identified using the aforementioned lexicons, as aspect terms. As we did with nouns, we also sought the best cutoff frequencies for adjectives and verbs.

3.2.2 Hu & Liu Method

Hu and Liu (2004) developed a hybrid, frequency and rule-based method for identifying aspects. The method has a first step similar to Freq-Baseline that identifies the most frequent nouns and noun phrases as candidate aspects. A first pruning mechanism eliminates less frequent composite candidates by comparing the frequency of the composite form with the frequency of each component. In a second pruning step, the algorithm analyzes the distance between each component of a composite aspect, eliminating those that are too far apart. Finally, the algorithm seeks to identify infrequent aspects, through the proximity of nouns of the opinion words, which are adjectives located close to an already identified aspect, including these nouns in the list of candidate aspects.

As in the Freq-Baseline method and as it was done in other studies (Pavlopoulos and Androutsopoulos 2014; Machado et al. 2017), we also used the Word2Vec model to eliminate aspects not related to the context under study. In the first experiment, we performed this pruning after the frequency step and in a second experiment after identifying infrequent aspects.

3.2.3 Panchendrajaran et al. Method

We also tested the method presented by Panchendrarajan et al. (2016), as it is one of the few to deal with the detection of implicit aspects. As mentioned in Section 2, the method uses two models, one using a maximum entropy classifier for the explicit aspects and another using word co-occurrence. As a result of the first model, two lists are created: one formed by candidate aspects and another

[d]Using static vectors from Word2Vec may appear to be an already old approach, but in fact they fit well for the purpose intended here.

containing the opinion words related to these candidate aspects. Selecting aspects from the list of candidate aspects is a two-step process:

- **Step 1:** an implicit aspect from the list of candidate aspect is chosen as a potential aspect candidate, using co-occurrence between the opinion word and other words in the sentence. If there is only one candidate aspect, it will be chosen as a potential candidate. Otherwise, for each candidate aspect, a score is calculated, and the higher one is selected.
- **Step 2:** from a potential candidate, its opinion target is extracted to verify if the prediction is correct. These targets are extracted using the double propagation approach proposed by Qiu et al. (2011), which propagates information between opinion words and targets using grammar rules, based on the dependency relationships between the words. If the opinion target identified is related to the potential candidate, it will be chosen as the winning candidate; otherwise, it will be discarded.

In this way, the method can select, filter, and validate the aspects found.

3.2.4 Traditional Machine Learning Methods

For supervised analysis, we chose some traditional machine learning algorithms in order to broaden the basis for comparison between the methods. In these experiments, we performed a grid search of the best parameters for each algorithm, using the training and testing sets mentioned above. To calculate the metrics, we used the same 5 folds used previously, calculating the mean and standard deviation of each experiment. For the experiments, we chose the following algorithms:

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

For the algorithms Multinomial Naive Bayes, Stochastic Gradient Descent, Perceptron, and Passive Aggressive, we used the bag-of-words model for features. For the Conditional Random Fields experiment, we used, in addition to the words, their neighbors and part of speech tags obtained by the pt_core_news_lg model[e] from the spaCy (Honnibal et al. 2020) module.

3.2.5 BERT-based Methods

Pre-trained models have provided great improvements in the results of several NLP tasks. These models are initially trained for simpler tasks, such as discovering the next word or sentence of a text. This type of task does not require an annotated corpus, allowing the use of large volumes of texts in training, and the created model allows a better understanding of the context in which each word appears by analyzing the text to the left and right of each word simultaneously. With some

[e]https://spacy.io/models/pt

examples and using a process called fine-tuning, it is possible to adjust this model for specific tasks such as named-entity recognition, polarity detection, or aspect extraction, which is our case.

In this work, we used two pre-trained models, a multilingual one [f] trained with Wikipedia texts in 102 languages (Devlin et al. 2019), and a second one (Souza et al. 2020) trained with texts only in Portuguese obtained from the brWaC corpus (Wagner Filho et al. 2018). For these experiments, we removed 20% of the texts from the training set to form a validation set, necessary for the fine-tuning process.

After performing some tests, we achieved satisfactory results with the following parameters: 40 epochs, a learning rate of 0.0001, and a batch size of 13. For each pre-trained model, we performed a fine-tuning for each domain individually and another one with data from all the domains.

3.2.6 Large Language Models-based Methods

LLMs enabled a new way of solving NLP tasks. Through the use of prompts, and guidance in natural language, it is possible to solve problems in an unsupervised way (zero-shot), or with a few examples (few-shot) (Brown et al. 2020). The so-called prompt engineering aims to find the most appropriate prompt that will allow a LLM to solve a given task (Liu et al. 2023).

In the experiments, we used the zero-shot and few-shot approaches, varying the number of examples and the format of the prompt sent to the LLM, as shown in Table 3.2.6. In the zero-shot, we bring a brief explanation of the problem and request the aspects of the sentence. In the first few-shot approach tested, we present some examples and then request the aspects of the sentence. In a second approach (Wang et al. 2023), we marked the aspects in the sentence using tags (@@ and ##), requesting that the LLM return the sentence with the same marking, this way limiting the responses to the content of the presented sentence.

Table 5. Examples of prompts.

zero-shot:
The sentence below is an example of opinion text for detecting aspects. What are its aspects?
Text: the price is compatible with the unbeatable quality of a Sony.
Aspects:
Few-shot:
Some examples of detecting aspects in evaluative texts:
Text: I didn't miss the resources it doesn't have.
Aspects: resources
...
Text: price is compatible with the unbeatable quality of a Sony.
Aspects:
Few-shot with special tags:
Some examples of detecting aspects in evaluative texts:
Input: I didn't miss the resources it doesn't have.
Output: I didn't miss the @@resources## it doesn't have.
...
Input: price is compatible with the unbeatable quality of a Sony.
Output:

[f] https://huggingface.co/bert-base-multilingual-uncased

To assemble the prompt, examples of phrases and aspects were chosen at random within the set of training folds. The number of examples was limited to the number of characters allowed by the model in use or experiment. In some cases, the LLM did not return an adequate response to the prompt or returned words that were not in the text as aspects, such as a noun that appears in the text in the plural, but the model returns its singular form. Due to the fact that we chose only sentences containing aspects for the experiments, to work around this problem, when the answer does not present any aspect present in the text, we present the prompt to the LLM again until we obtain a valid answer, repeating this process for a maximum of 10 times.

In the experiments, we used the following LLMs:

- GPT 3.5[g]: the experiments were carried out through the ChatGPT interface, using the davinci-002 model with prompts of up to 4,000 characters;
- Maritaca[h]: the model was trained using texts in Portuguese. The experiments were performed through its API with prompts of up to 2,000 characters;
- Llama 2[i]: due to the availability of this model by Meta, it was possible to conduct a greater number of experiments. We worked with the model of 7 billion parameters, with which we ran zero-shot and few-shot experiments with prompts of up to 2,000 and 4,000 characters, and aspects tagging. With the model of 13 billion parameters, we tested zero-shot and few-shot with up to 4,000 characters.

4. Results

Freq-Baseline is a simple method to implement and is commonly used as a basis for comparison with other methods. The use of adjectives, verbs, and their respective nominal forms slightly improved the detection of implicit aspects, but it was not enough to improve the f-measure, which found its best results in experiments with the Word2Vec pruning mechanism.

Table 4 presents the cutoff of each experiment carried out and the arithmetic means and standard deviation (in parentheses) of the 5-fold results for the precision, recall, f-measure, and implicit metrics. The precision, recall, and f-measure metrics are valid for all aspects, given that the classifiers do not separately detect implicit and explicit aspects. The implicit column presents the percentage of implicit aspects that each method identified.

The hybrid, frequency, and rule-based method, created by Hu & Liu, presented results close to those obtained by Freq-Baseline and a little below the Freq-Baseline with Word2Vec pruning. Despite its greater complexity, and better ability to detect composite and infrequent aspects, the method did not do well, as we can see in the Table 4.

Among the traditional methods of machine learning, the CRF stood out, which was already expected, since it, in addition to using other grammatical characteristics, also takes into account neighboring words. Only in the domain of books, its results were one point below the Perceptron, in addition to not having identified any implicit aspect, as shown in Table 4.

Another hybrid method was the one presented by Panchendrarajan et al. (2016), which combines machine learning and rules, with the aim of improving the detection of implicit aspects, an objective achieved as we can see in Table 4, given that the method was able to detect a significant amount of implicit aspects in all the analyzed domains.

In the experiments using the BERT pre-trained model, the results were close to those achieved with CRF, and, in general, BERTimbau performed better or very close to the multilingual model. The models also obtained positive results in the detection of implicit aspects, as shown in Table 4.

[g] https://chat.openai.com/
[h] https://chat.maritaca.ai/
[i] https://ai.meta.com/llama/

Table 6. Freq-Baseline.

Domain	Algorithm	Cutoff	Precision	Recall	F-measure	Implicits
book	Freq	0.04	**0.79 (0.04)**	0.55 (0.09)	0.64 (0.07)	0.00 (0.00)
book	Freq w2v	0.04	0.75 (0.04)	**0.74 (0.10)**	**0.74 (0.06)**	**0.09 (0.12)**
book	Freq adj	0.04	**0.79 (0.04)**	0.55 (0.09)	0.64 (0.07)	0.00 (0.00)
book	Freq adj nom	0.04	**0.79 (0.04)**	0.55 (0.09)	0.64 (0.07)	0.00 (0.00)
book	Freq verb	0.04	0.72 (0.05)	0.56 (0.10)	0.63 (0.08)	0.08 (0.11)
book	Freq verb nom	0.04	0.72 (0.05)	0.56 (0.10)	0.63 (0.08)	0.08 (0.11)
book	Freq adj verb	0.04	0.72 (0.05)	0.56 (0.10)	0.63 (0.08)	0.08 (0.11)
book	Freq adj nom verb	0.04	0.72 (0.05)	0.56 (0.10)	0.63 (0.08)	0.08 (0.11)
book	Freq adj nom verb w2v	0.03	0.76 (0.05)	0.57 (0.09)	0.65 (0.07)	0.08 (0.11)
camera	Freq	0.02	0.68 (0.05)	0.60 (0.04)	0.63 (0.04)	0.11 (0.08)
camera	Freq w2v	0.02	**0.78 (0.05)**	0.56 (0.03)	**0.65 (0.03)**	0.11 (0.08)
camera	Freq adj	0.02	0.63 (0.05)	**0.62 (0.04)**	0.62 (0.04)	**0.24 (0.06)**
camera	Freq adj nom	0.02	0.63 (0.05)	**0.62 (0.04)**	0.62 (0.04)	**0.24 (0.06)**
camera	Freq verb	0.03	0.68 (0.04)	0.54 (0.04)	0.60 (0.04)	0.11 (0.08)
camera	Freq verb nom	0.03	0.68 (0.04)	0.54 (0.04)	0.60 (0.04)	0.11 (0.08)
camera	Freq adj verb	0.03	0.64 (0.04)	0.56 (0.05)	0.60 (0.03)	**0.24 (0.06)**
camera	Freq adj nom verb	0.03	0.64 (0.04)	0.56 (0.05)	0.60 (0.03)	**0.24 (0.06)**
camera	Freq adj nom verb w2v	0.02	0.58 (0.05)	0.58 (0.03)	0.58 (0.03)	**0.24 (0.06)**
hotel	Freq	0.01	0.70 (0.03)	0.67 (0.02)	0.69 (0.02)	0.13 (0.02)
hotel	Freq w2v	0.01	**0.84 (0.02)**	0.65 (0.02)	**0.73 (0.02)**	0.10 (0.02)
hotel	Freq adj	0.01	0.68 (0.02)	0.73 (0.02)	0.70 (0.01)	0.19 (0.02)
hotel	Freq adj nom	0.01	0.68 (0.02)	0.73 (0.02)	0.70 (0.01)	0.19 (0.02)
hotel	Freq verb	0.01	0.65 (0.02)	0.69 (0.02)	0.67 (0.01)	0.15 (0.02)
hotel	Freq verb nom	0.01	0.65 (0.02)	0.69 (0.02)	0.67 (0.01)	0.15 (0.02)
hotel	Freq adj verbs	0.01	0.64 (0.02)	**0.74 (0.02)**	0.69 (0.01)	**0.21 (0.02)**
hotel	Freq adjs nom verbs	0.01	0.63 (0.02)	**0.74 (0.02)**	0.68 (0.01)	**0.21 (0.02)**
hotel	Freq adj nom verbs w2v	0.01	0.72 (0.02)	0.72 (0.02)	0.72 (0.01)	0.18 (0.01)
smartphone	Freq	0.02	0.51 (0.04)	0.54 (0.05)	0.52 (0.04)	0.02 (0.03)
smartphone	Freq w2v	0.04	0.64 (0.03)	0.52 (0.05)	**0.58 (0.04)**	0.02 (0.03)
smartphone	Freq adj	0.04	**0.65 (0.05)**	0.41 (0.06)	0.50 (0.06)	0.00 (0.00)
smartphone	Freq adj nom	0.04	**0.65 (0.05)**	0.41 (0.06)	0.50 (0.06)	0.00 (0.00)
smartphone	Freq verb	0.02	0.45 (0.04)	**0.55 (0.05)**	0.50 (0.04)	0.06 (0.04)
smartphone	Freq verb nom	0.04	0.63 (0.04)	0.42 (0.04)	0.51 (0.05)	0.04 (0.04)
smartphone	Freq adj verb	0.04	0.61 (0.05)	0.41 (0.06)	0.49 (0.05)	0.00 (0.00)
smartphone	Freq adj nom verb	0.04	0.62 (0.04)	0.42 (0.04)	0.50 (0.05)	0.04 (0.04)
smartphone	Freq adj nom verb w2v	0.02	0.50 (0.03)	**0.55 (0.03)**	0.52 (0.03)	**0.16 (0.05)**

Table 7. Hu & Liu method.

Domain	Algorithm	Cutoff	Precision	Recall	F-measure	Implicits
book	Hu Liu	0.03	0.69 (0.04)	**0.60 (0.10)**	0.64 (0.06)	0.00 (0.00)
book	Hu Liu w2v	0.03	0.71 (0.04)	**0.60 (0.10)**	0.65 (0.06)	0.00 (0.00)
book	Hu Liu infw2v	0.03	**0.73 (0.04)**	**0.60 (0.10)**	**0.66 (0.07)**	0.00 (0.00)
camera	Hu Liu	0.03	0.76 (0.04)	**0.50 (0.03)**	0.60 (0.03)	**0.11 (0.08)**
camera	Hu Liu w2v	0.03	0.76 (0.04)	**0.50 (0.03)**	0.60 (0.03)	**0.11 (0.08)**
camera	Hu Liu infw2v	0.03	**0.77 (0.04)**	**0.50 (0.03)**	**0.61 (0.04)**	**0.11 (0.08)**
hotel	Hu Liu	0.01	0.72 (0.03)	**0.67 (0.02)**	**0.70 (0.02)**	**0.13 (0.02)**
hotel	Hu Liu w2v	0.01	**0.73 (0.03)**	**0.67 (0.02)**	**0.70 (0.02)**	**0.13 (0.02)**
hotel	Hu Liu infw2v	0.01	**0.73 (0.02)**	**0.67 (0.03)**	**0.70 (0.02)**	**0.13 (0.02)**
smartphone	Hu Liu	0.02	0.56 (0.04)	0.52 (0.06)	0.54 (0.05)	**0.10 (0.04)**
smartphone	Hu Liu w2v	0.02	**0.57 (0.03)**	0.52 (0.06)	0.54 (0.04)	**0.10 (0.04)**
smartphone	Hu Liu infw2v	0.01	0.53 (0.03)	**0.58 (0.05)**	**0.55 (0.03)**	0.02 (0.03)

Table 8. Machine Learning-based methods.

Domain	Algorithm	Precision	Recall	F-measure	Implicits
book	CRF	0.92 (0.04)	**0.75 (0.07)**	0.82 (0.04)	0.00 (0.00)
book	MNB	**0.94 (0.04)**	0.66 (0.07)	0.78 (0.05)	0.00 (0.00)
book	PA	**0.94 (0.01)**	0.73 (0.06)	0.82 (0.03)	**0.05 (0.11)**
book	Perceptron	0.93 (0.03)	**0.75 (0.09)**	**0.83 (0.06)**	0.04 (0.09)
book	SGD	**0.94 (0.02)**	0.73 (0.07)	0.82 (0.04)	0.04 (0.09)
camera	CRF	**0.80 (0.04)**	**0.68 (0.06)**	**0.74 (0.05)**	**0.30 (0.10)**
camera	MNB	0.38 (0.04)	0.61 (0.07)	0.47 (0.05)	0.27 (0.13)
camera	PA	0.74 (0.12)	0.56 (0.08)	0.63 (0.10)	0.26 (0.10)
camera	Perceptron	0.72 (0.15)	0.63 (0.05)	0.66 (0.06)	0.29 (0.16)
camera	SGD	0.73 (0.13)	0.49 (0.10)	0.58 (0.11)	0.22 (0.11)
hotel	CRF	**0.95 (0.01)**	**0.86 (0.02)**	**0.90 (0.01)**	**0.31 (0.02)**
hotel	MNB	0.55 (0.03)	0.85 (0.02)	0.67 (0.02)	0.30 (0.04)
hotel	PA	0.92 (0.01)	0.83 (0.02)	0.87 (0.02)	0.27 (0.03)
hotel	Perceptron	0.85 (0.18)	0.67 (0.16)	0.73 (0.10)	0.14 (0.14)
hotel	SGD	0.92 (0.01)	0.84 (0.02)	0.87 (0.01)	0.27 (0.03)
smartphone	CRF	**0.81 (0.03)**	0.69 (0.04)	0.74 (0.03)	0.35 (0.11)
smartphone	MNB	0.44 (0.03)	**0.75 (0.07)**	0.55 (0.04)	**0.38 (0.13)**
smartphone	PA	**0.81 (0.03)**	0.69 (0.03)	**0.75 (0.02)**	0.33 (0.10)
smartphone	Perceptron	0.80 (0.04)	0.66 (0.08)	0.72 (0.05)	0.25 (0.05)
smartphone	SGD	0.80 (0.03)	0.67 (0.04)	0.73 (0.04)	0.34 (0.09)

Finally, Table 4 presents the results of the experiments using the LLMs, where 7B and 13B represent the models with 7 and 13 billion parameters respectively, 2k indicates the experiment with a prompt of 2,000 words, and tag indicates the experiment carried out with the special tags

Table 9. Method of Panchendrajaran et al.

Domain	Algorithm	Precision	Recall	F-measure	Implicits
book	Panchendrarajan et al.	0.92 (0.02)	0.76 (0.09)	0.83 (0.06)	0.17 (0.17)
camera	Panchendrarajan et al.	0.47 (0.06)	0.54 (0.04)	0.50 (0.05)	0.32 (0.11)
hotel	Panchendrarajan et al.	0.91 (0.01)	0.87 (0.01)	0.89 (0.01)	0.34 (0.02)
smartphone	Panchendrarajan et al.	0.78 (0.03)	0.69 (0.04)	0.73 (0.03)	0.37 (0.08)

Table 10. BERT-based methods.

Domain	Algorithm	Precision	Recall	F-measure	Implicits
book	BERT in-domain	0.84 (0.07)	0.80 (0.12)	0.82 (0.09)	0.08 (0.11)
book	BERT cross-domain	0.85 (0.04)	0.83 (0.04)	0.84 (0.04)	**0.17 (0.21)**
book	BERTimbau in-domain	**0.86 (0.04)**	**0.87 (0.06)**	**0.87 (0.04)**	**0.17 (0.17)**
book	BERTimbau cross-domain	**0.86 (0.05)**	0.83 (0.04)	0.84 (0.02)	0.03 (0.06)
camera	BERT in-domain	0.63 (0.09)	0.57 (0.04)	0.60 (0.06)	0.24 (0.06)
camera	BERT cross-domain	**0.66 (0.06)**	0.59 (0.07)	**0.62 (0.06)**	0.26 (0.04)
camera	BERTimbau in-domain	0.63 (0.06)	**0.62 (0.07)**	**0.62 (0.05)**	0.24 (0.16)
camera	BERTimbau cross-domain	**0.66 (0.05)**	0.59 (0.07)	**0.62 (0.05)**	**0.32 (0.14)**
hotel	BERT in-domain	**0.86 (0.02)**	0.87 (0.01)	**0.87 (0.01)**	0.36 (0.03)
hotel	BERT cross-domain	**0.86 (0.02)**	0.88 (0.02)	**0.87 (0.01)**	0.37 (0.05)
hotel	BERTimbau in-domain	0.85 (0.02)	**0.89 (0.02)**	**0.87 (0.01)**	0.38 (0.05)
hotel	BERTimbau cross-domain	0.85 (0.02)	**0.89 (0.02)**	**0.87 (0.02)**	**0.39 (0.05)**
smartphone	BERT in-domain	0.67 (0.05)	**0.71 (0.05)**	**0.69 (0.05)**	0.30 (0.08)
smartphone	BERT cross-domain	**0.70 (0.03)**	0.66 (0.06)	0.68 (0.04)	0.24 (0.05)
smartphone	BERTimbau in-domain	0.65 (0.03)	0.66 (0.06)	0.66 (0.04)	**0.33 (0.07)**
smartphone	BERTimbau cross-domain	0.69 (0.03)	0.67 (0.05)	0.68 (0.04)	0.27 (0.04)

for aspects. The LLMs did not achieve results close to the CRF in terms of f-measure, but, on the other hand, they performed well in detecting implicit aspects.

Regarding the Freq-Baseline, we can observe that the inclusion of adjectives, verbs, and their nominal forms had some positive effect on the results, but overall below the LLMs that obtained the best results in relation to the implicit ones.

The CRF that obtained the best results, in terms of f-measure, failed to detect some types of aspects. The same happened with the Panchendrarajan et al. and BERT-based methods, although both performed better than the CRF.

The highlight is for the methods based on LLMs, reaching 100% in some categories and detecting aspects in categories that none of the other methods could. A very important point when analyzing the results of the LLMs is that they were ”trained” with a much smaller number of examples when compared to the other methods.

Table 4 summarizes the results found. We selected the experiments that achieved the best f-measures or were highlighted in the detection of implicit aspects. From the data presented, we can observe that the CRF performed very well compared to other methods, not achieving the best result only in the domain of books, where the method could not detect any implicit aspect. Regarding implicit aspects, the methods using LLMs deserve to be highlighted, despite not achieving such high f-measures, they found significant amounts of implicit aspects.

Table 11. LLM-based methods.

Domain	Model	Precision	Recall	F-measure	Implicits
book	GPT-3.5	0.28 (0.12)	0.65 (0.07)	0.38 (0.08)	**0.33 (0.16)**
book	Maritaca	0.41 (0.16)	**0.69 (0.08)**	0.50 (0.09)	0.20 (0.25)
book	Llama 7B zero-shot	0.34 (0.12)	0.59 (0.07)	0.42 (0.08)	0.14 (0.22)
book	Llama 7B 2k	0.47 (0.14)	0.65 (0.13)	0.52 (0.05)	0.12 (0.11)
book	Llama 7B tags	**0.79 (0.16)**	0.62 (0.12)	**0.68 (0.06)**	0.08 (0.18)
book	Llama 7B	0.69 (0.24)	0.63 (0.10)	0.63 (0.08)	0.25 (0.20)
book	Llama 13B zero-shot	0.40 (0.15)	0.55 (0.09)	0.44 (0.08)	0.18 (0.25)
book	Llama 13B	0.60 (0.14)	0.59 (0.11)	0.59 (0.09)	0.29 (0.25)
camera	GPT-3.5	0.39 (0.10)	0.45 (0.12)	0.40 (0.07)	0.42 (0.16)
camera	Maritaca	0.55 (0.11)	0.60 (0.15)	0.55 (0.09)	0.36 (0.10)
camera	Llama 7B zero-shot	0.48 (0.21)	0.49 (0.10)	0.45 (0.03)	0.43 (0.09)
camera	Llama 7B 2k	0.54 (0.13)	**0.63 (0.15)**	0.56 (0.08)	0.40 (0.15)
camera	Llama 7B tags	**0.77 (0.20)**	0.50 (0.06)	0.59 (0.03)	0.26 (0.18)
camera	Llama 7B	0.72 (0.25)	0.57 (0.11)	**0.60 (0.04)**	**0.52 (0.19)**
camera	Llama 13B zero-shot	0.42 (0.17)	0.49 (0.11)	0.43 (0.07)	0.45 (0.19)
camera	Llama 13B	0.54 (0.12)	0.44 (0.13)	0.47 (0.08)	0.26 (0.04)
hotel	GPT-3.5	0.40 (0.13)	0.62 (0.04)	0.48 (0.08)	0.26 (0.02)
hotel	Maritaca	0.56 (0.12)	**0.80 (0.05)**	0.65 (0.06)	**0.30 (0.05)**
hotel	Llama 7B zero-shot	0.54 (0.13)	0.75 (0.06)	0.62 (0.06)	0.25 (0.04)
hotel	Llama 7B 2k	0.58 (0.14)	0.76 (0.04)	0.65 (0.08)	0.26 (0.03)
hotel	Llama 7B tags	**0.79 (0.15)**	0.66 (0.05)	**0.72 (0.06)**	0.22 (0.04)
hotel	Llama 7B	0.70 (0.23)	0.75 (0.04)	0.71 (0.10)	0.24 (0.03)
hotel	Llama 13B zero-shot	0.50 (0.13)	0.73 (0.06)	0.59 (0.07)	0.25 (0.04)
hotel	Llama 13B	0.69 (0.09)	0.73 (0.04)	0.70 (0.04)	0.24 (0.04)
smartphone	GPT-3.5	0.41 (0.17)	0.58 (0.08)	0.45 (0.07)	**0.48 (0.12)**
smartphone	Maritaca	0.54 (0.12)	**0.65 (0.10)**	0.58 (0.05)	0.40 (0.22)
smartphone	Llama 7B zero-shot	0.49 (0.14)	0.53 (0.12)	0.49 (0.06)	0.38 (0.15)
smartphone	Llama 7B 2k	0.58 (0.10)	**0.65 (0.14)**	**0.60 (0.07)**	0.32 (0.23)
smartphone	Llama 7B tags	**0.74 (0.17)**	0.46 (0.08)	0.56 (0.07)	0.24 (0.18)
smartphone	Llama 7B	0.69 (0.23)	0.58 (0.11)	**0.60 (0.06)**	0.21 (0.14)
smartphone	Llama 13B zero-shot	0.39 (0.09)	0.50 (0.20)	0.41 (0.08)	0.40 (0.24)
smartphone	Llama 13B	0.60 (0.14)	0.50 (0.09)	0.54 (0.07)	0.17 (0.12)

In addition to the metrics, we carry out other analyses of the results. Table 13 presents some examples of correctly, incorrectly, and not found implicit aspects. Next to each aspect, in parentheses, is the number of times that an algorithm detected (or not) each aspect, those without this number were detected only once. Analyzing the table, we can see that the aspects with just one

Table 12. Best results.

Domain	Algorithm	Precision	Recall	F-measure	Implicits
book	Freq	0.79 (0.04)	0.55 (0.09)	0.64 (0.07)	0.00 (0.00)
book	Freq adj nom verb w2v	0.76 (0.05)	0.57 (0.09)	0.65 (0.07)	0.08 (0.11)
book	CRF	**0.92** (0.04)	0.75 (0.07)	0.82 (0.04)	0.00 (0.00)
book	Panchendrarajan et al.	**0.92** (0.02)	0.76 (0.09)	0.83 (0.06)	0.17 (0.17)
book	BERTimbau in-domain	0.86 (0.04)	**0.87 (0.06)**	**0.87 (0.04)**	0.17 (0.17)
book	GPT-3.5	0.28 (0.12)	0.65 (0.07)	0.38 (0.08)	**0.33 (0.16)**
book	Maritaca	0.41 (0.16)	0.69 (0.08)	0.50 (0.09)	0.20 (0.25)
book	Llama 7B	0.69 (0.24)	0.63 (0.10)	0.63 (0.08)	0.25 (0.20)
camera	Freq	0.68 (0.05)	0.60 (0.04)	0.63 (0.04)	0.11 (0.08)
camera	Freq adj nom verb w2v	0.58 (0.05)	0.58 (0.03)	0.58 (0.03)	0.24 (0.06)
camera	CRF	**0.80 (0.04)**	**0.68 (0.06)**	**0.74 (0.05)**	0.30 (0.10)
camera	Panchendrarajan et al.	0.47 (0.06)	0.54 (0.04)	0.50 (0.05)	0.32 (0.11)
camera	BERTimbau in-domain	0.63 (0.06)	0.62 (0.07)	0.62 (0.05)	0.24 (0.16)
camera	GPT-3.5	0.39 (0.10)	0.45 (0.12)	0.40 (0.07)	0.42 (0.16)
camera	Maritaca	0.55 (0.11)	0.60 (0.15)	0.55 (0.09)	0.36 (0.10)
camera	Llama 7B	0.72 (0.25)	0.57 (0.11)	0.60 (0.04)	**0.52 (0.19)**
hotel	Freq	0.70 (0.03)	0.67 (0.02)	0.69 (0.02)	0.13 (0.02)
hotel	Freq adj nom verbs w2v	0.72 (0.02)	0.72 (0.02)	0.72 (0.01)	0.18 (0.01)
hotel	CRF	**0.95 (0.01)**	0.86 (0.02)	**0.90 (0.01)**	0.31 (0.02)
hotel	Panchendrarajan et al.	0.91 (0.01)	0.87 (0.01)	0.89 (0.01)	0.34 (0.02)
hotel	BERTimbau in-domain	0.85 (0.02)	**0.89 (0.02)**	0.87 (0.01)	**0.38 (0.05)**
hotel	GPT-3.5	0.40 (0.13)	0.62 (0.04)	0.48 (0.08)	0.26 (0.02)
hotel	Maritaca	0.56 (0.12)	0.80 (0.05)	0.65 (0.06)	0.30 (0.05)
hotel	Llama 7B	0.70 (0.23)	0.75 (0.04)	0.71 (0.10)	0.24 (0.03)
smartphone	Freq	0.51 (0.04)	0.54 (0.05)	0.52 (0.04)	0.02 (0.03)
smartphone	Freq adj nom verb w2v	0.50 (0.03)	0.55 (0.03)	0.52 (0.03)	0.16 (0.05)
smartphone	CRF	**0.81 (0.03)**	**0.69 (0.04)**	**0.74 (0.03)**	0.35 (0.11)
smartphone	Panchendrarajan et al.	0.78 (0.03)	**0.69 (0.04)**	0.73 (0.03)	0.37 (0.08)
smartphone	BERTimbau in-domain	0.65 (0.03)	0.66 (0.06)	0.66 (0.04)	0.33 (0.07)
smartphone	GPT-3.5	0.41 (0.17)	0.58 (0.08)	0.45 (0.07)	**0.48 (0.12)**
smartphone	Maritaca	0.54 (0.12)	**0.65 (0.10)**	0.58 (0.05)	0.40 (0.22)
smartphone	Llama 7B	0.69 (0.23)	0.58 (0.11)	0.60 (0.06)	0.21 (0.14)

term were those most identified by the algorithms. Among those that were located by fewer algorithms, we noticed that the majority of them were located by LLMs, demonstrating how these models really ”understand” the texts differently.

Among the incorrect ones, we can see some signs of problems in the manual annotation made. For example, we have the word “*livro*” (book) included as an aspect, but its plural form (“*livros*” books) does not. The name “Saramago” could also be noted as an aspect related to the author.

Other errors indicate problems in the classifier, such as "um livro" (a book) where the article "um" should not have been included in the identified aspect. On smartphones, we have some adjectives, a class that often represents implicit aspects, but in the cases such as "*qualidade*" (quality), "*bom*" (good), and "*problema*" (problem), for example, they alone cannot demonstrate which aspect they refer to.

Among those not found, we have many expressions, proving the greater difficulty in detecting compound terms. We can also consider the possibility of errors in the annotation as in "*fala de*" (speech of), "*pertinho de*" (near of), and "*dos recepcionistas*" (of the receptionists), in these cases, the prepositions should not be included in the manual annotation as they have no semantic value in identifying the aspect. In "*sabe como escrever*" (know how to write), the verb to write alone already refers to the author aspect and could have been tagged as an aspect.

Table 13. Examples of correct, incorrect, and not found implicit aspects

Domain	Correct	Incorrect	Not found
book	"*reflexão*" (9), "*ler*" (7), ..., "*termina*", "*retrata*", "*escreve*"	"*livros*" (21), "*vida*" (20), "*saramago*" (18) "*um livro*"	"*conta coisas por as quais*" (30), "*sabe como escrever*" (30)
camera	"*megapixels*" (28), "*qualidade*" (26), "*compacta*" (23), ..., "*demora a responder*" (2), "*facil manuseio*"	"*fácil*" (25), "*foto*" (24), "*bom*" (20), "*excelente*" (18), "*uso*" (17)	"*linda*" (30), "*cada centavo investido*" (30), "*opções para fazer a foto*" (29), "*intuitiva*" (29)
hotel	"*custo*" (30), "*banheiro*" (30), "*cama*" (30), ..., "*diárias acessíveis*", "*lugar simples*"	"*ar*" (26), "*local*" (25), "*porta*" (24), "*lugar*" (23), "*banho*" (23)	"*sabia que era simples*" (30), "*localizado no*" (30), "*pertinho do*" (30), "*dos recepcionistas*" (30)
smartphone	"*moderno*" (19), "*bonito*" (18), "*trava*" ... "*modulo para filmar*", "*volumoso*", "*cheios de bugs*"	"*fácil*" (25), "*foto*" (23), "*excelente*" (23), "*qualidade*" (23), "*facil*" (22), "*bom*" (20)	"*funciona em qualquer lugar*" (29), "*navegar na internet*" (29), "*a recepção é muito boa*" (29)

In addition to the metrics presented, we analyzed the implicit aspects detected using the mentioned typology of aspects. Table 14 indicates the percentage of aspects detected in each typology category. Cells marked with – are those in which there are no aspects in the dataset in that category, as shown in Table 3.1.1. We analyzed the experiments of each group that obtained the best results in terms of implicit detection, excluding Hu & Liu for having results equal to those obtained by Freq Baseline.

In the domain of books, as mentioned previously, most of the implicit aspects are concentrated in the subcategory of verbs in expressions such as "*ler*" (read), "*sabe escrever*" (knows how to write), and "*fala de*" (speaks of), with more than half of these being detected by the LLM GPT-3.5. Within the contextual category, we find expressions that would be very difficult to detect, as they refer to specific characteristics of the book being analyzed. For example, "*caricatura cega de amor ou ódio*" (blind caricature of love or hate) and "*história sobre big brother*" (story about big brother), only with knowledge of the stories told, it would be possible to link these expressions to the plot aspect.

The camera and smartphone domains, which have similar distributions in relation to the categories, also had similar results in detecting implicit aspects when analyzed through typology. In the subcategory Verb, expressions such as "*fácil de usar*" (easy to use), and "*fácil de operar*" (easy to operate), and in Nominal Form, terms like "*facilidde de uso*" (ease of use) and "*tempo de resposta*" (response time) were detected. In the qualification category, adjectives such as "*barato*" (cheap), "*pequeno*" (small), and "*compacta*" (compact) and nominal forms such as "*qualidade*" (quality), "*beleza*" (beauty), and "*velocidade*" (speed) were also identified with good results by the LLMs.

Finally, for the hotel domain, the highlight was the methods that used the BERT model. Composed mainly of aspects belonging to the Feature categories, aspects from the subcategories Equivalence ("*apartamento*" room and"*equipe*" team), Is-a ("*café da manhã*" breakfast and "*jantar*" dinner), and Part-of ("*academia*" gym and "*cama*" bed) were detected. Among the contextual ones, we have those related to the location of the hotel ("*fica no centro*" it is in downtown "*localizado no centro*" located in downtown) that were well identified by BERT. For the related subcategory Related ("*cheiro de modo*" musty smell and "*infiltrações*" infiltrations) the methods BERT, Panchendrarajan, and LLMs achieved good detection rates.

5. Final Remarks

Detecting implicit aspects is a proven difficult task. Few methods deal directly with the issue of implicit aspects, as we could analyze the algorithm developed by Panchendrarajan et al. With the understanding of the formation of the implicit aspects through the typology, it was possible to improve the Freq-Baseline detection. The recent LLMs, despite not reaching such high f-measures, detected a significantly greater amount of implicit aspects and required much less manually annotated text. Given the rapid evolution of these models, expectations are high that they will achieve increasingly better results. The tested models, if observed in chronological order (GPT-3.5, Maritaca, and Llama-2), demonstrate this.

Although much more present in the literature, the detection of explicit aspects should also be analyzed more carefully. Most papers just calculate the traditional metrics (precision, recall, and f-measure) and finish the work. A greater understanding of the explicit aspects can help point out new ways to improve its detection. Another way to achieve better results is error analysis. With some observations we made, it is possible to implement adjustments in post-processing that would bring better results.

Suggestions for improvements for future work would be the development of a typology of explicit aspects and another for error analysis, which would help, for example, to point out the type of error in the classifier or manual annotation. Regarding methods, as mentioned, LLMs showed promise, especially in implicit aspect detection, a possible path would be the combination of traditional methods with LLMs.

References

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

Bottou, L. 2012. Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pp. 421–436. Springer.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA. Curran Associates

Table 14. Typological analysis of the detected implicit aspects

	Event		Feature				Qualification			Contextual	
Algorithm	NV	Verb	Att	Eq	Is-a	Part	Adj	Eq	NF	Loc	Rel
Domain: book											
Freq	–	0.0%	–	0.0%	0.0%	–	–	–	–	–	0.0%
Freq adj nom verb w2v	–	23.1%	–	0.0%	0.0%	–	–	–	–	–	0.0%
CRF	–	0.0%	–	0.0%	0.0%	–	–	–	–	–	0.0%
Panchendrarajan	–	23.1%	–	**40.0%**	0.0%	–	–	–	–	–	0.0%
BERTimbau in-domain	–	30.8%	–	20.0%	0.0%	–	–	–	–	–	0.0%
GPT-3.5	–	**53.8%**	–	20.0%	**100.0%**	–	–	–	–	–	0.0%
Maritaca	–	23.1%	–	**40.0%**	**100.0%**	–	–	–	–	–	0.0%
Llama 7B	–	30.8%	–	**40.0%**	0.0%	–	–	–	–	–	0.0%
Domain: camera											
Freq	0.0%	0.0%	62.5%	0.0%	–	–	0.0%	0.0%	33.3%	–	0.0%
Freq adj nom verb w2v	0.0%	0.0%	62.5%	0.0%	–	–	32.0%	0.0%	33.3%	–	0.0%
CRF	0.0%	11.1%	62.5%	0.0%	–	–	44.0%	0.0%	0.0%	–	0.0%
Panchendrarajan	**33.3%**	0.0%	62.5%	0.0%	–	–	40.0%	0.0%	33.3%	–	0.0%
GPT-3.5	**33.3%**	11.1%	**75.0%**	0.0%	–	–	48.0%	0.0%	**100.0%**	–	0.0%
BERTimbau in-domain	16.7%	0.0%	62.5%	0.0%	–	–	32.0%	0.0%	66.7%	–	0.0%
Maritaca	16.7%	22.2%	62.5%	0.0%	–	–	40.0%	0.0%	66.7%	–	0.0%
Llama 7B	**33.3%**	**33.3%**	**75.0%**	0.0%	–	–	**60.0%**	0.0%	**100.0%**	–	0.0%
Domain: hotel											
Freq	0.0%	0.0%	–	19.2%	52.8%	77.4%	0.0%	0.0%	0.0%	0.0%	29.3%
Freq adj nom verb w2v	0.0%	80.0%	–	39.7%	52.8%	64.5%	**100.0%**	0.0%	0.0%	0.0%	15.5%
CRF	0.0%	0.0%	–	48.7%	64.2%	85.5%	**100.0%**	0.0%	**100.0%**	44.4%	55.2%
Panchendrarajan	0.0%	80.0%	–	51.3%	**75.5%**	87.1%	**100.0%**	0.0%	**100.0%**	0.0%	65.5%
BERTimbau in-domain	0.0%	40.0%	–	**55.1%**	67.9%	**90.3%**	**100.0%**	10.0%	**100.0%**	**47.2%**	65.5%
GPT-3.5	0.0%	**100.0%**	–	30.8%	41.5%	71.8%	**100.0%**	10.0%	**100.0%**	5.6%	**69.0%**
Maritaca	0.0%	80.0%	–	32.1%	64.2%	88.7%	**100.0%**	**20.0%**	**100.0%**	0.0%	67.2%
Llama 7B	0.0%	40.0%	–	28.2%	56.6%	82.3%	**100.0%**	0.0%	**100.0%**	8.3%	62.1%
Domain: smartphone											
Freq	0.0%	0.0%	0.0%	–	0.0%	–	11.4%	–	0.0%	–	0.0%
Freq adj nom verb w2v	0.0%	33.3%	0.0%	–	0.0%	–	34.3%	–	0.0%	–	0.0%
CRF	0.0%	55.6%	0.0%	–	0.0%	–	42.9%	–	50.0%	–	0.0%
Panchendrarajan	0.0%	33.3%	0.0%	–	0.0%	–	**68.6%**	–	33.3%	–	0.0%
BERTimbau in-domain	0.0%	55.6%	**50.0%**	–	0.0%	–	34.3%	–	83.3%	–	0.0%
GPT-3.5	**33.3%**	55.6%	0.0%	–	0.0%	–	65.7%	–	83.3%	–	**66.7%**
Maritaca	**33.3%**	**61.1%**	0.0%	–	0.0%	–	48.6%	–	**100.0%**	–	33.3%
Llama 7B	**33.3%**	22.2%	0.0%	–	0.0%	–	25.7%	–	33.3%	–	0.0%

Inc.

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*, pp. 833–843, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Costa, R. and Pardo, T. 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*, pp. 61–72, Porto Alegre, RS, Brasil. SBC.

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.

De Freitas, L. A. and Vieira, R. 2015. Exploring resources for sentiment analysis in portuguese language. In *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 152–156.

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, pp. 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 *XI Encontro de Linguística de Corpus (ELC 2012)*, São Paulo.

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. International Conference on Computational Intelligence and Data Science.

Hartmann, N., Avanço, L., Balage, P., Duran, M., das Graças Volpe Nunes, M., Pardo, T., and Aluísio, S. 2014. A large corpus of product reviews in Portuguese: tackling out-of-vocabulary words. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 3865–3871, Reykjavik, Iceland. European Language Resources Association (ELRA).

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. 2020. spaCy: industrial-strength natural language processing in python.

Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, 168–177, New York, NY, USA. Association for Computing Machinery.

Liu, B. and others 2010. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. 2023. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.*, 55(9).

Machado, M. and Pardo, T. A. S. 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*, pp. 3819–3828, Marseille, France. European Language Resources Association.

Machado, M. T., Pardo, T. A. S., and Ruiz, E. E. S. 2017. Analysis of unsupervised aspect term identification methods for portuguese reviews. *Anais do XIV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC*, pp. 239–249.

Machado, M. T., Pardo, T. A. S., Ruiz, E. E. S., Di Felippo, A., and Vargas, F. 2022. Implicit opinion aspect clues in portuguese texts: analysis and categorization. In *Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings*, 68–78, Berlin, Heidelberg. Springer-Verlag.

Manning, C. D., Raghavan, P., and Schütze, H. 2008. *Introduction to information retrieval*. Cambridge University Press.

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*, NIPS'13, 3111–3119, Red Hook, NY, USA. Curran Associates Inc.

Panchendrarajan, R., Ahamed, N., Murugaiah, B., Sivakumar, P., Ranathunga, S., and Pemasiri, A. 2016. Implicit aspect detection in restaurant reviews using cooccurence of words. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 128–136, San Diego, California. 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 (LASM)*, pp. 44–52, Gothenburg, Sweden. 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 (SemEval-2016)*, pp. 19–30, San Diego, California. 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 (SemEval 2015)*, pp. 486–495, Denver, Colorado. Association for Computational Linguistics.

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 (SemEval 2014)*, pp. 27–35, Dublin, Ireland. Association for Computational Linguistics.

Qiu, G., Liu, B., Bu, J., and Chen, C. 2011. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27.

Souza, F., Nogueira, R., and Lotufo, R. 2020. Bertimbau: pretrained bert models for brazilian portuguese. In **Cerri, R. and**

Prati, R. C., editors, *Intelligent Systems*, pp. 403–417, Cham. Springer International Publishing.

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, oppportunities, and open challenges. *Information Processing & Management*, 54(4):545–563.

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*, pp. 342–351. Editora Unisinos.

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*, pp. 4339–4344, Miyazaki, Japan. European Language Resources Association.

Wang, S., **Sun, X.**, **Li, X.**, **Ouyang, R.**, **Wu, F.**, **Zhang, T.**, **Li, J.**, **and Wang, G.** 2023. Gpt-ner: named entity recognition via large language models. *arXiv preprint arXiv:2304.10428*.

Wang, X., **Jiang, Y.**, **Bach, N.**, **Wang, T.**, **Huang, Z.**, **Huang, F.**, **and Tu, K.** 2021. Automated concatenation of embeddings for structured prediction. In *the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (**ACL-IJCNLP 2021**)*. Association for Computational Linguistics.

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

Zhang, Y. and Zhu, W. 2013. Extracting implicit features in online customer reviews for opinion mining. In *Proceedings of the 22nd International Conference on World Wide Web*, pp. 103–104.

8 CONCLUSION

We started our research work by studying the rule-based method developed in (PORIA *et al.*, 2014). Although we did not achieve satisfactory results with the adaptation to Portuguese of the rules developed by the authors, this work motivated us to develop our own method for creating rules based on PoS tagging and dependency analysis, as seen in (MACHADO *et al.*, 2021).

Even though the analysis we carried out of the percentage of implicit aspects detected was simple, we did not find any studies in the reviewed literature that further explored the implicit aspects issue. To reduce this gap, we analyzed data from two corpora on four domains, aiming to understand what kind of knowledge was necessary to link a given implicit aspect clue with its respective explicit aspect. Our typology of implicit aspects was born from this study, which allowed deeper analysis of the results of aspect extraction methods. This study also resulted in an extension of the corpus presented by (FREITAS; VIEIRA, 2015), where we manually annotated the implicit aspects clues.

Our participation in ABSAPT task resulted in a technical report evaluating the Conditional Random Fields machine learning method, in which we carried out a detailed error analysis using the implicit aspects typology and trying to understand the origin of the error. Based on this analysis, we also implemented a post-processing step to improve the results.

At LREC we presented, implemented, and tested frequency, hybrid, machine learning and BERT methods. Once again, we analyzed the results using our proposed typology, which allowed us to see and understand the strengths and weaknesses of each method in detecting aspects.

In our latest work, submitted to the Natural Language Engineering journal, we improved the evaluation of methods using 5-fold cross validation, which allowed all implicit aspects of the datasets to be evaluated and not just those belonging to the test group. Furthermore, we carried out several experiments using LLMs.

A method we tested in all our papers was the Freq-Baseline (HU; LIU, 2004). As mentioned, it is a simple method that achieves reasonable results, thus being a good basis for comparison with other methods. It is important to highlight that, unlike the studies of other authors who use a fixed cutoff frequency of 1%, we always searched for the best cutoff frequency in the training sets, making the method achieve better results. We can conclude that this is an excellent baseline for comparison, because if a more elaborate method does not achieve the results of one so simple method, perhaps it is not worth it.

Our Corpus-based rules method (MACHADO *et al.*, 2021) inspired in Poria *et* *al.* (2014) achieved results close to the Freq-Baseline but demonstrated better efficiency in detecting implicit aspects. As we mentioned, due to the minor difference between the results from Freq-Baseline and our rule search method, we concluded that it would not be worth continuing to explore this path of rules in future works.

A second method we evaluated in our experiments was the Hu & Liu (HU; LIU, 2004) hybrid method. From its results, we observed something like what happened with the Poria et al. rule-based method. Satisfactory results in English datasets do not mean satisfactory results in Portuguese datasets. The results also showed the importance of a more comprehensive validation, when we expanded the evaluation to 5-folds the results of the method had a small drop.

Among the machine learning methods, the highlight was Conditional Random Fields, which obtained the best results (or something close to that) in all the experiments carried out. Analyzing only machine learning methods, this result was already expected, as it is a more suitable algorithm for sequence labeling, and due to it allows the use of more resources in classification, such as neighboring words and PoS tagging. Surprisingly, the method performed better than more recent methods such as BERT and the LLMs, demonstrating that, as happened with Hu & Liu, more complexity is not a guarantee of better results.

With the BERT pre-trained models, we obtained results close to the Conditional Random Fields. As a positive point of the experiments with this method, we can highlight a better performance concerning the implicit aspects and as a negative point the greater need for processing required by the method.

The Panchendrajaran et al. (PANCHENDRARAJAN *et al.*, 2016) method was one of the few that the authors were specifically concerned with the issue of implicit, and we could see this in the results. Unlike other methods adapted from the English language, this one also obtained reliable results in Portuguese (MACHADO; PARDO, 2022a), standing out in detecting implicit aspects.

In our last work, we analyzed the large language models. Despite the much lower f-measure results, the methods deserve to be highlighted due to the great advances in implicit detection. Another point in favor of the method concerns the number of manually annotated examples required for the few-shot approach, a much smaller quantity than that required for training machine learning models, for example. This ends up being a great advantage, given the difficulty and costs of the manual annotation process.

Finally, we rescue the hypotheses presented in the introduction of this thesis. Our first hypothesis was regarding pre-trained models. As we can see, these obtained results that are close to the best results of traditional machine learning models, with a small improvement in implicit detection. Perhaps its use is not justified, given its greater complexity and need for processing.

In our second hypothesis, we mentioned the possibility of extending methods to improve the detection of implicit aspects. This hypothesis was proven in the work of Panchendrajaran et al. (PANCHENDRARAJAN *et al.*, 2016), which expanded a machine learning method with an implicit detection stage, and in the experiment with Freq-baseline, in which using other PoS tagging categories and our lexicon of nominal forms to detect aspects, we can also improve the detection of implicit aspects.

The third hypothesis presented brings the need to model domain-specific knowledge to identify implicit aspects. The typology of implicit aspects proved this, demonstrating that distinct categories of implicit aspects stood out for each domain.

As future works, we hope to develop a typology for explicit aspects, seeking to categorize them in some way as we did with the implicit aspects. We hope to develop an error analysis methodology like the one we used in ABSAPT, with the aim of identifying what caused each error in the aspect identification process. Regarding methods, as mentioned, LLMs have shown promise, especially in implicit aspect detection, we hope to combine LLMs with other methods and hope to achieve better results. A method that aroused our interest was the one presented in Chapter 3(WANG *et al.*, 2021), which combines different embeddings and uses them in classification and achieved the best results in the revised bibliography. We believe that by adapting to Portuguese this method, we will also achieve superior results.

Finally, we hope that the methodology for analyzing implicit and resources that we created (datasets, typology, lexicons, implemented extraction methods, and evaluations of these) can contribute to advancing the state-of-the-art in the challenging task of detecting aspects, especially the implicit ones.

REFERENCES

Aires, J. P. *et al.* A deep learning approach to classify aspect-level sentiment using small datasets. *In*: **2018 International Joint Conference on Neural Networks (IJCNN)**. [*S.l.: s.n.*], 2018. p. 1–8.

AKBIK, A.; BLYTHE, D.; VOLLGRAF, R. Contextual string embeddings for sequence labeling. *In*: **Proceedings of the 27th International Conference on Computational Linguistics**. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. p. 1638–1649.

ALY, R.; REMUS, S.; BIEMANN, C. Hierarchical multi-label classification of text with capsule networks. *In*: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop**. Florence, Italy: Association for Computational Linguistics, 2019. p. 323–330.

BAKER, C. F.; FILLMORE, C. J.; LOWE, J. B. The Berkeley FrameNet project. *In*: **36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1**. Montreal, Quebec, Canada: Association for Computational Linguistics, 1998. p. 86–90.

Balage Filho, P.; PARDO, T. Nilc_usp: aspect extraction using semantic labels. *In*: **Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)**. [*S.l.: s.n.*], 2014. p. 433–436.

BOJANOWSKI, P. *et al.* Enriching Word Vectors with Subword Information. **Transactions of the Association for Computational Linguistics**, v. 5, p. 135–146, 06 2017.

BOTTOU, L. Stochastic gradient descent tricks. *In*: _____. **Neural Networks: Tricks of the Trade: Second Edition**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 421–436.

BRUNA, J. *et al.* Spectral networks and deep locally connected networks on graphs. *In*: **2nd International Conference on Learning Representations, ICLR 2014**. [*S.l.: s.n.*], 2014.

BURNAP, P. *et al.* Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. **Social Network Analysis and Mining**, Springer, v. 4, n. 1, p. 206, 2014.

CAI, H. *et al.* Aspect-category based sentiment analysis with hierarchical graph convolutional network. *In*: **Proceedings of the 28th International Conference on Computational Linguistics**. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020. p. 833–843.

CALZOLARI, N. *et al.* (ed.). **Proceedings of the Thirteenth Language Resources and Evaluation Conference**. Marseille, France: European Language Resources Association, 2022.

CAMBRIA, E. Affective computing and sentiment analysis. **IEEE intelligent systems**, IEEE, v. 31, n. 2, p. 102–107, 2016.

CAMBRIA, E. *et al.* An elm-based model for affective analogical reasoning. **Neurocomputing**, Elsevier, v. 149, p. 443–455, 2015.

CAMBRIA, E.; OLSHER, D.; RAJAGOPAL, D. Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. *In*: **Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence**. [*S.l.: s.n.*]: AAAI Press, 2014. (AAAI'14), p. 1515–1521.

CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems and their applications**, IEEE, v. 14, n. 1, p. 20–26, 1999.

COLLOBERT, R. Deep learning for efficient discriminative parsing. *In*: **Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics**. Fort Lauderdale, FL, USA: PMLR, 2011. (Proceedings of Machine Learning Research, v. 15), p. 224–232.

CONDORI, R. E. L.; PARDO, T. A. S. Opinion summarization methods: Comparing and extending extractive and abstractive approaches. **Expert Systems with Applications**, Elsevier, v. 78, p. 124–134, 2017.

CONNEAU, A. *et al.* Unsupervised cross-lingual representation learning at scale. *In*: **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. Online: Association for Computational Linguistics, 2020. p. 8440–8451.

COSTA, R.; PARDO, T. Métodos baseados em léxico para extração de aspectos de opiniões em português. *In*: **Anais do IX Brazilian Workshop on Social Network Analysis and Mining**. Porto Alegre, RS, Brasil: SBC, 2020. p. 61–72.

CRAMMER, K. *et al.* Online passive-aggressive algorithms. **Journal of Machine Learning Research**, v. 7, n. 19, p. 551–585, 2006.

CRUZ, I.; GELBUKH, A. F.; SIDOROV, G. Implicit aspect indicator extraction for aspect based opinion mining. **International Journal of Computational Linguistics and Applications**, v. 5, n. 2, p. 135–152, 2014.

DAS, D. *et al.* Semafor 1.0: A probabilistic frame-semantic parser. **Language Technologies Institute, School of Computer Science, Carnegie Mellon University**, 2010.

DAVE, K.; LAWRENCE, S.; PENNOCK, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *In*: **Proceedings of the 12th International Conference on World Wide Web**. New York, NY, USA: Association for Computing Machinery, 2003. (WWW '03), p. 519–528.

DEVLIN, J. *et al.* **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. 2019.

DING, X.; LIU, B.; YU, P. S. A holistic lexicon-based approach to opinion mining. *In*: **Proceedings of the 2008 international conference on web search and data mining**. [*S.l.: s.n.*], 2008. p. 231–240.

DROBA, D. D. Methods for measuring attitudes. **Psychological Bulletin**, Psychological Review Company, v. 29, n. 5, p. 309, 1932.

FREITAS, C. *et al.* Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. *In*: **XI Encontro de Linguística de Corpus (ELC 2012)**. São Paulo: [*S.l.: s.n.*], 2012.

FREITAS, L. A. D.; VIEIRA, R. Exploring resources for sentiment analysis in portuguese language. *In*: **2015 Brazilian Conference on Intelligent Systems (BRACIS)**. [*S.l.: s.n.*], 2015. p. 152–156.

FREUND, Y.; SCHAPIRE, R. E. Large margin classification using the perceptron algorithm. **Machine learning**, Springer, v. 37, n. 3, p. 277–296, 1999.

GANDHI, H.; ATTAR, V. Extracting aspect terms using crf and bi-lstm models. **Procedia Computer Science**, v. 167, p. 2486–2495, 2020. International Conference on Computational Intelligence and Data Science.

GANGANWAR, V.; RAJALAKSHMI, R. Implicit aspect extraction for sentiment analysis: A survey of recent approaches. **Procedia Computer Science**, Elsevier, v. 165, p. 485–491, 2019.

GANU, G.; ELHADAD, N.; MARIAN, A. Beyond the stars: improving rating predictions using review text content. *In*: **WebDB**. [*S.l.: s.n.*], 2009. v. 9, p. 1–6.

GóMEZ, M. M. y *et al.* (ed.). **Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022)**, (CEUR Workshop Proceedings). [*S.l.: s.n.*]: CEUR-WS.org, 2022.

HARTMANN, N. *et al.* A large corpus of product reviews in Portuguese: tackling out-of-vocabulary words. *In*: **Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)**. Reykjavik, Iceland: European Language Resources Association (ELRA), 2014. p. 3865–3871.

HU, M.; LIU, B. Mining and summarizing customer reviews. *In*: **Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2004. (KDD '04), p. 168–177.

JAIN, L. C.; MEDSKER, L. R. **Recurrent Neural Networks: Design and Applications**. 1st. ed. USA: CRC Press, Inc., 1999.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 1st. ed. USA: Prentice Hall PTR, 2000.

KARLGREN, J.; SAHLGREN, M. From words to understanding. **Foundations of real-world intelligence**, CSLI publications, p. 294–308, 2001.

KIPF, T. N.; WELLING, M. Semi-supervised classification with graph convolutional networks. *In*: **5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings**. [*S.l.: s.n.*]: OpenReview.net, 2017.

KNUTSON, A. L. Japanese opinion surveys: the special need and the special difficulties. **Public Opinion Quarterly**, Oxford University Press, v. 9, n. 3, p. 313–319, 1945.

LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *In*: **Proceedings of the Eighteenth International Conference on Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. (ICML '01), p. 282–289.

LAMPLE, G. *et al.* Neural architectures for named entity recognition. *In*: **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. San Diego, California: Association for Computational Linguistics, 2016. p. 260–270.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, May 2015.

LIU, B. Sentiment analysis and opinion mining. **Synthesis Lectures on Human Language Technologies**, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.

LIU, B.; HU, M.; CHENG, J. Opinion observer: analyzing and comparing opinions on the web. *In*: **Proceedings of the 14th international conference on World Wide Web**. [*S.l.: s.n.*], 2005. p. 342–351.

LONG, C.; ZHANG, J.; ZHUT, X. A review selection approach for accurate feature rating estimation. *In*: **Proceedings of the 23rd International Conference on Computational Linguistics: Posters**. USA: Association for Computational Linguistics, 2010. (COLING '10), p. 766–774.

MACHADO, M. *et al.* Learning rules for automatic identification of implicit aspects in portuguese. *In*: **Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana**. Porto Alegre, RS, Brasil: SBC, 2021. p. 82–91.

MACHADO, M.; PARDO, T. A. S. 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**. Marseille, France: European Language Resources Association, 2022. p. 3819–3828.

MACHADO, M. T.; PARDO, T. A. S. Nilc at absapt 2022: Aspect extraction for portuguese. *In*: **Proceedings**. [*S.l.: s.n.*], 2022.

MACHADO, M. T.; PARDO, T. A. S. Evaluating from the classical frequency to the large language model-based methods for aspect extraction in portuguese and their performance on detecting typologically different implicit aspects. Submitted to Natural Language Engineering. 2023.

MACHADO, M. T.; PARDO, T. A. S.; RUIZ, E. E. S. Analysis of unsupervised aspect term identification methods for portuguese reviews. **Anais do XIV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), SBC**, p. 239–249, 2017.

MACHADO, M. T. *et al.* Implicit opinion aspect clues in portuguese texts: analysis and categorization. *In*: **Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings**. Berlin, Heidelberg: Springer-Verlag, 2022. p. 68–78.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. [*S.l.: s.n.*]: Cambridge University Press, 2008.

MÄNTYLÄ, M. V.; GRAZIOTIN, D.; KUUTILA, M. The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. **Computer Science Review**, v. 27, p. 16–32, 2018.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. **Ain Shams engineering journal**, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.

MIKOLOV, T. *et al.* **Efficient Estimation of Word Representations in Vector Space**. 2013.

MIKOLOV, T. *et al.* Distributed representations of words and phrases and their compositionality. *In*: **Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2**. Red Hook, NY, USA: Curran Associates Inc., 2013. (NIPS'13), p. 3111–3119.

MILLER, G. A. Wordnet: a lexical database for english. **Communications of the ACM**, ACM New York, NY, USA, v. 38, n. 11, p. 39–41, 1995.

MILLER, G. A. **WordNet: An electronic lexical database**. [*S.l.: s.n.*]: MIT press, 1998.

NASSIRTOUSSI, A. K. *et al.* Text mining for market prediction: A systematic review. **Expert Systems with Applications**, Elsevier, v. 41, n. 16, p. 7653–7670, 2014.

PANCHENDRARAJAN, R. *et al.* Implicit aspect detection in restaurant reviews using cooccurence of words. *In*: **Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis**. San Diego, California: Association for Computational Linguistics, 2016. p. 128–136.

PANG, B.; LEE, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *In*: **Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)**. Ann Arbor, Michigan: Association for Computational Linguistics, 2005. p. 115–124.

PAVLOPOULOS, J.; ANDROUTSOPOULOS, I. 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 (LASM)**. Gothenburg, Sweden: Association for Computational Linguistics, 2014. p. 44–52.

PENNINGTON, J.; SOCHER, R.; MANNING, C. GloVe: Global vectors for word representation. *In*: **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Doha, Qatar: Association for Computational Linguistics, 2014. p. 1532–1543.

PEREIRA, D. A. A survey of sentiment analysis in the portuguese language. **Artificial Intelligence Review**, Springer, p. 1–29, 2020.

PETERS, M. E. *et al.* Deep contextualized word representations. *In*: **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**. New Orleans, Louisiana: Association for Computational Linguistics, 2018. p. 2227–2237.

PINHEIRO, V. *et al.* (ed.). **Computational Processing of the Portuguese Language - 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21-23, 2022, Proceedings**, v. 13208 de **Lecture Notes in Computer Science**, (Lecture Notes in Computer Science, v. 13208). [*S.l.: s.n.*]: Springer, 2022.

PONTIKI, M. *et al.* SemEval-2015 task 12: aspect based sentiment analysis. *In*: **Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)**. Denver, Colorado: Association for Computational Linguistics, 2015. p. 486–495.

PONTIKI, M. *et al.* SemEval-2016 task 5: aspect based sentiment analysis. *In*: **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)**. San Diego, California: Association for Computational Linguistics, 2016. p. 19–30.

PONTIKI, M. *et al.* SemEval-2014 task 4: aspect based sentiment analysis. *In*: **Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)**. Dublin, Ireland: Association for Computational Linguistics, 2014. p. 27–35.

POPESCU, A.-M.; ETZIONI, O. Extracting product features and opinions from reviews. *In*: ______. **Natural Language Processing and Text Mining**. London: Springer London, 2007. p. 9–28.

PORIA, S. *et al.* A rule-based approach to aspect extraction from product reviews. *In*: LIN, S. de *et al.* (ed.). **SocialNLP 2014 - 2nd Workshop on Natural Language Processing for Social Media, in conjunction with COLING 2014**. [*S.l.: s.n.*]: Association for Computational Linguistics (ACL), 2014. (SocialNLP 2014 - 2nd Workshop on Natural Language Processing for Social Media, in conjunction with COLING 2014), p. 28–37.

QIU, G. *et al.* Opinion word expansion and target extraction through double propagation. **Computational Linguistics**, MIT Press, Cambridge, MA, v. 37, n. 1, p. 9–27, mar. 2011.

RANA, T. A.; CHEAH, Y.-N. Aspect extraction in sentiment analysis: comparative analysis and survey. **Artificial Intelligence Review**, Springer, v. 46, n. 4, p. 459–483, 2016.

RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. **Knowledge-Based Systems**, v. 89, p. 14 – 46, 2015.

RICHMOND, J. A. Spies in ancient greece. **Greece and Rome**, Cambridge University Press, v. 45, n. 1, p. 1–18, 1998.

RUIZ, E. E. S.; TORRENT, T. T. Proceedings of the 13th brazilian symposium in information and human language technology. *In*: **Proceedings of the 13th Brazilian Symposium in Information and Human Language Technology**. [*S.l.: s.n.*], 2021.

SCHMIDHUBER, J.; HOCHREITER, S. Long short-term memory. **Neural Comput**, v. 9, n. 8, p. 1735–1780, 1997.

SCHMITT, M. *et al.* Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. *In*: **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**. Brussels, Belgium: Association for Computational Linguistics, 2018. p. 1109–1114.

SCHOUTEN, K. *et al.* Semantics-driven implicit aspect detection in consumer reviews. *In*: **Proceedings of the 24th International Conference on World Wide Web**. New York, NY, USA: Association for Computing Machinery, 2015. (WWW '15 Companion), p. 109–110.

SILLA, C. N.; FREITAS, A. A. A survey of hierarchical classification across different application domains. **Data Mining and Knowledge Discovery**, Springer, v. 22, n. 1-2, p. 31–72, 2011.

SNYDER, B.; BARZILAY, R. Multiple aspect ranking using the good grief algorithm. *In*: **Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference**. Rochester, New York: Association for Computational Linguistics, 2007. p. 300–307.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. Bertimbau: pretrained bert models for brazilian portuguese. *In*: CERRI, R.; PRATI, R. C. (ed.). **Intelligent Systems**. Cham: Springer International Publishing, 2020. p. 403–417.

STAGNER, R. The cross-out technique as a method in public opinion analysis. **The Journal of Social Psychology**, Taylor & Francis, v. 11, n. 1, p. 79–90, 1940.

TABOADA, M. Sentiment analysis: an overview from linguistics. **Annual Review of Linguistics**, v. 2, n. 1, p. 325–347, 2016.

THET, T. T.; NA, J.-C.; KHOO, C. S. Aspect-based sentiment analysis of movie reviews on discussion boards. **Journal of information science**, Sage Publications Sage UK: London, England, v. 36, n. 6, p. 823–848, 2010.

THORLEY, J. **Athenian democracy**. [*S.l.: s.n.*]: Psychology Press, 2004.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, Springer, v. 24, n. 3, p. 478–514, 2012.

TUBISHAT, M.; IDRIS, N.; ABUSHARIAH, M. A. Implicit aspect extraction in sentiment analysis: review, taxonomy, oppportunities, and open challenges. **Information Processing & Management**, v. 54, n. 4, p. 545–563, 2018.

VARGAS, F. A.; PARDO, T. A. S. Clustering and hierarchical organization of opinion aspects: a corpus study. *In*: **Proceedings of the 14th Corpus Linguistics Meeting**. [*S.l.: s.n.*]: Editora Unisinos, 2017. p. 342–351.

VARGAS, F. A.; PARDO, T. A. S. Aspect Clustering Methods for Sentiment Analysis. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 11122 LNAI, p. 365–374, 2018.

WANG, X. *et al.* Structure-level knowledge distillation for multilingual sequence labeling. *In*: **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. Online: Association for Computational Linguistics, 2020. p. 3317–3330.

WANG, X. *et al.* Automated concatenation of embeddings for structured prediction. *In*: ZONG, C. *et al.* (ed.). **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021**. [*S.l.: s.n.*]: Association for Computational Linguistics, 2021. p. 2643–2660.

WEI, Z. *et al.* Don't eclipse your arts due to small discrepancies: Boundary repositioning with a pointer network for aspect extraction. *In*: **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. Online: Association for Computational Linguistics, 2020. p. 3678–3684.

XU, H. *et al.* Double embeddings and CNN-based sequence labeling for aspect extraction. *In*: **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**. Melbourne, Australia: Association for Computational Linguistics, 2018. p. 592–598.

XU, H. *et al.* BERT post-training for review reading comprehension and aspect-based sentiment analysis. *In*: **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 2324–2335.

YADOLLAHI, A.; SHAHRAKI, A. G.; ZAIANE, O. R. Current state of text sentiment analysis from opinion to emotion mining. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 50, n. 2, p. 1–33, 2017.

YOUNG, T. *et al.* Recent trends in deep learning based natural language processing. **IEEE Computational intelligenCe magazine**, IEEE, v. 13, n. 3, p. 55–75, 2018.

YU, J. *et al.* Aspect ranking: Identifying important product aspects from online consumer reviews. *In*: **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1**. USA: Association for Computational Linguistics, 2011. (HLT '11), p. 1496–1505.

ZHAI, Z. *et al.* Grouping product features using semi-supervised learning with soft-constraints. *In*: **Proceedings of the 23rd International Conference on Computational Linguistics**. USA: Association for Computational Linguistics, 2010. (COLING '10), p. 1272–1280.

ZHAI, Z. *et al.* Clustering product features for opinion mining. *In*: **Proceedings of the Fourth ACM International Conference on Web Search and Data Mining**. New

York, NY, USA: Association for Computing Machinery, 2011. (WSDM '11), p. 347–354. ISBN 9781450304931.

ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, p. e1253, 2018.

ZHANG, Y.; ZHU, W. Extracting implicit features in online customer reviews for opinion mining. *In*: **Proceedings of the 22nd International Conference on World Wide Web**. New York, NY, USA: Association for Computing Machinery, 2013. (WWW '13 Companion), p. 103–104.

APPENDIX

APPENDIX A – NILC AT ABSAPT 2022: ASPECT EXTRACTION FOR PORTUGUESE

Our third work was the technical report entitled "NILC at ABSAPT 2022: Aspect Extraction for Portuguese" (MACHADO; PARDO, 2022b) which presented the method implemented for the shared task Aspect-Based Sentiment Analysis in Portuguese (ABSAPT 2022), which was part of the Iberian Languages Evaluation Forum (IberLEF 2022) (GóMEZ *et al.*, 2022) and took place in September 20, 2022. The technical report was published in the proceedings of the event[1].

The Aspect-Based Sentiment Analysis in Portuguese was the first shared task dedicated to identifying aspects and extracting the polarity in texts written in Portuguese. The IberLEF is a shared evaluation campaign for Natural Language Processing (NLP) systems in Spanish and other Iberian languages, consisting of several challenges that are executed with a large international participation of research groups from academia and industry.

In this work, we tested the methods Freq-Baseline (HU; LIU, 2004), Hu & Liu (HU; LIU, 2004), both extended with a Word2Vec (MIKOLOV *et al.*, 2013b) pruning step (PAVLOPOULOS; ANDROUTSOPOULOS, 2014; MACHADO; PARDO; RUIZ, 2017), Multinomial Naive Bayes (MANNING; RAGHAVAN; SCHÜTZE, 2008), Passive Aggressive (CRAMMER *et al.*, 2006), Perceptron (FREUND; SCHAPIRE, 1999), Stochastic Gradient Descent (BOTTOU, 2012), Conditional Random Fields (CRF) (GANDHI; ATTAR, 2020), and BERT pre-trained models (DEVLIN *et al.*, 2019; SOUZA; NOGUEIRA; LOTUFO, 2020). After these initial tests, we chose CRF to process the ABSAPT dataset. After an error analysis, we implemented some improvements in the way of post-processing. The final results were presented, and we carried out an error analysis of the results, containing incorrectly detected and undetected aspects and their categories according to the typology presented in the previous work. Below is the complete work, inserted here with the permission of the editor.

[1] https://ceur-ws.org/Vol-3202/

NILC at ABSAPT 2022: Aspect Extraction for Portuguese

Mateus Tarcinalli Machado[1], Thiago Alexandre Salgueiro Pardo[1]

[1]*University of São Paulo, Interinstitutional Center for Computational Linguistics (NILC), Institute of Mathematical and Computer Sciences, Av Trabalhador São-carlense, 400, São Carlos, São Paulo, Brazil*

Abstract

In this paper, we describe the adopted method to extract aspects of the Aspect-Based Sentiment Analysis in Portuguese (ABSAPT) task proposed in the context of the evaluation campaign of IberLEF 2022. We used the Conditional Random Fields (CRF) machine learning algorithm combined with a post-processing step achieving an accuracy of 0.97. We performed a detailed analysis of the erroneously detected and non-detected aspects, with both lists formed only by implicit aspects. For the analysis, we used a refined typology of the kind of knowledge necessary to identify implicit aspects. Finally, we conclude that with a simpler machine learning algorithm like CRF it is possible to achieve good results with a lower processing cost than deep learning methods. The main aspect identification errors are related to implicit aspects, evidencing the greater difficulty in finding this type of aspect.

Keywords

Aspect-based sentiment analysis, Aspect Extraction, Implicit aspects

1. Introduction

We can define sentiment analysis as an area of study at the intersection between Computer Science and Linguistics that aims to automatically analyze text [1] to understand people's opinions, feelings, assessments, attitudes, and emotions concerning products, services, organizations, individuals, issues and topics [2].

One of the first areas of application of sentiment analysis was the evaluation of products, conducted through the analysis of opinion texts obtained from social networks or specialized websites. This analysis can be done at a more general (about the product) or more specific level (about the characteristics of the product). From this need to qualify specific characteristics of each item or object analyzed, the area known as aspect-based sentiment analysis (ABSA) was born. ABSA performs deeper analysis on opinion texts, identifying and relating sentiments and aspects [3] of a given entity.

Most systems that perform aspect-based sentiment analysis divide the problem into three steps [4]:

- Extraction of aspect terms: extracts the characteristics (aspects) about the analyzed entity (product). For example, about a hotel, we may have aspects related to its facilities, location,

IberLEF 2022, September 2022, A Coruña, Spain.
mateusmachado@usp.br (M. T. Machado); taspardo@icmc.usp.br (T. A. S. Pardo)
0000-0003-1848-3082 (M. T. Machado); 0000-0003-2111-1319 (T. A. S. Pardo)

and employees, among others [2, 5, 6, 7].

- Aspect term sentiment estimation: this step estimates the polarity (positive, negative, or neutral) and possibly the intensity (e.g., strongly negative, moderately positive) of opinions for each aspect term of the target entity [2, 8, 9].
- Aspect aggregation: some systems group aspect terms that are synonymous ("price", "cost") or, more generally, group aspect terms to obtain aspects of a more general granularity ("concierge", "manager" and "receptionist" can be replaced by "employees") [2, 5, 10, 11].

Something that directly influences aspect-based sentiment analysis results is how aspects appear in texts. They can be explicitly mentioned, as in the example "The location of the hotel is excellent". In other situations, the aspect is implicitly mentioned, as in "It's close to the city center", where the same location aspect is indirectly mentioned, making the aspect detection task more challenging.

This paper concerns our participation at IberLEF 2022 [12], a shared evaluation campaign for Natural Language Processing (NLP) systems In Spanish and other Iberian languages. In particular, we have participated in the Aspect-Based Sentiment Analysis in Portuguese (ABSAPT) task in its Aspect Term Extraction sub-task. We describe our aspect extraction method using a classifier based on the Conditional Random Fields algorithm and a post-processing step. We also present here a detailed analysis of the errors found by the proposed method, especially regarding the implicit aspects and a refined typology of necessary knowledge to identify them.

The rest of the paper is organized as follows. Section 2 presents the data and methods used for the task. Section 3 shows the results and analysis of errors. Section 4 presents some final remarks.

2. Data and methods

In our most recent work [13], we analyzed fourteen methods for aspect extraction, with a focus on implicit aspects. We processed opinion texts from camera, book, smartphone, and hotel domains. In the hotel domain (which is the one related to the ABSAPT evaluation), we analyzed the dataset of the work of [14] in conjunction with the annotation of implicit aspects that we conducted [15].

For the task of extraction of aspects, we tested frequency-based (FreqBaseline [16]) and frequency- and rule-based hybrid methods (Hu & Liu [16]), both extended with the use of a pruning step to remove unrelated aspects using the Word2Vec [17] distributional model [4, 18]. We also applied the Multinomial Naive Bayes [19], Passive Aggressive [20], Perceptron [21], and Stochastic Gradient Descent [22] machine learning algorithms using the bag-of-words model as features. For the Conditional Random Fields [23] algorithm, we evaluated the tokens and their neighbors with their respective POS tags as features. In the last experiment, we tested pre-trained BERT models, where we performed in-domain and cross-domain experiments with multilanguage[1] [24] and Portuguese[2] [25] models.

[1]https://huggingface.co/bert-base-multilingual-uncased
[2]https://huggingface.co/neuralmind/bert-base-portuguese-cased

Table 1
Results for the hotel domain

Method	Precision	Recall	F1	% implicit aspects
FreqBaseline	0.74	0.51	0.61	0.12
FreqBaseline + Word2Vec	0.72	0.52	0.60	0.15
Hu & Liu	0.78	0.46	0.58	0.08
Hu & Liu + Word2Vec	0.79	0.47	0.59	0.09
Hu & Liu + Infrequent Word2Vec	0.82	0.45	0.58	0.08
Multinomial Naive Bayes	0.49	0.56	0.52	0.14
Passive Aggressive	**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
Conditional Random Fields	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 1 presents the results obtained in these experiments for the hotel domain, and we can see that the Passive Aggressive machine learning method had the best precision, the pre-trained BERT language model had the best recall, and the Conditional Random Fields (CRF) had the best F-measure. The last column shows the percentage of implicit aspects that each method identified. As the F-measure is a harmonic average between the two previous measures, we chose the CRF to carry out the experiments of ABSAPT task, in addition to having the possibility of analyzing other combinations of features in the training of models with this algorithm. CRF was also the best method for extracting the implicit aspects.

2.1. Corpus

For the ABSAPT task, the organizers made available a corpus developed by Freitas [26] and Correa [27], containing reviews about accommodation service companies, written in Portuguese. The authors annotated the aspects found and their respective polarities in each review, following the same guidelines in both works. The final dataset contains 847 reviews, comprising 3,090 aspects (with 77 unique aspects).

2.2. Pre-processing

In the pre-processing step, we prepare the texts for applying the machine learning method. In this experiment, we divided the reviews into sentences and then into tokens. We also did the morphosyntactic analysis using the part of speech tagger of the spaCy [28] module with the model pt_core_news_lg[3]. After this step, we converted this data to the IOB format (short for

[3]https://spacy.io/models/pt

Inside, Outside, and Beginning annotation format). To identify the aspects, we used the tags "B-ASP" for Begin of Aspect, "I-ASP" for Inside of Aspect, and "O" for Outside of Aspect.

2.3. Method for aspect extraction

As we explained earlier, we chose the Conditional Random Fields algorithm for this experiment. It uses probabilistic graph models that consider characteristics of words and their surroundings, suitable for segmenting and labeling data sequences [23], which are desired characteristics for the aspect identification task.

In the experiment, we used the module sklearn-crfsuite [29]. As features for training the model, we choose the tokens, their neighbors and respective morphosyntactic labels. We divided the dataset into a training set with 70% of the sentences and a test set with the remaining 30%. For optimization, we used the random search method, given the continuous nature of the parameters c_1 and c_2 of the algorithm implementation. We performed the training and validation steps 1,000 times with random values for these variables and selected the best result by the accuracy measure, which was 0.93 for c1 of 0.3914 and c2 of 0.0593, all approximated values.

The next step was to divide the dataset into five parts to perform and validate the tests in 5-folds. Using the parameters found with the random search, we found an average accuracy of 0.96 as shown in Table 2. We then analyzed the sets of aspects found, not found, and those erroneously found, as follows.

- Correctly identified aspects : '*aeroporto*', '*almoço*', '*apartamento*', '*apartamentos*', '*atendimento*', '***banheira***', '*cafeteira*', '*café da manhã*', '*cama*', '*camas*', '*carpete*', '*carpetes*', '*casal*', '*cassino*', '*cassinos*', '*centro da cidade*', '***chuveiro***', '***cidade***', '*colchão*', '*conforto*', '*corredor*', '*corredores*', '*cozinha*', '*custo-benefício*', '*ducha*', '*duchas*', '*elevador*', '*elevadores*', '*escadas*', '***estacionamento***', '*família*', '*famílias*', '*farmácia*', '*farmácias*', '*frigobar*', '*funcionários*', '*garagem*', '***gerente***', '*horário*', '*horários*', '*hotel*', '*instalações*', '*internet*', '***jantar***', '*lavanderia*', '*limpeza*', '*localização*', '*lojas*', '*luxo*', '***museu***', '*móveis*', '*padrão*', '*pia*', '*piscina*', '*piscinas*', '*praia*', '***praça***', '*preço*', '*preços*', '*quarto*', '*quartos*', '*recepção*', '*rodoviária*', '*roupa de cama*', '*rua*', '*ruas*', '***secador***', '*serviço*', '*serviço de quarto*', '*serviços*', '***shopping***', '*suíte*', '*suítes*', '*teatro*', '*teatros*', '*telefone*', '*televisão*', '*toalha*', '*toalhas*', '*tv*'
- Incorrectly identified aspects: '*centro de cidade*', '*colchão da cama*', '*serviço de café da manhã*'
- Aspects not found: '*aquecimento*', '***banheiras***', '*calefação*', '***chuveiros***', '***cidades***', '*cortinas*', '*escada*', '*estabelecimento*', '***estacionamentos***', '***gerentes***', '*gerência*', '*iluminação*', '***jantares***', '*jardim*', '*motel*', '***museus***', '*parque*', '*porteiro*', '***praças***', '***secadores***', '***shoppings***', '*tapete*', '*tomada*', '*tomadas*', '*torneira*', '*torneiras*', '*travesseiros*', '*turistas*'

In the list of aspects that were not found, we could find some terms in plural and, looking at the list of correctly identified aspects, we realized that ten of them were there in their singular form (aspects in bold in the lists). We then carried out some tests that did not prove to be effective. In a first attempt to identify these aspects, we added the lemma of each token as a feature to the classifier, so that plural and singular words would be converted to the same lemma. In a second attempt, we replaced the token feature with its respective lemma. Finally, we added

Table 2
Results with average and standard deviation in the experiments with 5 folds.

Method	Precision		Recall		F Measure		Accuracy	
	Avg.	Dev.	Avg.	Dev.	Avg.	Dev.	Avg.	Dev.
CRF	**0.9933**	0.0027	0.9579	0.0058	0.9753	0.0021	0.9579	0.0058
CRF + Pos-Processing	0.9914	0.0025	**0.9663**	0.0068	**0.9787**	0.0029	**0.9663**	0.0068

a post-processing step comparing the lemmas of the words of the texts with the lemmas of the aspects found by the classifier. Although not effective, this last approach had the best result, which led us to implement a post-processing step analyzing the plural form of the aspects found. To do this, we converted all aspects that were in singular form to their respective plural form and included them as possible aspects to be checked to compose the list of aspects found. This way, we could reduce the list of aspects not found (indicated below in strikethrough style), but three new aspects were added to the incorrect list (aspects in bold).

- Incorrectly identified aspects: '*centro de cidade*', '*colchão da cama*', '***lavanderias***', '***luxos***', '*serviço de café da manhã*', '***tvs***'
- Aspects not found: '*aquecimento*', '~~***banheiras***~~', '*calefação*', '~~***chuveiros***~~', '~~***cidades***~~', '*cortinas*', '*escada*', '*estabelecimento*', '~~***estacionamentos***~~', '~~***gerentes***~~', '*gerência*', '*iluminação*', '~~***jantares***~~', '*jardim*', '*motel*', '***museus***', '*parque*', '*porteiro*', '~~***praças***~~', '~~***secadores***~~', '~~***shoppings***~~', '*tapete*', '*tomada*', '*tomadas*', '*torneira*', '*torneiras*', '*travesseiros*', '*turistas*'

3. Results and error analysis

We calculated the average and standard deviation of the metrics found in the 5 folds based on the obtained results. We computed precision, recall, f-measure, and accuracy. We can see that, despite our post-processing method having located nine new aspects, the difference between the results was minimal, as we can see in Table 2. Anyway, when we deal with large volumes of data, all performance gain makes a difference.

After we analyzed the results of the CRF, we conduct an error analysis of the aspects not found and of those erroneously marked at the end of the post-processing step. A first important observation is that all aspects of the two lists fit into the classification of implicit aspects that we presented, a fact that highlights the difficulty of finding such aspects. This fact made it possible to use our typology of implicit aspects [15] to carry out this analysis. In this typology, we aim to characterize the implicit aspects in categories and subcategories according to the knowledge necessary for their identification. Table 3 presents a summary of the categories and subcategories in the typology.

In Tables 4 and 5, we present the aspect terms with their respective aspects, the categories and subcategories found in the typology and the possible types of errors that led to the incorrect classification. The aspects followed the annotation performed in Freitas and Vieira[14].

The first case of incorrectly detected aspect was the expression '*centro de cidade*'. This classification was certainly produced because we have the expression '*centro da cidade*' often

Table 3
Typology of knowledge necessary to identify implicit aspects: categories, subcategories, aspects and examples

Category	Subcategory	Aspect	Example
Event	Non-verbal form	Price	*pagamento* (payment)
	Verb	Meal	*comer* (to eat)
Feature	Attribute	Design	*material* (material)
	Equivalence	Employee	*equipe* (team)
	Is-a	Employee	*recepcionista* (receptionist)
	Part-of	Facilities	*banheiro* (bathroom)
Qualification	Adjective	Installation	*mobiliado* (furnished)
	Equivalence	Installation	*hotel simples* (simple hotel)
	Nominal form	Location	*proximidade* (proximity)
Contextual	Location	Location	*fácil acesso ao centro* (easy access to downtown)
	Related	Cleanliness	*cheira a mofo* (musty smell)

Table 4
Incorrect aspects.

Aspect term	Aspect	Category	Subcategory	Error
centro de cidade (center of a city)	Location	Contextual	Local	Classifier
colchão da cama (bed mattress)	Sleep quality	Contextual	Related	Classifier
lavanderias (laundries)	Location	Contextual	Local	Annotation
luxos (luxuries)	Facilities/Room	Qualification	Equivalence	Annotation
serviço de café da manhã (breakfast service)	Meal	Feature	is-a	Classifier

appearing related to the hotel's aspect location. Interestingly, both forms appear in the same sentence: "*Fica bem localizado, no* ***centro da cidade****, o que é bom para quem vai resolver coisas por perto, mas ruim por ser uma área bagunçada como todo bom* ***centro de cidade*** *brasileira.*" (It is well located, in the **center of the city**, which is good for those who are going to solve things nearby, but bad for being a messy area like all good **city center** in Brazil.). The first occurrence refers to the location aspect, and the second to the appearance of the center of Brazilian cities in general. The error is related to the classifier, and it would be difficult for the classifier to be able to differentiate the two expressions without understanding the context of each one.

The next case was the expression *"colchão da cama"*. Here the expression is formed by two distinct aspects, *"colchão"* and *"cama"*, which led the classifier to make the mistake of uniting both in a single aspect. The same happened with the expression *"serviço de café da manhã"*, where the term *"café da manhã"* was connected to the word *"serviço"* that is found in other aspects.

As possible dataset annotation errors, we have the aspects *"lavanderias"*, *"luxos"*, and *"tvs"*.

All of them were annotated only in their singular forms. The word *"lavanderia"* is an interesting case because the term appears related to two distinct aspects: location (hotel close to a laundry) and the hotel's laundry service. Given that the word in the singular form appears as an aspect, we believe that its respective plural form should also have been annotated, as it was the case with other aspects. It is worth mentioning that these three errors were made in the post-processing step. The classifier possibly did not make such errors due to the lower frequency of plural forms.

Table 5
Aspects not found.

Term aspect	QTY	Aspect	Category	Subcategory	Error
aquecimento (heating)	8	Facilities	Feature	Part-of	Classifier
calefação (calefaction)	2	Facilities	Feature	Part-of	Classifier
cortinas (curtains)	4	Facilities	Feature	Part-of	Classifier
escada (stairs)	3	Facilities	Feature	Part-of	Classifier
estabelecimento (establishment)	2	Facilities	Feature	Equivalence	Classifier
gerência (management)	1	Employees	Feature	Is-a	Classifier
iluminação (lighting)	4	Facilities	Feature	Part-of	Classifier
jardim (garden)	4	Facilities	Feature	Part-of	Classifier
motel (motel)	1	Facilities	Contextual	Related	Annotation
museus (museums)	3	Location	Contextual	Local	Classifier
parque (park)	4	Location	Contextual	Local	Classifier
porteiro (concierge)	1	Employees	Feature	Is-a	Classifier
tapete (carpet)	2	Facilities	Feature	Part-of	Classifier
tomada (socket)	2	Facilities	Feature	Part-of	Classifier
tomadas (sockets)	1	Facilities	Feature	Part-of	Classifier
torneira (faucet)	3	Facilities	Feature	Part-of	Classifier
torneiras (faucets)	1	Facilities	Feature	Part-of	Classifier
travesseiros (pillows)	5	Facilities	Feature	Part-of	Classifier
turistas (tourists)	8	Facilities	Contextual	Related	Classifier

Analyzing the table of aspects not found (Table 5), we can see that most cases were related to classifier errors. This is probably related to a smaller number of these aspects. A caveat for the aspect "*museus*": this aspect appears in its singular form, and probably should not have been detected due to validation in the 5 folds, appearing in a different set in its singular form, thus making its detection impossible.

As exceptions to the issue of the number of occurrences, we have the aspects "*aquecimento*" and "*turistas*". Regarding the "*aquecimento*" aspect, we have no hypothesis about the fact that it was not located, as it appears eight times in a similar context to other aspects related to the hotel's facilities, which in theory would not be an obstacle to its detection. Regarding the aspect '*turistas*', it is a more difficult aspect to detect since it is necessary to understand the context in which it appears. It is necessary to understand that, for some people, a hotel full of tourists is something negative, as in the example "*Muito cheio de turistas*" (Too full of tourists), as well as its positive absence "*já que não tem muitos turistas por lá*" (since there are not many tourists there). Regarding the facilities, it is necessary to know that a hotel that attracts tourists must have certain appealing characteristics, as in the sentence "*Local ideal para turistas*" (Ideal place for tourists).

One last aspect not found that deserves to be commented on is the aspect “motel”. In our view, it seems more like a case of an error in the annotation. The word appears ironic in the sentence “*Economize o seu dinheiro e vá para um motel*” (Save your money, and go to a motel), not dealing with a motel itself. Therefore, we believe that it would be necessary to mark the entire sentence to be able to understand the negative criticism presented to the hotel’s facilities.

4. Final remarks

Our experiments show that a good performance in the aspect extraction task may be achieved with a relatively simple strategy. Our error analysis reveals the difficulty in detecting implicit aspects, especially those that depend on additional information about the context in which they are inserted. In general, classifier errors were related to terms not annotated as aspects but very similar to other aspects found in the dataset. The aspects that were not found were mostly due to their low frequency in the dataset, which harms most of the machine learning-based strategies.

Acknowledgments

The authors of this work would like to thank the Center for Artificial Intelligence (C4AI-USP) and the support from the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and from the IBM Corporation.

References

[1] M. Taboada, Sentiment analysis: An overview from linguistics, Annual Review of Linguistics 2 (2016) 325–347.
[2] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5 (2012) 1–167.
[3] T. T. Thet, J.-C. Na, C. S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, Journal of information science 36 (2010) 823–848.
[4] J. Pavlopoulos, I. Androutsopoulos, 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, Association for Computational Linguistics, Gothenburg, Sweden, 2014, pp. 44–52.
[5] C. Long, J. Zhang, X. Zhu, A review selection approach for accurate feature rating estimation, in: Coling 2010: Posters, 2010, pp. 766–774.
[6] B. Snyder, R. Barzilay, Multiple aspect ranking using the good grief algorithm, in: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, 2007, pp. 300–307.
[7] J. Yu, Z.-J. Zha, M. Wang, T.-S. Chua, Aspect ranking: identifying important product aspects from online consumer reviews, in: Proceedings of the 49th annual meeting of

the association for computational linguistics: human language technologies, 2011, pp. 1496–1505.
[8] B. Pang, L. Lee, Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales, in: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005, pp. 115–124.
[9] M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, Data Mining and Knowledge Discovery 24 (2012) 478–514.
[10] Z. Zhai, B. Liu, H. Xu, P. Jia, Grouping product features using semi-supervised learning with soft-constraints, in: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), 2010, pp. 1272–1280.
[11] Z. Zhai, B. Liu, H. Xu, P. Jia, Clustering product features for opinion mining, in: Proceedings of the fourth ACM international conference on Web search and data mining, 2011, pp. 347–354.
[12] 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 (Eds.), Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022), CEUR Workshop Proceedings, CEUR-WS.org, 2022.
[13] M. T. Machado, T. A. S. Pardo, Evaluating methods for extraction of aspect terms in opinion texts in portuguese - the challenges of implicit aspects, in: Language Resources and Evaluation Conference, ELRA, 2022, pp. 3819–3828.
[14] L. A. de Freitas A, R. Vieira, Exploring resources for sentiment analysis in portuguese language, in: Proceedings of the 4th Brazilian Conference on Intelligent Systems, IEEE Computer Society, Natal, Brazil, 2015, pp. 152–156.
[15] M. T. Machado, T. A. S. Pardo, E. E. S. Ruiz, A. D. Felippo, F. Vargas, Implicit opinion aspect clues in portuguese texts: Analysis and categorization, in: Proceedings of the 15th International Conference on Computational Processing of the Portuguese, Springer, Fortaleza, Brazil, 2022, pp. 68–78.
[16] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, United States of America, 2004, pp. 168–177.
[17] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, volume 2, Curran Associates Inc., Lake Tahoe, United States of America, 2013, p. 3111–3119.
[18] M. T. Machado, T. A. S. Pardo, E. E. S. Ruiz, Analysis of unsupervised aspect term identification methods for portuguese reviews, in: Proceedings of the 14th National Meeting on Artificial and Computational Intelligence, Brazilian Computing Society, Uberlândia, Brazil, 2017, pp. 239–249.
[19] C. D. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, Cambridge University Press, 2008.
[20] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, Y. Singer, Online passive-aggressive algorithms, Journal of Machine Learning Research 7 (2006) 551–585.
[21] Y. Freund, R. E. Schapire, Large margin classification using the perceptron algorithm,

Machine learning 37 (1999) 277–296.
[22] L. Bottou, Stochastic gradient descent tricks, in: Neural networks: Tricks of the trade, Springer Berlin Heidelberg, 2012, pp. 421–436.
[23] H. Gandhi, V. Attar, Extracting aspect terms using crf and bi-lstm models, Procedia Computer Science 167 (2020) 2486–2495.
[24] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, 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, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.
[25] F. Souza, R. Nogueira, R. Lotufo, BERTimbau: pretrained BERT models for Brazilian Portuguese, in: Proceedings of the 9th Brazilian Conference on Intelligent Systems, Springer, 2020, pp. 403–417.
[26] L. A. de Freitas, Feature-level sentiment analysis applied to brazilian portuguese reviews, Ph.D. thesis, Pontifícia Universidade Católica do Rio Grande do Sul, 2015.
[27] U. B. Corrêa, Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa, Ph.D. thesis, Universidade Federal de Pelotas, 2021.
[28] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python, 2020.
[29] M. Korobov, sklearn-crfsuite, 2015. URL: https://github.com/TeamHG-Memex/sklearn-crfsuite.

A. Online Resources

More information and the source codes for the NILC participation are available via

- GitHub,
- POeTiSA, POrtuguese processing - Towards Syntactic Analysis and parsing.