



Pós-Graduação em Ciência da Computação

ARTHUR CAÍQUE BEZERRA VIEIRA

**EXTRAÇÃO DE TERMOS DE ASPECTOS PARA A MINERAÇÃO
DE OPINIÃO APLICADA À LÍNGUA PORTUGUESA: uma adaptação
do método Double Propagation**



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

Recife
2018

Arthur Caique Bezerra Vieira

**EXTRAÇÃO DE TERMOS DE ASPECTOS PARA A MINERAÇÃO DE OPINIÃO
APLICADA À LÍNGUA PORTUGUESA: uma adaptação do método Double
Propagation**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador: Ricardo Bastos Prudêncio Cavalcante

Coorientadora: Flavia de Almeida Barros

Recife
2018

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

V658e Vieira, Arthur Caíque Bezerra
Extração de termos de aspectos para a mineração de opinião aplicada à
língua portuguesa: uma adaptação do método *Double Propagation* / Arthur
Caíque Bezerra Vieira. – 2018.
70 f.: il., fig., tab.

Orientador: Ricardo Bastos Prudêncio Cavalcante.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn,
Ciência da Computação, Recife, 2018.
Inclui referências.

1. Inteligência computacional. 2. Mineração de opinião. I. Cavalcante,
Ricardo Bastos Prudêncio (orientador). II. Título.

006.3

CDD (23. ed.)

UFPE- MEI 2019-016

Arthur Caíque Bezerra Vieira

Extração de Termos de Aspectos para a Mineração de Opinião Aplicada à Língua Portuguesa: Uma Adaptação do Método Double Propagation

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 27/08/2018.

Orientador: Prof. Dr. Ricardo Bastos Cavalcanti Prudêncio

BANCA EXAMINADORA

Prof. Dr. Luciano de Andrade Barbosa
Centro de Informática / UFPE

Prof. Dr. Rafael Ferreira Leite de Mello
Departamento de Estatística e Informática / UFRPE

Profa. Dra. Flávia de Almeida Barros
Centro de Informática/UFPE

AGRADECIMENTOS

Primeiramente, gostaria de agradecer à minha mãe, por sempre ter me guiado por um bom caminho.

Gostaria de agradecer aos professores Ricardo Prudêncio e Flávia Barros, por todo o apoio e confiança.

Também gostaria de agradecer à Profa. Elyda Freitas, e por todo o apoio dado para que eu pudesse ingressar no mestrado.

Agradeço à Beatriz, por fazer os meus dias mais felizes.

Agradeço aos meus amigos, pois sem eles a vida não teria graça alguma.

E, por último, mas não menos importante, agradeço à FACEPE, pelo apoio financeiro.

RESUMO

A popularização de fóruns online e do e-commerce favoreceu o aumento do número de comentários/avaliações na Web sobre produtos e serviços, sendo impraticável analisar manualmente essa enorme quantidade de comentários. Para vencer esse desafio, lançamos mão de uma área da computação denominada de Mineração de Opinião (MO) – também conhecida como Análise de Sentimento. A MO busca extrair de textos em linguagem natural opiniões sobre entidades (produtos, serviços, pessoas, marcas, eventos, etc) e seus aspectos (características detalhadas de cada entidade - e.g., tamanho e peso de um celular). O objetivo final é classificar as opiniões extraídas entre positivas ou negativas, a fim de identificar como cada entidade e seus aspectos estão sendo avaliados pelos usuários. Este trabalho de mestrado investigou a MO a partir de resenhas em Português sobre produtos, com foco principal na primeira etapa da MO, isto é, na extração de termos referentes a entidades e seus aspectos. Devido à escassez de dados (comentários/resenhas) rotulados com informações de termos de aspecto na língua portuguesa, torna-se difícil a utilização de técnicas supervisionadas para a extração de termos de aspectos. Assim, este trabalho foi desenvolvido dentro da abordagem não supervisionada (baseadas em conhecimento - regras explícitas considerando informações estatísticas e/ou linguísticas, ontologias, etc), que tem obtido desempenho comparável à abordagem supervisionada pra a língua portuguesa. Este trabalho utilizou como ponto de partida um algoritmo originalmente construído para a língua inglesa que apresenta altas taxas de cobertura na tarefa de extração: o Double Propagation (DP). Esse algoritmo se baseia na classe gramatical das palavras nas frases, e em relações de dependência sintática entre as palavras nas frases para realizar a extração de termos de aspectos. Como contribuição principal deste trabalho, o algoritmo DP original foi adaptado para a língua portuguesa. A seguir, foram propostos mais tipos de relações de dependência sintática no processo de extração considerando mais classes gramaticais do que o algoritmo base. Como contribuição secundária, este trabalho também investigou a efetividade do uso de normalizadores de texto no processamento de resenhas oriundas da Web (que geralmente contêm erros ortográficos e gramaticais, e também erros de pontuação). Os experimentos realizados compararam diversas configurações do processo de extração, variando tanto o algoritmo de extração (o original e o ampliado com novas regras) quanto os recursos externos (e.g., normalizadores de texto, configurações dos métodos de poda/pruning e léxicos de sentimento). Os resultados obtidos foram bastante promissores, sendo comparáveis a outras abordagens já utilizadas em trabalhos com foco na língua portuguesa.

Palavras-chave: Mineração de opinião. Extração de termos de aspectos. Língua portuguesa. Double propagation.

ABSTRACT

The popularization of online forums and e-commerce benefited the increase of the number of comments/evaluations about products and services on the Web, making impractical to manually analyze the huge amount of comments. To overcome this challenge, we use an area of computing named Opinion Mining - also known as Sentiment Analysis. The Opinion Mining aims to extract opinions about entities (products, services, people, brands, events, etc) and their aspects (attributes of each entity - e.g., size and weight of a smartphone) from texts in natural language. The ultimate goal is to classify the extracted opinions between positive or negative, in order to identify how each entity and their aspects are evaluated by the users. This Master's thesis investigated the Opinion Mining from portuguese reviews of products, focusing in the first stage of Opinion Mining, which means, in the terms extraction of entities and their aspects. Given the lack of data (comments/reviews) labeled with information of aspect terms in portuguese, it becomes difficult to apply a supervised approach for the aspect term extraction of reviews written in brazilian portuguese. Thus, this work was developed in a unsupervised manner (based on knowledge - explicit rules taking into account statistics and/or linguistics information, ontologies, etc.), which have been getting comparable performance to the supervised approach for the portuguese language. In this work, we used as a starting point an algorithm originally constructed for the english language, which presented high recall on the task of aspect term extraction: the Double Propagation. This algorithm uses the Part-of-speech tagging and dependency relations of words in sentences to accomplish the extraction process. As the main contribution of this work, we adapted the original Double Propagation algorithm for the portuguese language. Next, we proposed the addition of more types of dependency relations in the process of extraction, and we considered more Part-of-Speech tags than the base algorithm. Another core contribution of this work was the investigation of the effectiveness of text normalizers in the processing of reviews from the Web (which, in its nature comes with spelling, grammatical and punctuation errors). The conducted experiments compared several configurations of the extraction process, varying the extraction algorithm (original and the proposed - with the additional rules), the external resources (text normalizers, sentiment lexicons and configuration of the pruning methods). The obtained results were very promising, being comparable to others approaches used in works with a focus on the portuguese language.

Keywords: Opinion mining. Aspect terms extraction. Portuguese language. Double propagation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura para ASNA com ontologias	33
Figura 2 – Processo de criação do <i>parser</i> sintático	46
Figura 3 – Processo de Extração de Termos de Aspectos	46
Figura 4 – Números de produtos por tipo de produto	51
Figura 5 – Números de comentários por tipo de produto	51
Figura 6 – Média do número de comentários por seção	52
Figura 7 – Exemplo do sistema auxiliar para anotação de termos de aspectos .	53
Figura 8 – Número de resenhas por produto no corpus de teste	53
Figura 9 – Número de termos de aspectos por produto no corpus de teste . . .	54

LISTA DE TABELAS

Tabela 1 – Dados dos Léxicos de Sentimento	23
Tabela 2 – Avaliação da Correlação dos Léxicos de Sentimento	23
Tabela 3 – Resumo dos trabalhos apresentados	31
Tabela 4 – Regras para extração de termos de aspecto e palavras opinativas - 1 (QIU et al., 2011)	39
Tabela 5 – Regras para extração de termos de aspecto e palavras opinativas - 2 (QIU et al., 2011)	40
Tabela 6 – Novas regras adicionadas ao Double Propagation	48
Tabela 7 – Resultados da Efetividade dos Léxicos de Sentimento no Processo de Extração	56
Tabela 8 – Avaliação dos normalizadores de texto	57
Tabela 9 – Poda Baseada no P-support dos Termos	57
Tabela 10 – Poda Baseada na Frequência de Termos No Corpus com Comentá- rios sobre todos os Artigos de um Produto	58
Tabela 11 – Resultados da Nova Regra Proposta Para o Processo de Extração .	58
Tabela 12 – Resultados obtidos para as resenhas de notebook sem normalização	59
Tabela 13 – Resultados obtidos para as resenhas de notebook com o UGC Normal	60
Tabela 14 – Resultados obtidos para as resenhas de Smartvts sem normalização	60
Tabela 15 – Resultados obtidos para as resenhas de Smartvts com o UGC Normal	61
Tabela 16 – Resultados obtidos para as resenhas de Smartphones sem normali- zador de textos	61
Tabela 17 – Resultados obtidos para as resenhas de Smartphones com o UGC Normal	62
Tabela 18 – Avaliação final	63

LISTA DE ABREVIATURAS E SIGLAS

AS	Análise de Sentimento
ASNA	Análise de Sentimento no Nível de Aspecto
BPUD	Brazilian Portuguese Universal Dependencies
DP	Double Propagation
ETA	Extração de Termos de Aspecto
MO	Mineração de Opinião
PD	Propagação Dupla
PLN	Processamento de Linguagem Natural
PT	Português
PT-BR	Português do Brasil

SUMÁRIO

1	INTRODUÇÃO	12
1.1	CONTEXTO GERAL	12
1.2	CONTEXTO ESPECÍFICO	13
1.3	CONTRIBUIÇÕES	13
1.4	ORGANIZAÇÃO DA DISSERTAÇÃO	15
2	ANÁLISE DE SENTIMENTO/MINERAÇÃO DE OPINIÃO	16
2.1	DEFINIÇÕES E CONCEITOS	16
2.2	ELEMENTOS DA ANÁLISE DE SENTIMENTO	16
2.2.1	Opinião e Sentimento	17
2.2.2	Níveis da Análise de Sentimento	19
2.2.3	Etapas da Análise de Sentimento	20
2.2.3.1	Classificação de Subjetividade	21
2.2.3.2	Classificação de Sentimentos	21
2.2.3.3	Sumarização	21
2.3	ANÁLISE DE SENTIMENTO NA LÍNGUA PORTUGUESA	21
2.3.1	Recursos Disponíveis para a Análise de Sentimento Aplicada à Língua Portuguesa	22
2.3.1.1	Léxicos de Sentimento	22
2.3.1.1.1	<i>Brazilian LIWC Dictionary</i>	24
2.3.1.1.2	<i>Opinion Lexicon</i>	24
2.3.1.1.3	<i>SentiLex-PT</i>	25
2.3.1.2	Corpora	25
2.3.2	Abordagens Baseadas em Léxicos de Sentimento	26
2.3.3	Abordagem híbrida	27
2.4	DIFICULDADES E OPORTUNIDADES	28
2.5	CONSIDERAÇÕES FINAIS	29
3	ANÁLISE DE SENTIMENTO NO NÍVEL DE ASPECTO	30
3.1	EXTRAÇÃO DE TERMOS DE ASPECTOS	31
3.1.1	Extração Supervisionada de Termos de Aspectos	32
3.1.2	Extração Não Supervisionada de Termos de Aspectos	32
3.1.2.1	Ontologias	33
3.1.2.2	Métodos Estatísticos	34
3.1.2.3	Relações Sintáticas	36
3.2	CLASSIFICAÇÃO DE SENTIMENTO NO NÍVEL DE ASPECTO	36
3.3	CONSIDERAÇÕES FINAIS	37

4	DOUBLE PROPAGATION PTBR	38
4.1	DOUBLE PROPAGATION	38
4.1.1	Dificuldades na Replicação do <i>Double Propagation</i>	44
4.2	DOUBLE PROPAGATION PTBR	45
4.2.1	Visão Geral	45
4.2.2	Criação do <i>Parser Sintático</i>	45
4.2.3	Processo de Extração de Termos de Aspectos	46
4.3	CONSIDERAÇÕES FINAIS	49
5	EXPERIMENTOS E RESULTADOS	50
5.1	CRIAÇÃO DO CORPUS DE TESTE	50
5.1.1	Etiquetagem do Corpus de Teste	52
5.2	EXPERIMENTOS E RESULTADOS	54
5.2.1	Métricas de Avaliação	54
5.2.2	Experimentos	55
5.2.2.1	Experimentos DP PTBR	55
5.2.2.1.1	<i>Avaliação da Efetividade dos Léxicos de Sentimento</i>	55
5.2.2.1.2	<i>Avaliação da Efetividade do Uso de Normalizadores de Texto</i>	56
5.2.2.1.3	<i>Avaliação da Efetividade dos Métodos de Poda</i>	57
5.2.2.1.4	<i>Avaliação da Nova Regra Proposta para a Extração de Termos</i>	58
5.2.2.2	Comparação com outras abordagens	59
5.2.3	Discussão dos resultados	62
5.3	CONSIDERAÇÕES FINAIS	64
6	CONCLUSÃO	65
6.1	PRINCIPAIS CONTRIBUIÇÕES	65
6.2	TRABALHOS FUTUROS	66
	REFERÊNCIAS	67

1 INTRODUÇÃO

Neste capítulo, será apresentado o contexto ao qual este trabalho de mestrado está inserido, uma breve descrição do trabalho realizado, e a organização do restante do trabalho.

1.1 CONTEXTO GERAL

Nos últimos anos, tem sido cada vez mais notável o crescimento do volume de conteúdo gerado por usuário (CGU)¹ na internet. Desse conteúdo, existe uma parcela relacionada a opiniões emitidas pelos usuários sobre algum produto adquirido por eles, ou algum evento do qual eles participaram, ou ainda sobre alguma pessoa ou organização, dentre outros temas.

Como esse volume de CGU tem aumentado significativamente, torna-se impossível para humanos examinarem, manualmente, todas as opiniões emitidas sobre alguma entidade, e ainda avaliarem se as opiniões emitidas são positivas ou negativas.

Esse tipo de informação pode ser útil para marcas e organizações saberem o que os seus clientes pensam em relação aos seus produtos e aos produtos de seus concorrentes. Usuários também se beneficiam dessas informações, nos processos individuais de tomada de decisão (por exemplo, que produto comprar, ir ou não a um evento, etc).

Diante da necessidade de processar automaticamente essa massa de conteúdo disponível na Web, pesquisadores e desenvolvedores lançam mão de métodos e técnicas oriundos da área da Computação denominada de Mineração de Opinião (também conhecida como Análise de Sentimento). Essa área de trabalho tem por objetivo extrair, a partir dos textos disponíveis, os trechos que contêm opiniões, classificando-as entre positivas ou negativas.

A Análise de Sentimento (AS) pode ser realizada em diferentes níveis de granularidade: no nível de documento, no qual considera-se o sentimento geral expressado em um texto. No nível de frase, que considera o sentimento expressado em cada frase. E, quando deseja-se saber o sentimento do usuário a respeito de cada aspecto mencionado no texto, a AS é realizada no nível de aspecto. Por exemplo, na AS no nível de aspecto, a partir da frase “A tela do iPhone é bem grande, mas a bateria descarrega rápido demais”, é desejável que se identifique o sentimento expressado em relação à tela e à bateria do iPhone, que são positivo e negativo, respectivamente.

Para realizar a AS no nível de aspecto (ASNA), é necessário determinar a princípio quais são os aspectos das entidades que estão sendo avaliadas. Quando não

¹ Do inglês, *User-Generated Content* (UGC).

se conhece, de antemão, os termos que representam os aspectos das entidades sob análise, é necessário realizar a tarefa de Extração de Termos de Aspectos (ETA).

1.2 CONTEXTO ESPECÍFICO

Este trabalho de mestrado foi desenvolvido dentro do terceiro nível de granularidade da AS - nível de aspecto, tendo como objetivo geral oferecer um processo automático para extração de aspectos a partir de textos de avaliações/comentários em português.

A motivação para desenvolver este trabalho originou-se de duas constatações:

(1) Na área da AS em geral, em relação à língua inglesa, pouco trabalho focado na língua portuguesa foi publicado. Isso se dá, principalmente, pela escassez de dados na língua portuguesa rotulados manualmente. Com isso, foi identificada uma lacuna na área, abrindo a oportunidade de se investigar a ETA através de uma abordagem não supervisionada aplicada à língua portuguesa.

(2) O crescimento da popularidade de sites de *e-commerce* foi um dos fatores que mais contribuíram para o aumento do volume de CGU, justamente por permitir que os seus usuários adicionem comentários avaliando os produtos que adquiriram.

Diante desse cenário, investigamos a ETA a partir de resenhas em português de produtos eletrônicos coletadas de um site de *e-commerce*. Na próxima seção é dada uma visão geral sobre o trabalho realizado.

1.3 CONTRIBUIÇÕES

As tarefas realizadas neste trabalho foram divididas nas atividades a seguir:

- **Estudo aprofundado do estado da arte nesse tema**

Esse estudo apontou para a lacuna existente em relação a trabalhos sobre ETA para a língua portuguesa. Além disso, o estudo revelou que o algoritmo Double Propagation (DP) apresenta altas taxas de cobertura para a língua inglesa. A partir daí, surgiu a ideia de se adaptar esse algoritmo para a Língua Portuguesa.

- **Implementação do algoritmo *Double Propagation***

Como a proposta deste trabalho de mestrado é a realização de uma adaptação do algoritmo *Double Propagation* para a língua portuguesa, a primeira tarefa realizada foi a implementação do algoritmo original.

- **Adaptação do algoritmo *Double Propagation* para a língua portuguesa**

Foi realizada uma adaptação do *Double Propagation* para a língua portuguesa, principal contribuição deste trabalho.

- **Investigação do uso de relações de dependências sintáticas para a ETA aplicada à língua portuguesa**

Visto que o *DP* realiza a ETA a partir de relações de dependência sintática, ficou como mais uma contribuição deste trabalho de mestrado a investigação do uso de relações de dependências sintáticas para a ETA aplicada à língua portuguesa.

- **Criação de um *parser* sintático**

Um *parser* sintático se faz necessário para a extração de relações de dependência das frases dos textos. Neste trabalho, foi implementado um *parser*, o qual será descrito na seção 4.2.2.

- **Criação de uma base de teste para a língua portuguesa**

Foi criada uma base de teste para avaliar o desempenho da adaptação proposta do *Double Propagation*, nomeado de DP PTBR (DP para o Português Brasileiro). Foi necessária a implementação de um *crawler* capaz de coletar comentários em um site de *e-commerce*. E, também, da etiquetagem manual dos termos de aspectos presentes nos comentários.

- **Extração de termos de aspectos a partir de expressões emotivas**

Como veremos ao longo deste trabalho, o *Double Propagation* original só é capaz de realizar a extração de termos a partir de expressões racionais, já a extensão desse algoritmo, proposta neste trabalho, também é capaz de extrair termos de aspectos a partir de expressões emotivas. Na seção 2.2, veremos a diferença entre expressões racionais e emotivas.

- **Investigação da efetividade da aplicação de um normalizador de texto ao processo de ETA**

Visto que a nossa base de teste é composta por textos advindos da *web*, que podem conter muitos ruídos, como erros de ortografia e o mal uso de pontuações, foi investigada a efetividade do uso de um normalizador de texto aplicado aos textos antes do processo de ETA.

- **Avaliação de diferentes léxicos de sentimento no processo de ETA**

Como será visto de maneira mais aprofundada na seção 4.1, o algoritmo DP utiliza um dicionário de sentimento, a partir do qual são obtidas palavras opinativas 'semente' usadas na extração dos aspectos. Neste trabalho, foram avaliados quatro léxicos de sentimento: SentiLex-PT Lemma, SentiLex-PT Flex, LIWC 2007 PT e OPLex.

- **Avaliação de diferentes métodos de poda**

Durante o processo de ETA, muitos ruídos podem ser extraídos como sendo aspectos. Dessa forma, se faz necessária a eliminação desses ruídos. Como será visto na seção 4.2, foram avaliados diferentes métodos de poda.

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

Quando à organização deste documento, no capítulo 2 são apresentadas as principais definições e conceitos que formalizam a AS e Mineração de Opinião, bem como também são apresentados os elementos que compõem a área, os níveis em que se pode realizar a AS, e as etapas para a sua realização. Também é apresentada a AS aplicada à língua portuguesa, os recursos disponíveis para a sua aplicação, bem como trabalhos que realizaram a AS aplicada a textos escritos na língua portuguesa.

Já o capítulo 3 é focado na Análise de Sentimento no Nível de Aspecto, apresentando trabalhos que realizam a ETA tanto de maneira supervisionada como não-supervisionada. Também é apresentada a classificação de sentimento no nível de aspecto.

A grande contribuição deste trabalho se encontra no capítulo 4, no qual é apresentado o algoritmo *Double Propagation* e suas propriedades, as dificuldades em implementá-lo, bem como são detalhadas a sua implementação, e adaptação para a língua portuguesa, o DP PTBR.

Por sua vez, no capítulo 5 é apresentada a criação da base de teste, a qual serviu para avaliar os algoritmos de ETA, desde a aquisição da base e seleção das resenhas até a rotulação manual dos termos de aspectos. Nesse capítulo, ainda são apresentados os experimentos realizados junto com os resultados obtidos. Comparamos o desempenho do DP PTBR, com o *Double Propagation* e com outras abordagens não supervisionadas baseadas em frequência no corpus e ontologias.

Por fim, no capítulo 6 concluímos este trabalho apresentando uma discussão final, bem como direções futuras a serem seguidas a partir deste trabalho.

2 ANÁLISE DE SENTIMENTO/MINERAÇÃO DE OPINIÃO

Neste capítulo, serão apresentadas as principais definições e conceitos que formalizam a Análise de Sentimento (AS), bem como os seus níveis, as suas abordagens, etapas, dificuldades e oportunidades.

2.1 DEFINIÇÕES E CONCEITOS

Segundo Bing Liu, a AS, também conhecida como Mineração de Opinião (MO), pode ser definida como o campo de estudo que analisa as opiniões, sentimentos, apreciações, atitudes e emoções das pessoas em relação a entidades e seus atributos expressados em texto escrito. Essas entidades podem ser produtos, serviços, pessoas, marcas, eventos, etc¹(LIU, 2015).

Enquanto Li e Hovy definem a AS como uma aplicação de Processamento de Linguagem Natural (PLN) que foca em identificar expressões que refletem a opinião dos autores baseada em seus posicionamentos (i.e., gosta ou não gosta) em relação a entidades e/ou suas propriedades (i.e., preço, qualidade)² (LI; HOVY, 2017).

Contudo, Liu (2015) defende que, ao invés de ser vista como uma subárea de PLN, a AS seja vista como uma pequena versão ou um caso especial do PLN, visto que, por ter apresentado muitos problemas de pesquisa desafiadores que não haviam sido investigados antes de sua concepção, ela expandiu significativamente a pesquisa em PLN. Liu ainda afirma que todo subproblema do PLN também é um subproblema da AS. O reconhecimento de entidades nomeadas e resolução de correferência podem ser citados como alguns casos de subproblema da AS que também é um subproblema do PLN.

Dessa forma, podemos definir a AS como uma tarefa ou área de pesquisa ligada ao PLN que busca identificar a opinião e o sentimento de pessoas em relação a entidades e seus atributos (aspectos).

Nas próximas seções, serão apresentados os elementos e os níveis da AS.

2.2 ELEMENTOS DA ANÁLISE DE SENTIMENTO

Para se ter uma visão mais ampla do que é a AS, é necessário conhecer os seus elementos. Dessa forma, nesta seção, serão apresentados os principais elementos que compõem e formalizam a AS. E, para auxiliar na compreensão desses elementos,

¹ Livre tradução do autor

² Livre tradução do autor.

utilizaremos o exemplo a seguir, o qual foi criado pelo autor deste trabalho, e que será identificado como **RES1** ao longo do texto:

“(1) Ontem, finalmente, comprei o iPhone X. (2) Apesar do preço ser muito alto, estou extremamente satisfeito, principalmente com a qualidade da câmera. (3) Quando a minha namorada viu o meu iPhone, ela quis comprar um na hora. (4) Ah, a capacidade de armazenamento é fenomenal, vai levar alguns anos até ficar sem memória disponível. (5) O único ponto negativo é a duração da bateria, duas horas de uso direto e já tive que colocar para recarregar. (6) Ah, sua tela é mais frágil que a do Moto G, mas ele é leve como uma pena. (7) Enfim, amei o meu iPhone X.”³

2.2.1 Opinião e Sentimento

Como vimos na seção 2.1, a área sob estudo é conhecida, principalmente, por Análise de Sentimento ou Mineração de Opinião. Logo, podemos dizer que sentimento e opinião são os principais objetos dessa área. No entanto, segundo Liu (2015), pela área ter surgido na Ciência da Computação, pouco esforço se deu para discutir o que é opinião, e o que é sentimento, e quando o uso de cada uma das duas nomenclaturas seria mais adequado. Então, para esclarecer um pouco dos dois conceitos, recorreremos às definições de opinião e sentimento dadas pela versão online do dicionário Michaelis (MICHAELIS, 2017).

Destacamos algumas definições para opinião segundo o dicionário citado, as quais podem ser vistas a seguir:

- 1) “Modo de pensar, de julgar, de ver”.
- 2) “Ponto de vista ou posição tomada sobre assunto em particular (social, político, religioso, etc.)”.
- 3) “Parecer emitido sobre determinado assunto em que muito se refletiu e deliberou”.
- 4) “Juízo de valor que se faz sobre alguém ou alguma coisa”.

Por sua vez, sentimento é definido como:

- 1) “Ato ou efeito de sentir(-se)”.
- 2) “Capacidade ou aptidão para sentir; disposição para ser facilmente comovido ou impressionado; sensibilidade”.

³ Resenha criada pelo autor, não se tratando de uma avaliação real sobre o produto.

- 3) “Faculdade de conhecer, apreciar, perceber; noção, senso”.
- 4) “Atitude moral ou mental que se caracteriza pelo estado afetivo”.
- 5) “Ligação afetiva e amorosa em relação a alguém ou algo; afeição, afeto, amor”.
- 6) “Percepção intuitiva; conhecimento imediato; intuição, pressentimento, suspeita”.
- 7) “Demonstração viva, animada; emoção, entusiasmo”.
- 8) “Estado afetivo de desprazer; mágoa, pesar, tristeza”.

Logo, utilizamos as definições dadas para o termo *opinião* para adotarmos o termo *Mineração de Opinião* como a tarefa de identificar opiniões em textos, enquanto que utilizamos as definições dadas para o termo *sentimento* para adotarmos o termo *Análise de Sentimento* como a tarefa de se classificar o sentimento presente em uma opinião.

Adicionalmente, Liu (2015) define a opinião como sendo uma tupla contendo dois componentes chave: (g, s) . No qual, g se refere ao alvo da opinião, que pode ser uma entidade (ou seus aspectos) sob a qual a opinião foi expressada, enquanto s é o sentimento expressado em relação ao alvo da opinião. Por exemplo, na frase 2 em **RES1**, na qual o autor diz “Apesar do preço ser muito alto”, a opinião seria a tupla (preço, negativo), onde “preço” seria referente a g , ou seja, o alvo da opinião, e o sentimento s seria negativo pois, ao avaliar como “*muito alto*”, o opinante dá uma opinião negativa em relação ao preço do *iPhone*.

Em alguns trabalhos, o sentimento pode ser chamado de orientação da opinião, polaridade ou orientação semântica. Além disso, o sentimento também pode ser uma avaliação numérica, como o sistema de 1 a 5 estrelas, onde 1 é o sentimento mais negativo e 5 o mais positivo sobre a entidade/aspecto sob análise.

Em relação à opinião, temos dois tipos comuns, que são: as opiniões implícitas e as opiniões explícitas, que são definidas a seguir:

- **Opinião explícita:** É quando o opinante deixa claro qual aspecto ele está avaliando. Por exemplo, na frase 5 em **RES1**, o autor fala que “o único ponto negativo é a duração da bateria”, logo fica claro sobre qual aspecto ele está emitindo uma opinião.
- **Opinião implícita:** Enquanto que, na opinião implícita, o autor avalia um aspecto sem mencioná-lo, que é o caso na frase 6 em **RES1**, que ele afirma que o iPhone é “leve como uma pena”. Implicitamente, com a expressão “leve como uma pena”, o opinante avalia o peso do iPhone, sem mencionar o termo “peso”.

Um caso especial de opinião a ser tratado na AS é a opinião comparativa. Para (LIU, 2015), a opinião comparativa é aquela que compara várias entidades que possuem aspectos semelhantes. Nesse caso, a tarefa é identificar quais entidades e aspectos dessas entidades estão sendo comparados, e o sentimento relativo a cada um desses aspectos. Na frase 6 de **RES1** temos um exemplo de opinião comparativa, onde o autor afirma que a tela do *iPhone* é “mais frágil que a do *Moto G*”. Ou seja, ele compara um aspecto comum (a tela) entre duas entidades distintas (*iPhone* e *Moto G*).

Em relação aos sentimentos, (LIU, 2015) utiliza uma abordagem baseada em pesquisa do consumidor para classificar os tipos de sentimento, que são: sentimento racional e sentimento emotivo.

- **Sentimento racional:** vem do raciocínio lógico, são crenças tangíveis. Por exemplo, quando em **RES1**, o autor afirma que o iPhone é “leve como uma pena”, ele está expressando um sentimento racional em relação ao peso do produto, que ele segurou, sentiu o seu peso para então formar uma opinião em relação ao mesmo.
- **Sentimento emotivo:** vem do intangível e de respostas emotivo, são opiniões que expressam emoções. Por exemplo, Eu amei o design do notebook; Odiei as teclas.

No trabalho de Qiu et al. (2011), o qual este trabalho de mestrado toma como base, os autores utilizam uma abordagem não supervisionada baseada em conhecimento - regras explícitas considerando informações linguísticas para a realização da tarefa de extração de termos de aspectos (ETA). Os termos são extraídos apenas a partir de expressões racionais, enquanto a abordagem deste trabalho estende a abordagem de Guanq Qiu et al. para extrair também aspectos mencionados a partir de expressões emotivas. Isso será detalhado no capítulo 4.

2.2.2 Níveis da Análise de Sentimento

A AS pode ser realizada em vários níveis de granularidade. Para (LIU, 2015), a AS pode ser executada em três níveis: nível de documento, nível de frase ou nível de aspecto.

- **Nível de documento:** Neste nível de análise, o objetivo é descobrir se o texto, documento ou resenha, por completo, expressa uma opinião positiva, negativa ou neutra sobre alguma entidade.
No caso da **RES1**, a entidade avaliada é o *iPhone X*. No geral, o autor expressa contentamento com o produto, ou seja, a saída de um sistema que

realiza a AS no nível de documento para RES1 deveria ser um sentimento positivo.

- **Nível de frase:** Na AS no nível de frase, o objetivo é classificar separadamente o sentimento geral de cada frase de um documento. Logo, em um sistema que realiza a AS no nível de frase, a saída para cada resenha deve ser o sentimento associado a cada uma de suas frases.

Na frase **1** em **RES1**, o autor não expressa nenhuma opinião em relação ao iPhone X. Enquanto, as frases **2** e **4** expressam principalmente opiniões positivas sobre a entidade avaliada. É válido observar que, na frase **2**, o autor expressa uma opinião negativa sobre o preço do produto, no entanto ele diz que está extremamente satisfeito com o produto, dessa forma, podemos inferir que o sentimento positivo é dominante, pois ele avalia com satisfação a entidade geral. Por sua vez, ainda que a frase **3** expresse um fato, podemos inferir que foi expressado um sentimento positivo por parte da namorada do autor. Já a frase **5** expressa uma opinião negativa. Logo, um sistema que realiza a classificação de sentimento no nível de frase deve apresentar um *score* de sentimento associado a cada uma das frases.

- **Nível de aspecto:** A AS nos níveis de documento e frase podem não ser suficientes para algumas aplicações como, por exemplo, quando se deseja saber exatamente qual opinião dos usuários sobre cada aspecto do produto individualmente. Dessa forma, a AS no nível de aspecto se faz necessária, pois seu objetivo é identificar/extrair quais aspectos estão sendo retratados em cada documento, e qual a opinião do autor sobre cada um desses aspectos.

Como observamos anteriormente, em **RES1**, as frases opinativas são as 2, 3, 4 e 5. Então, em um sistema que executa a Análise de Sentimento no Nível de Aspecto (ASNA), a saída em **2** deveria ser (*preço*, **NEGATIVO**), (*qualidade da câmera*, **POSITIVO**). Já em **3**, a saída seria (*iPhone*, **POSITIVO**). Enquanto que na frase **4**, deveria ser extraído (*capacidade de armazenamento*, **POSITIVO**). E, por fim, em **5**, a extração deveria trazer (*duração da bateria*, **NEGATIVO**).

2.2.3 Etapas da Análise de Sentimento

Para a realização da AS, é desejável a realização de algumas etapas de acordo com o nível da análise a ser realizada. A seguir, detalhamos as principais etapas, que são: **Classificação de subjetividade; Extração de Termos de Aspecto; Classificação de Sentimento; e Sumarização.**

2.2.3.1 Classificação de Subjetividade

Ravi e Ravi (2015) definem a classificação de subjetividade como o identificação da linguagem orientada a opinião. A detecção de subjetividade é uma tarefa a ser realizada quando se deseja executar a AS no nível de frase.

Liu (2015) defende que, por existir a possibilidade de um documento conter muitas frases que o opinante não expressa opinião, é necessário separar as frases que carregam subjetividade das que não carregam, ou seja, separar as frases subjetivas (opinativas) das frases objetivas (que não contém opinião).

Em **RES1**, um sistema de classificação de subjetividade deveria identificar a frase **1** como objetiva, ou seja, que não carrega opinião. Enquanto que deveria identificar todas as outras frases como subjetivas, ou seja, que carregam opinião.

(MORAES et al., 2016) realizaram um estudo comparativo entre métodos para classificação de subjetividade em *tweets* escritos na língua portuguesa. Foram comparadas abordagens baseadas em léxicos de sentimento e aprendizagem de máquina. Foram utilizados os léxicos *WordNetAffectBR* (PASQUALOTTI; VIEIRA,) e *SentiLex-PT* (MÁRIO J. SILVA et al., 2010), enquanto os algoritmos de aprendizagem de máquina utilizados foram o *NaïveBayes* e o *Sequential Minimal Optimization* (SMO). Por fim, os resultados mostraram que a abordagem utilizando o algoritmo SMO obteve um melhor desempenho.

2.2.3.2 Classificação de Sentimentos

A classificação do sentimento é a análise do sentimento propriamente dita. É a principal tarefa da AS, independente do nível que deseja-se realizar a análise, visto que descobrir o sentimento do opinante em relação a entidades e seus aspectos é o objetivo da área. A classificação de sentimento varia dependendo do nível que a AS está sendo realizada.

2.2.3.3 Sumarização

Após a realização da classificação de sentimento, é necessário exibir ao usuário em formato gráfico ou textual quais os sentimentos dos opinantes em relação às entidades e seus aspectos sob análise.

2.3 ANÁLISE DE SENTIMENTO NA LÍNGUA PORTUGUESA

Nesta seção, abordaremos a tarefa da AS aplicada à língua portuguesa. Primeiro, abordaremos os recursos disponíveis utilizados para realizar a AS, como: léxicos de

sentimento e *corpora*, e algoritmos. Depois, veremos trabalhos na área de AS aplicados à língua portuguesa. Os trabalhos que realizam a AS no nível de aspecto serão discutidos no Capítulo 3.

2.3.1 Recursos Disponíveis para a Análise de Sentimento Aplicada à Língua Portuguesa

Nesta seção, apresentaremos os recursos disponíveis para a realização da AS na língua portuguesa. Será possível observar que ainda existem poucos recursos disponíveis, e que isso pode se dar pelo fato de que a pesquisa dessa área no Brasil ainda estar no início (FREITAS; VIEIRA, 2015). Vargas e Pardo (2017) corroboram esse pensamento ao afirmar que em relação ao inglês, poucos trabalhos nessa área foram desenvolvidos com foco na língua portuguesa.

2.3.1.1 Léxicos de Sentimento

Com relação aos léxicos de sentimento, Carvalho e Silva (2015) afirmam que “em geral, a informação de sentimento descrita nestes recursos corresponde à orientação semântica ou polaridade das palavras ou expressões. Neste âmbito, os traços mais comumente utilizados são os de negativo, positivo e neutro”.

Os principais léxicos de sentimento para a língua portuguesa são: *Brazilian Portuguese LIWC 2007 Dictionary* (BALANGE FILHO; PARDO; ALUÍSIO, 2013), *Opinion Lexicon* (SOUZA et al., 2012) e *SentiLex-PT* (MÁRIO J. SILVA et al., 2010) (SILVA; CARVALHO; SARMENTO, 2012).

A tabela 1 traz algumas informações com respeito aos três léxicos acima citados. É válido mencionar que o SentiLex-PT possui dois dicionários, um com as formas flexionadas dos termos, e outra com suas respectivas formas lematizadas. Ainda que o *LIWC-PT* possua mais entradas que os outros dois léxicos, a maioria de suas entradas dizem respeito a outras categorias além das polaridades positiva e negativa. Tanto o *LIWC-PT* quanto os dicionários do *SentiLex-PT* possuem entradas que são atribuídas a mais de uma polaridade, ou seja, sistemas que utilizam esses dicionários necessitam de análise do contexto para que possam atribuir a polaridade correta ao termo ambíguo. É possível observar que em todos os dicionários, há uma maior representatividade de termos pertencentes à polaridade negativa. As próximas seções, trazem mais detalhes acerca dos léxicos citados, como: forma de construção, métodos para construção, domínio, língua, e trabalhos que os utilizaram.

Tabela 1 – Dados dos Léxicos de Sentimento

Léxico	Total de entradas	Entradas negativas	Entradas neutras	Entradas positivas	Entradas com mais de uma polaridade	Possui outras categorias	Idioma
LIWC-PT	127.149	14.613	0	12.376	502	Sim	PT-BR
OP-Lex	32.191	14.210	8.987	8.457	0	Não	PT-BR
SentiLex-PT (Lema)	7.014	4.424	776	1.358	456	Não	PT
SentiLex-PT (Flex)	82.347	43.640	2.676	9.236	26.795	Não	PT

Em (BALANGE FILHO; PARDO; ALUÍSIO, 2013), foi realizado um estudo para comparar a correlação ou concordância entre os três léxicos acima citados, os resultados estão presentes na tabela 2. Os autores normalizaram os léxicos de uma forma que os três ficassem no mesmo formato, a fim de se realizar uma avaliação justa. Por exemplo, como o *LIWC-PT* não tem a categoria neutra, a comparação levou em consideração apenas as palavras com valores de polaridade positiva ou negativa. Eles analisaram a concordância de pares entre os léxicos, ou seja, o número de entradas com valores de polaridade igual.

Tabela 2 – Avaliação da Correlação dos Léxicos de Sentimento

Concordância	LIWC-PT	Op-Lex	SentiLex-PT
LIWC-PT	X	80,17% (1871 entradas)	74,83% (7310 entradas)
Op-Lex	X	X	97,04% (13.880 entradas)
SentiLex-PT	X	X	X

Como podemos ver na tabela 2, os léxicos OP-Lex e Senti são os que tem o maior grau de concordância entre as suas entradas, enquanto LIWC-PT e Senti são os que apresentam o menor grau de concordância.

2.3.1.1.1 Brazilian LIWC Dictionary

Baseados em estudos que evidenciavam que a saúde física e mental poderiam ser previstas pelas palavras que as pessoas utilizam, sendo o que elas falam e escrevem fornecedores de seus estados emocionais e cognitivos, Pennebaker, Francis e Booth (2001) buscaram fornecer um método eficiente para estudar os vários componentes emocionais, cognitivos e estruturais presentes na fala e escrita de pessoas. Dessa forma, eles criaram o *Linguistic Inquiry and Word Count (LIWC)*.

O *Brazilian LIWC Dictionary (LIWC-PT)* é uma adaptação da versão de 2007 do *English LIWC Dictionary* (PENNEBAKER; FRANCIS; BOOTH, 2001), (PENNEBAKER et al., 2007) para a língua portuguesa (BALANGE FILHO; PARDO; ALUÍSIO, 2013).

Essa adaptação foi realizada por três equipes via tradução utilizando vários dicionários bilíngues Português-Inglês (LIWC, 2017). A cada uma das entradas do LIWC-PT pode ser atribuída uma ou mais categorias, dentre elas as categorias *posemo* (*positive emotion*) e *negemo* (*negative emotion*), cada uma constituindo de 12.376 e 14.613 entradas, respectivamente. Porém, dependendo da aplicação, algumas outras entradas categorizadas como *affect* e *anger*, por exemplo, também podem ser úteis para a AS.

Em (BALANGE FILHO; PARDO; ALUÍSIO, 2013), o *LIWC-PT* foi comparado aos léxicos *OpLex* e *SentiLex* na tarefa de classificação de sentimentos nos níveis de frase e aspecto, obtendo melhores resultados na classificação de sentimento no nível de frase.

2.3.1.1.2 Opinion Lexicon

O *Opinion Lexicon (OP-Lex)* foi construído a partir da combinação de três métodos geradores de léxicos: baseado em corpus (**M1**), baseado em tesouro (**M2**) e baseado em tradução (**M3**). Cada um desses métodos gerou um léxico, e o resultado final foi a combinação dos três.

Para a extração de expressões de sentimento em **M1**, foi utilizado um corpus que consistia em 346 resenhas de filmes extraídos dos sites CinePlayers⁴ e Cinema com Rapadura⁵, além de 970 resenhas jornalísticas sobre temas diferentes extraídos do corpus PLN-BR CATEG (BRUCKSCHEN et al., 2008). E, para a atribuição de polaridade às expressões de sentimento, foi utilizado o método **PMI** (TURNERY, 2002).

Em **M2**, os autores utilizaram uma função baseada em distância para medir a distância de expressões de sentimento em relação a dois conjuntos sementes, um de

⁴ <http://www.cineplayers.com/>

⁵ <http://cinemacomrapadura.com.br/>

palavras positivas e outro de palavras negativas. A polaridade atribuída ao termo é aquela da qual o termo apresentou a menor distância de seu conjunto de palavras semente.

Em **M3**, os autores utilizaram o *Google Translate*⁶ para traduzir as entradas do léxico inglês de Liu (HU; LIU, 2004) para a língua portuguesa, as entradas que não puderam ser traduzidas pela ferramenta foram descartadas manualmente.

O *OP-Lex* foi utilizado nas tarefas de classificação de sentimento no nível de aspecto em (CHAVES et al., 2012) e (FREITAS; VIEIRA, 2013a).

2.3.1.1.3 *SentiLex-PT*

Atualmente na versão 02, o *SentiLex-PT* é definido por Carvalho e Silva (2015) como “um léxico de sentimento especificamente concebido para a análise de sentimento e opinião sobre entidades humanas em textos redigidos em português”.

O *SentiLex-PT* foi utilizado para realizar a tarefa de classificação de subjetividade em (MORAES et al., 2016), obtendo um desempenho um pouco inferior às abordagens baseadas em aprendizagem de máquina.

Em (AVANÇO; BRUM; NUNES, 2016), o Senti obteve um desempenho ao classificar resenhas de produto no nível de documento.

O *SentiLex-PT* foi utilizado na tarefa de classificação de sentimento no nível de aspecto no setor de acomodações em (FREITAS; VIEIRA, 2015). Em (FREITAS; VIEIRA, 2013a), ele obteve um desempenho melhor que o *Op-Lex* na tarefa de classificação de sentimento no nível de aspecto de resenhas de filmes.

2.3.1.2 Corpora

Uma das maiores dificuldades em se trabalhar com AS na língua portuguesa é a escassez de *corpora* anotados disponíveis para a avaliação de algoritmos e ferramentas. Esse é um fato reportado por diversos autores da literatura em vários trabalhos, como (BALANGE FILHO; PARDO; ALUÍSIO, 2013), (FREITAS, 2015), (SOUZA et al., 2016) e (BALAGE FILHO, 2017). Balange Filho, Pardo e Aluísio (2013), por exemplo, afirmam que existem poucos *corpora* disponíveis para a classificação de texto na língua portuguesa. Adicionalmente, Souza et al. (2016) afirmam que mesmo quando bons resultados são alcançados, raramente a base de dados utilizada é disponibilizada.

Para a AS na língua portuguesa, temos os *corpus*: ReLi e Priberam, os quais são descritos a seguir.

⁶ <https://translate.google.com/>

O *corpus* ReLi consiste de 1600 resenhas atribuídas a 13 livros de 7 autores diferentes. Os anotadores desse *corpus* relatam em (FREITAS et al., 2012) que a escolha pela variação de autores e livros para a construção dessa base reflete na variação da linguagem utilizada nas resenhas dos livros, eles dizem “há desde registros altamente informais e com grande uso de gírias, abreviações, neologismos e emoticons, até resenhas mais formais, com um vocabulário mais rebuscado”. Por fim, esse corpus é anotado com orientação semântica em dois níveis: de frase, e de sintagma nominal. No nível de frase, apenas uma polaridade é atribuída à frase, enquanto na anotação no nível de sintagma nominal, é possível identificar o que exatamente o autor da resenha está avaliando, para a AS essa é uma anotação no nível de aspecto.

O *corpus* opinativo *priberam* é composto por 80 resenhas no domínio de notícias extraídas do *corpus* sumarização *priberam* (ALMEIDA et al., 2015).

2.3.2 Abordagens Baseadas em Léxicos de Sentimento

Balange Filho, Pardo e Aluísio (2013) realizaram um estudo para medir o impacto do *LIWC-PT* numa tarefa de classificação de sentimento em textos na língua portuguesa do Brasil. Para a classificação de sentimento, foi utilizada uma adaptação do algoritmo *SO-CAL* (TABOADA et al., 2011), e o corpus utilizado foi o ReLi (FREITAS et al., 2012), o qual é descrito na seção 2.3.1.2. A classificação de sentimento foi realizada em dois níveis: frase e aspecto. Na análise no nível de frase, o algoritmo obteve um desempenho melhor quando foi utilizado o *LIWC-PT*, enquanto que no nível de aspecto o uso do *SentiLex* fez com que o algoritmo alcançasse um desempenho melhor.

Os autores também realizaram uma avaliação da classificação de cada classe (positiva, negativa) individualmente, e tanto no nível de aspecto quanto no nível de frase, o dicionário *LIWC-PT* teve dificuldade em classificar a classe negativa, com os autores afirmando que “o dicionário *LIWC* funciona melhor indicando positividade do que negatividade”.

O trabalho de Avanço e Nunes (2014) também utilizou os léxicos *LIWC-PT*, *Senti* e *Op-Lex* na tarefa de classificação de sentimento. Nesse trabalho, é relatado o desempenho de 3 versões diferentes de um classificador baseado em léxico. A primeira versão leva em consideração apenas a polaridade a priori das palavras de sentimento. A segunda versão adiciona contextos de negação à primeira, enquanto a terceira adiciona contextos de intensificação em relação às duas primeiras. Para a avaliação dos classificadores, foram selecionadas 2000 resenhas sobre celulares do site Buscapé⁷, sendo 1000 com opiniões positivas e 1000 com opiniões negativas em relação aos

⁷ <http://www.buscape.com.br/>

produtos nelas avaliados. Esses rótulos positivos e negativos foram dados com base na recomendação (ou não) do usuário em relação ao produto avaliado. A fim de verificar se seria confiável depender da recomendação do usuário em relação ao produto para rotular as resenhas, dois autores independentes realizaram manualmente a classificação desses comentários. Com um grau de concordância de 80.6% entre os dois autores, as resenhas ficaram classificadas como: 1085 positivas, 748 negativas, 71 neutras e 96 contendo tanto características positivas como negativas. Foi avaliada cada combinação: versão do classificador + léxico. A versão 3 do classificador + SentiLex foi a que obteve melhores resultados.

(CIRQUEIRA et al., 2017) realizaram uma comparação entre versões na língua portuguesa e inglesa de 4 métodos de classificação de sentimento baseados em léxico, três já existentes nas duas línguas, e a adaptação de um método da língua inglesa para a portuguesa. O método adaptado foi o *Vader*, enquanto os existentes foram o *SO-CAL*, *SentiStrength* e o *OpinionLexicon*. Foi utilizada uma base de dados construída a partir de comentários de páginas de instituições públicas no Facebook, e de *tweets* que mencionaram algumas dessas instituições. Como a base de dados foi constituída apenas de textos em português, foi criada outra versão dessa base em inglês através de tradução automática, a fim de se realizar a comparação. Os experimentos mostraram que as versões construídas para a língua portuguesa tiveram um desempenho bem abaixo das versões dos algoritmos para a língua inglesa. Os autores concluíram que os métodos em língua portuguesa ainda não têm maturidade suficiente para classificar textos oriundos de redes sociais, visto que as versões em inglês tenderam a ter um melhor desempenho, mesmo com os textos na língua inglesa tendo sido adquiridos via tradução automática.

2.3.3 Abordagem híbrida

(AVANÇO; BRUM; NUNES, 2016) trabalham com a ideia de combinar os melhores classificadores de cada abordagem (baseada em léxico e baseada em AM) para a classificação de polaridade no nível de documento. Para a abordagem baseada em léxico, três métodos foram examinados: (1) *Baseline*, no qual a classificação é dada pela soma da orientação semântica de todas as palavras de sentimento de um léxico presentes no texto; (2) *Baseline-VSM*, que é similar ao método (1), no entanto nenhum léxico de sentimento é necessário, pois a polaridade das palavras é calculada através da similaridade com palavras semente através de vetores de palavras; e (3) *LBC-VSM*, que é uma adaptação do classificador *LBC* de (AVANÇO; NUNES, 2014), porém a polaridade das palavras de sentimento são atribuídas através da similaridade com outras palavras semente, e seu valor pode mudar se ela está sob influência de negação ou de intensificadores.

Já na abordagem baseada em AM, os autores realizaram experimentos com os algoritmos Naive-Bayes e Support Vector Machines, os quais eles chamaram de C-NB e C-SVM, respectivamente. Os atributos utilizados para o treinamento dos classificadores de AM foram a presença/ausência de termos (*bag-of-words*) nos documentos.

Para os experimentos e avaliações dos classificadores das duas abordagens, os autores utilizaram resenhas de produtos eletrônicos do Buscapé e Mercado Livre, e apenas nos classificadores baseados em léxico, os autores utilizaram também o corpus ReLi (FREITAS et al., 2012). Os autores relatam também que “experimentos foram realizados sobre todos os léxicos de sentimento disponíveis para a língua portuguesa. E os melhores resultados foram obtidos utilizando o SentiLex”. Porém, eles não relatam quais foram exatamente os outros léxicos utilizados, e o desempenho de cada um separadamente. Por fim, o método baseado em léxico com o melhor desempenho foi o *LBC-SVM*, enquanto que melhor o baseado em AM foi o *C-SVM*. A abordagem híbrida consistia em que cada documento fosse classificado pelo *C-SVM*, e caso o resultado não fosse muito distante de um limiar, então o documento era classificado pelo *C-SVM*.

2.4 DIFICULDADES E OPORTUNIDADES

Por ser uma área relativamente nova, ainda que a pesquisa em AS tenha evoluído muito nos últimos, há ainda muitas dificuldades a serem superadas, principalmente quando se trata de da aplicação da AS/MO em textos na língua portuguesa. No entanto, essas dificuldades trazem novas oportunidades para investigação e pesquisa. Esta seção tem o objetivo de levantar algumas dificuldades e oportunidades da pesquisa e aplicação da AS.

Uma das principais dificuldades da AS é a obtenção de dados, principalmente rotulados, para a avaliação de algoritmos e aplicações. Quando se trata de AS para o português brasileiro, a escassez desses dados disponíveis publicamente é ainda maior. Então, a estratégia para a obtenção de dados é através de solicitação de bases de dados publicados em outros trabalhos, ou coletar uma base de dados na *web* e manualmente etiquetá-la, que é uma tarefa árdua, e que consome muito tempo.

Dessa forma, a criação de bases de teste rotuladas manualmente por vários anotadores,

Spamming em resenhas online é um problema apontado por (LIU, 2015), sobre o qual ele afirma que pessoas com intenções maliciosas criam perfis para inserir opiniões falsas para promover ou denegrir produtos e serviços. Dessa forma, surge a oportunidade de se realizar pesquisas para a identificação de perfis que tem apenas como objetivo postar *spams* em redes sociais e sites de *e-commerces*.

2.5 CONSIDERAÇÕES FINAIS

Neste capítulo, tivemos uma visão geral do que é a AS e a MO, os elementos que a compõem, os seus níveis de aplicação, e as etapas a serem realizadas para a sua aplicação.

Vimos alguns trabalhos que aplicam a AS na língua portuguesa, bem como os recursos disponíveis para a mesma. Assim como dificuldades e oportunidades de pesquisa.

No próximo capítulo, será apresentada a AS no nível de aspecto.

3 ANÁLISE DE SENTIMENTO NO NÍVEL DE ASPECTO

Como definido em (LIU, 2015), a AS no nível de documento ou frase não é adequada quando deseja-se saber exatamente qual atributo (aspecto) de uma entidade (pessoa, produto ou evento, etc) está sendo avaliado em um texto. Então, quando é desejável realizar uma análise mais sofisticada, é necessário descobrir quais entidades e seus aspectos estão sendo avaliados, e qual o sentimento do opinante em relação a esses aspectos e entidades. Essa tarefa é denominada análise de sentimento no nível de aspecto (ASNA).

As principais tarefas para a realização da ASNA são: a extração dos termos que representam os aspectos avaliados, o agrupamento de termos de aspectos, a classificação do sentimento expressado em relação a esses aspectos e a sumarização das opiniões. Visto que o objetivo deste trabalho é aplicar a ETA na língua portuguesa, neste capítulo daremos mais ênfase aos trabalhos que aplicam as tarefas relacionadas à ASNA na língua portuguesa.

A tabela 3 apresenta os trabalhos que são discutidos nas seções a seguir.

Tabela 3 – Resumo dos trabalhos apresentados

Algoritmo	Língua	Domínio
Hu e Liu (2004)	Inglês	Resenhas de produtos eletrônicos
Double Propagation (QIU et al., 2009; QIU et al., 2011)	Inglês	Resenhas de produtos eletrônicos
WhatMatter (SIQUEIRA; BARROS, 2010)	Português	Resenhas de serviços
PIRPO (CHAVES et al., 2012)	Português	Resenhas de hotéis
(FREITAS; VIEIRA, 2013a)	Português	Resenhas de filmes
(FREITAS; VIEIRA, 2013b)	Português	Resenhas de filmes e hotéis
(FREITAS, 2015)	Português	Resenhas de hotéis
(FREITAS, 2015)	Português	Resenhas de hotéis
Sápiras (2015)	Português	Notícias
CRF + Informações Sintáticas (BALAGE FILHO, 2017)	Português	Resenhas de livros
CRF-L (SHU; XU; LIU, 2017)	Inglês	Resenhas de produtos eletrônicos

Criada pelo autor

3.1 EXTRAÇÃO DE TERMOS DE ASPECTOS

A extração de termos de aspectos (ETA) nada mais é do que a identificação dos termos que representam aspectos de entidades avaliadas em textos, e é a tarefa mais difícil da ASNA. Ravi e Ravi (2015) citam a dificuldade de extrair esses termos por causa dos ruídos provenientes dos dados adquiridos em diferentes domínios.

Para a ETA, existem as abordagens supervisionadas e as não supervisionadas. Ao longo desta seção, discutiremos sobre cada uma delas, abordando tanto trabalhos aplicados à língua inglesa como, principalmente, aqueles aplicados à língua portuguesa.

Nesta seção, discutimos exclusivamente sobre a ETA, mesmo que os trabalhos citados realizem outras tarefas, como a classificação do sentimento ou sumarização das opiniões, pois essas serão abordadas nas outras seções deste capítulo.

3.1.1 Extração Supervisionada de Termos de Aspectos

A extração supervisionada de termos de aspecto é realizada pela aplicação de algoritmos supervisionados de aprendizagem de máquina. Liu (2015) afirma que os principais algoritmos para essa tarefa são os modelos de aprendizado sequencial, como o *Conditional Random Fields (CRF)* (LAFFERTY; MCCALLUM; PEREIRA, 2001) e o *Hidden Markov Models (HMM)*. Para a aplicação dessas técnicas, faz-se necessário um conjunto de dados rotulados, os quais são palavras e termos que são rotulados como aspectos ou não aspectos. Esses dados são, então, utilizados para treinar modelos de aprendizagem de máquina para que possam identificar padrões a serem reconhecidos em dados não rotulados.

Balage Filho (2017) aplicou o *Conditional Random Fields (CRF)* para a ETA no corpus ReLi, avaliando o impacto de um conjunto de características (*features*) sintáticas e semânticas no treinamento do modelo. O autor alcançou 0.35 de *f-score* como o melhor resultado ao aplicar informações semânticas do *parser* PALAVRAS aos dados. Ele observou ainda que *features* semânticas como *clusters* de palavras e *word2vec* não causaram nenhum impacto na ETA. Além disso, concluiu que o uso de características sintáticas não era promissor para esta tarefa.

Shu, Xu e Liu (2017) propõem o *CRF-L*, que é o CRF aplicado ao *framework* de aprendizagem de máquina perpétua (*Lifelong Machine Learning*). A premissa do *framework* é fazer com que um modelo de aprendizagem de máquina, mesmo depois de treinado, possa acumular, além do conhecimento dos dados rotulados no treinamento em um domínio, o conhecimento de novos dados em diferentes domínios aos quais ele for apresentado, diminuindo assim, o esforço para a rotulação de dados para cada novo domínio.

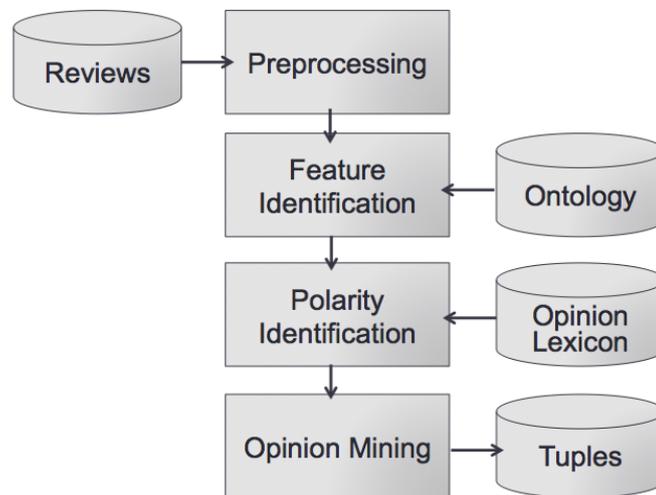
3.1.2 Extração Não Supervisionada de Termos de Aspectos

Geralmente, trabalhos que utilizam uma abordagem não supervisionada para a ETA constituem-se de regras explícitas baseadas em conhecimento (SIQUEIRA, 2010), (FREITAS; VIEIRA, 2013b) e (FREITAS, 2015). Em alguns casos, necessitam de recursos externos como ontologias e/ou léxicos de sentimento. As principais técnicas que fazem parte dessa abordagem são baseadas em: ontologias, frequência dos termos no corpus, e relação alvo da opinião com palavra opinativa. Essas abordagens serão detalhadas na próximas seções.

3.1.2.1 Ontologias

Geralmente, as abordagens para a realização da ASNA baseadas em ontologias seguem o fluxo da figura 1. Primeiro, realiza-se um pré-processamento dos textos, etapa que vai da segmentação de frases e *tokenização* ao *pos-tagging*. Depois, realiza-se a ETA, tarefa na qual as palavras e expressões do texto são buscadas em uma ontologia com informações de aspectos para então extraí-los. Após a identificação dos termos de aspectos, é utilizada uma combinação de léxico de sentimento e regras para a identificação do sentimento do opinante em relação aos aspectos identificados. Por fim, tuplas contendo os termos de aspectos e os sentimentos associados a eles são retornados ao usuário do sistema.

Figura 1 – Arquitetura para ASNA com ontologias



FREITAS;VIEIRA (2013)

Chaves et al. (2012) propuseram o PIRPO, um algoritmo independente de domínio, dando início a estudos para ASNA aplicada à língua portuguesa utilizando ontologias e léxicos de sentimento. Eles aplicaram o algoritmo em resenhas de hotéis do TripAdvisor¹ e Booking². Para a identificação de aspectos, foi utilizada a *HOntology* (CHAVES; TROJAHN, 2010), uma ontologia multilíngue para o domínio de hotéis.

Na etapa de pré processamento do PIRPO, é realizada a segmentação de frases e a *tokenização*. A próxima tarefa a ser realizada é ETA, na qual é utilizada a *HOntology*. Nessa fase do PIRPO, os conceitos da ontologia identificados nos textos são os termos de aspectos a serem extraídos.

Freitas e Vieira (2013a) utilizaram uma abordagem parecida com a de Chaves et al. (2012), porém aplicada ao domínio de filmes. Mais tarde, em (FREITAS;

¹ www.tripadvisor.com

² www.booking.com

VIEIRA, 2013b), mais uma vez as autoras combinaram o uso de ontologias e léxicos de sentimento para a ASNA, porém, dessa vez, também aplicada ao domínio de hotéis.

Por fim, (FREITAS, 2015) é uma extensão de (FREITAS; VIEIRA, 2015), e se propõe a identificar também aspectos implícitos.

3.1.2.2 Métodos Estatísticos

Os métodos baseados em frequência de termos para a extração de aspectos são definidos por Balage Filho (2017) como os mais simples para a realização dessa tarefa, servindo de *baseline* para outros métodos mais sofisticados.

Geralmente, esses métodos consideram como aspectos os substantivos e sintagmas nominais mais frequentes em um *corpus* (LIU, 2015). Segundo Balage Filho (2017), essa premissa é baseada no fato de que os principais aspectos de um produto são os mais comentados pelos opinantes das avaliações.

Hu e Liu (2004) realizaram a extração baseada nos 1% de substantivos e sintagmas nominais mais frequentes em um *corpus* de avaliações de produtos. Ao concluir que apenas os substantivos mais frequentes não garantiam uma boa cobertura, eles realizaram a identificação do que seriam aspectos não frequentes a partir da proximidade de substantivos não frequentes com adjetivos. No experimento, o algoritmo foi avaliado em resenhas de 5 produtos eletrônicos.

Em (SIQUEIRA, 2010), é apresentado o *WhatMatter*, um sistema que tem o objetivo de extrair, classificar o sentimento e sumarizar os aspectos avaliados em comentários sobre serviços de *e-commerce*. Através de um *crawler*, o autor coletou 2200 comentários na língua portuguesa sobre avaliações de serviços de *e-commerce* no Ebit³, sendo que 2000 foram utilizados para a construção do sistema, e os outros 200 para a realização dos experimentos.

Na tarefa de extração de aspectos, 4 etapas foram executadas: **(1)** Identificação de substantivos frequentes; **(2)** Identificação de substantivos relevantes; **(3)** Mapeamento de indicadores de aspecto; e **(4)** Remoção de substantivos não relacionados.

Na etapa **(1)**, o *WhatMatter* extrai os 3% substantivos mais frequentes como candidatos a aspectos. Esse limiar foi determinado empiricamente com base no conhecimento do autor em relação ao *corpus* adquirido. No entanto, Siqueira (2010) chegou à conclusão de que a lista de substantivos mais frequentes nem sempre cobre todos os aspectos. Então, o autor utilizou uma abordagem parecida com a de (HU; LIU, 2004) para identificar aspectos não frequentes na etapa **2** do *WhatMatter*. Porém, enquanto Hu e Liu (2004) identificaram os substantivos mais próximos aos adjetivos em

³ <https://www.ebit.com.br/>

uma frase, sendo ou não adjacentes, o trabalho de Siqueira (2010) considera apenas os substantivos adjacentes aos adjetivos, ou apenas separados por *stopwords*.

Na etapa (3) do *WhatMatter*, uma lista de 20 indicadores de aspectos, que compreende tanto advérbios como adjetivos, é utilizada para identificar aspectos implícitos. Por exemplo, o adjetivo *grande* indica o aspecto *tamanho*. No entanto, ele alerta que esse mapeamento requer cuidado extra, pois vários adjetivos podem ser muito versáteis, e seu significado pode ser dependente de domínio.

Em (BALAGE FILHO, 2017), é realizada uma avaliação acerca dessa tarefa utilizando o *corpus* ReLi, e também o *corpora* do SemEval 2015 e SemEval 2016, sendo o primeiro na língua portuguesa e os dois últimos na língua inglesa. O objetivo foi avaliar a efetividade dos métodos que utilizam essa abordagem tanto para a língua inglesa como para a portuguesa. Cada *corpus* foi dividido em dois conjuntos, um de treino e outro de teste.

Foram utilizadas 4 variações desse método. A primeira (1) identifica todos os aspectos do *corpus* de treinamento no *corpus* de teste. A segunda (2) é uma variação da (1), porém com a remoção de *stopwords*. Na terceira variação (3), foram observados limiares de corte em relacionados à frequência relativa entre os aspectos mencionados no conjunto de treinamento. Após a investigação, Balage concluiu que o limiar em 7% alcançou o melhor resultado para o *corpus* ReLi, e nenhum corte para o *corpora* do SemEval 2015 e SemEval 2016. O último método (4) é uma variação do (3), no entanto, aqui o limiar de corte é em relação à frequência relativa entre todos os termos do *corpus* de treinamento, sendo identificado como 67% o limiar para o ReLi, 30% para o SemEval 2015 e 45% para o SemEval 2016.

Balage Filho (2017) conclui que métodos baseados em frequência podem oferecer um bom ponto de partida, servindo de *baseline* para métodos mais sofisticados.

Em (SÁPIRAS, 2015), foi realizado um estudo de caso acerca da mineração de opinião no nível de aspecto em fontes de opinião fracamente estruturadas. Sápiras (2015) utilizou, como fonte de opinião, notícias e comentários na Folha de São Paulo sobre os três candidatos mais populares nas eleições de 2012 para a Prefeitura de São Paulo. Esse estudo realizou a identificação de termos de aspectos relacionados a planos eleitorais desses candidatos, assim como as percepções dos usuários em relação a esses aspectos. Foi avaliada a efetividade de três métodos estatísticos para a extração de termos relacionados a dois aspectos: saúde e educação. Os métodos avaliados foram: *Expected Mutual Information Measure* (EMIM), *Phi-squared* e Indexação Semântica Latente. Para o aspecto *saúde*, o *phi-squared* obteve o melhor resultado, enquanto que para o aspecto *educação*, o EMIM obteve um resultado superior.

3.1.2.3 Relações Sintáticas

Qiu et al. (QIU et al., 2009), (QIU et al., 2011) propuseram o algoritmo *Double Propagation* (DP), que é um algoritmo do tipo *bootstrapping* que foi desenvolvido com o propósito de simultaneamente extrair termos de aspectos e expandir léxicos de sentimento.

Um processo de *bootstrapping* é um processo auto-sustentável, no qual, a partir do momento que foi inicializado o algoritmo, ele não precisará mais de intervenção externa para funcionar ou para chamar uma outra função para continuar ou parar a sua execução. Para iniciar o seu processo de *bootstrapping*, o DP necessita apenas que um agente externo (usuário) forneça um conjunto de palavras opinativas semente (léxico de sentimento).

A partir de regras baseadas em relações de dependência sintática, o DP utiliza um conjunto de palavras opinativas semente para extrair alvos de opinião (termos que representam aspectos) e a seguir utiliza os aspectos extraídos para identificar e extrair novas palavras opinativas. Termos que representam aspectos também são usados para extrair novos aspectos, bem como palavras opinativas conhecidas também são usadas para extrair novas palavras opinativas.

Visto que o objetivo deste trabalho de mestrado é adaptar o DP, esse algoritmo será melhor detalhado no capítulo 4.

3.2 CLASSIFICAÇÃO DE SENTIMENTO NO NÍVEL DE ASPECTO

A classificação de sentimento no nível de aspecto é, como o nome já diz, a tarefa de classificação do sentimento em relação a cada termo de aspecto identificado nos textos sob análise. Nesta seção, discutiremos sobre alguns dos métodos que realizam essa tarefa.

Com o objetivo de identificar quais recursos disponíveis para a língua portuguesa eram mais adequados para a realização da ASNA, Freitas e Vieira (2015) realizaram uma pesquisa mais ampla utilizando ontologias e léxicos de sentimento. Nesse estudo, elas variaram os *pos-taggers* (*TreeTagger*, *FreeLing* e *CitiusTagger*), léxicos de sentimento (*LIWC-PT*, *SentiLex-PT*, *OpLexicon*, *Onto.PT*), e métodos para a atribuição de polaridade (levando ou não em consideração a posição dos adjetivos).

As autoras utilizaram um método que leva em consideração janelas de palavras próximas ao termo de aspecto. Primeiro, verifica-se a existência de adjetivos à esquerda do termo de aspecto. Caso nenhum adjetivo seja encontrado na primeira etapa, verifica-se a existência de adjetivos após o termo, até que outro aspecto seja encontrado ou a frase termine. Ao final, são adicionadas regras de negação que podem alterar o

sentimento em relação ao aspecto.

A configuração que alcançou os melhores resultados foi a que utilizou o *Tre- eTagger* para o pré-processamento, o *Onto.PT* para atribuição de polaridades, e o método baseado na posição dos adjetivos para o cálculo das polaridades. No entanto, ao fim do estudo, elas chegaram à conclusão de que todos os *pos-taggers*, léxicos e métodos de atribuição de polaridade utilizados são adequados para a tarefa.

Já em (SÁPIRAS, 2015), o autor utilizou a polaridade das frases que mencionam os termos de aspectos como o sentimento associado a eles.

3.3 CONSIDERAÇÕES FINAIS

Neste capítulo, foram apresentados alguns trabalhos que realizam tarefas relacionadas à ASNA como a ETA e a classificação de sentimento no nível de aspecto. No próximo capítulo, será apresentado o processo proposto para a ETA baseado no algoritmo *Double Propagation*.

4 DOUBLE PROPAGATION PTBR

Os capítulos anteriores apresentaram o contexto no qual está inserido este trabalho de mestrado, iniciando com uma visão geral da Análise de Sentimento (Cap. 2), e aprofundando a AS no nível do aspecto, com foco nos processos de extração de termos de aspectos (Cap. 3).

Veremos agora o processo de ETA proposto para o Português do Brasil, o *Double Propagation PTBR*, principal contribuição deste trabalho de mestrado. Este trabalho foi inspirado em um algoritmo de extração de termos para a língua inglesa baseado em relações sintáticas (QIU et al., 2009) e (QIU et al., 2011), que tem obtido ótimos resultados na tarefa de extração de aspectos, sendo considerados por muitos pesquisadores como o ‘estado da arte’ na área (seção 4.1).

A seguir temos uma apresentação detalhada do algoritmo DP original, seguida da sua adaptação para a língua portuguesa (o DP PTBR). Por fim, temos uma breve conclusão deste capítulo.

4.1 DOUBLE PROPAGATION

Proposto inicialmente em (QIU et al., 2009) , e estendido em (QIU et al., 2011), o algoritmo *Double Propagation* (DP), ou Propagação Dupla (PD), é um algoritmo do tipo *bootstrapping* que foi desenvolvido com o propósito de simultaneamente extrair termos de aspectos e expandir léxicos de sentimento.

A partir de regras baseadas em relações de dependência sintáticas, como observado em 3.1.2.3, o PD utiliza um conjunto de palavras opinativas semente para extrair termos de aspectos e a seguir utiliza os aspectos extraídos para identificar e extrair novas palavras opinativas. Termos que representam aspectos também são utilizados para extrair novos aspectos, assim como palavras opinativas conhecidas também são usadas para extrair novas palavras opinativas.

O método possui essa nomenclatura justamente por propagar informações entre termos de aspectos e palavras opinativas. Esse é um processo iterativo, que termina apenas quando não é encontrado mais nenhum novo termo ou palavra opinativa.

O conjunto de regras baseadas em relações de dependência sintáticas para a realização das extrações pode ser visto na tabela 4. Para a aplicação dessas regras de extração, o texto de entrada (a partir do qual os termos serão extraídos) deve ser pré-processado por um *parser*, que irá realizar a segmentação das frases do texto, identificar as relações de dependência sintática de cada frase e etiquetar cada termo do texto com sua classe gramatical correspondente.

Nas tabelas 4 e 5, podem ser observadas as regras para a extração dos termos de aspecto e de palavras opinativas. Com respeito à nomenclatura dessas tabelas: O e A se referem à palavra opinativa e alvo de opinião, respectivamente; {L} se refere ao conjunto de palavras opinativas semente; CG(O/A) se refere à classe gramatical de O ou A; O(A)-Dep significa que O(A) é o dependente na respectiva relação de dependência; O -> O-Dep -> A significa que O é dependente de A; {MR} é o conjunto de dependências sintáticas utilizadas para a extração; {NN} é o conjunto de classes gramaticais que se referem a substantivos, enquanto que {JJ} é o conjunto de classes gramaticais que se referem a adjetivos; Por fim, {CONJ} é a relação de dependência sintática CONJ.

Tabela 4 – Regras para extração de termos de aspecto e palavras opinativas - 1 (QIU et al., 2011)

IDs das regras	Regras	Saídas	Exemplos
R1 ¹	O -> O-Dep -> A, O em {L}, O-Dep em {MR}, CG(A) em {NN}	A	O iPhone tem uma “tela” <u>grande</u> . (grande -> AMOD -> tela)
R1 ²	O -> O-Dep -> H <- A-Dep <- A, O em {O}, O/T-Dep em {MR}, CG(A) em {NN}	A	O “iPhone” é o <u>melhor</u> celular. (melhor -> AMOD -> celular <- NSUBJ <- iPhone)
R2 ¹	O -> O-Dep -> A, A em {A}, O-Dep em {MR}, CG(O) {JJ}	O	O mesmo que R1 ¹ , com tela como a palavra conhecida, e grande como a palavra a ser extraída.
R2 ²	O -> O-Dep -> H <- A-Dep <- A, A em {A}, O/T-Dep em {MR}, CG(O) em {JJ}	O	O mesmo que R1 ² , sendo “iPhone” o aspecto conhecido, e “melhor” a palavra a ser extraída.

Qiu et al. (2011)

Por questões de espaço, a tabela foi dividida em duas.

Tabela 5 – Regras para extração de termos de aspecto e palavras opinativas - 2 (QIU et al., 2011)

IDs das regras	Regras	Saídas	Exemplos
R3 ¹	Ai(j) -> Ai(j)-Dep -> Aj(i), Ai(j) em {A}, Ai(j)-Dep em {CONJ}, CG(Aj(i)) em {NN}	Ai(j)	O “teclado” e a <u>tela</u> são ótimos. (tela -> CONJ -> teclado)
R3 ²	Ai -> Ai-Dep -> H <- Aj-Dep <- Aj, Ai em {A}, Ai-Dep == Aj-Dep, CG(Aj) em {NN}	Aj	O iPhone “X” tem uma ótima “câmera”. (câmera -> dobj -> tem <- nsubj <- X)
R4 ¹	Oi(j) -> Oi(j)-Dep -> Oj(i), Oi(j) em {O}, Oi(j)-Dep em {CONJ}, CG(Oj(i)) em {JJ}	Oi(j)	A tela é “grande” e <u>bonita</u> . (bonita -> CONJ -> grande)
R4 ²	Oi -> Oi-Dep -> H <- Oj-Dep <- Oj, Oi em {O}, Oi-Dep == Oj-Dep, CG(Oj) em {JJ}	Oj	Se quiser comprar um telefone bom, bonito e barato, escolha o Xiaomi. (bom -> amod -> telefone <- amod <- bonito)

O conjunto de relações de dependência sintáticas {MR} é composto pelas relações: *AMOD*, *PREP*, *NSUBJ*, *CSUBJ*, *XSUBJ*, *DOBJ* e *IOBJ*.

Como pode ser observado nas tabelas 4 e 5, as regras R1 e R3 são utilizadas para a ETA, sendo as do tipo R1 para a ETA a partir de palavras opinativas e as do tipo R3 para a ETA a partir de termos de aspectos conhecidos. Por sua vez, os tipos de regra R2 e R4 são realizadas extrações de palavras opinativas para a expansão de léxicos de sentimento, sendo o primeiro tipo para extração de palavras opinativas a partir de alvos de opinião, e a última para extração de palavras opinativas a partir de palavras opinativas já conhecidas.

Na R1¹, um alvo de opinião é identificado a partir de uma relação de dependência pertencente a {MR}, e que possua uma palavra opinativa conhecida como dependente e a classe gramatical do governante da relação pertença a {NN}. Em R1², por sua vez, a extração ocorre quando uma palavras opinativa conhecida e o possível termo de aspecto, que precisa ter a classe gramatical pertencente a {NN}, sejam sintaticamente dependentes de uma mesma palavra.

Já em R3¹, um termo de aspecto é extraído quando é uma palavra substantiva e dependente sintaticamente de um termo de aspecto conhecido através da relação CONJ. Já em R3², a extração é realizada quando duas palavras são dependentes de uma mesma palavra a partir de relações de dependência equivalentes e uma das palavras dependentes é um termo de aspecto conhecido, e o possível termo de aspecto é um substantivo.

Com respeito à extração de palavras opinativas, o método R2¹ é o mesmo que R1¹, porém voltada para a extração de palavras opinativas, sendo verificado se o possível termo opinativo é um adjetivo dependente de um termo de aspecto conhecido. Por sua vez, o R2² é o mesmo que a R1², porém também voltada para a extração de palavras opinativas, na qual é verificada se o possível novo termo opinativo é dependente de uma mesma palavra que uma palavra opinativa conhecida.

As regras do tipo R4 também são utilizadas para a extração de termos opinativos, e são similares às regras do tipo R3. Em R4¹, um termo opinativo é extraído quando é uma palavra adjetiva dependente sintaticamente de uma palavra opinativa conhecida através da relação CONJ. Já em R4², a extração é realizada quando duas palavras são dependentes de uma mesma palavras através de relações de dependência equivalentes e uma das palavras dependentes é uma palavra opinativa conhecida, e o possível termo opinativo é um adjetivo.

A seguir, temos o pseudo código do algoritmo *Double Propagation*.

Entrada: Léxico de sentimento $\{L\}$, Conjunto de resenhas $\{R\}$

Função:

01. $L\text{-exp} = \{L\}$
02. $\{A_i\} = \emptyset$
03. $\{L_i\} = \emptyset$
04. PARA CADA frase EM $\{R\}$:
05. $A_i = \text{aplicar}(R1.1) + \text{aplicar}(R1.2)$
06. $L_i = \text{aplicar}(R4.1) + \text{aplicar}(R4.2)$
07. FIM DO PARA
08. $\{A\} = \{A\} + \{A_i\}$
09. $\{L\text{-exp}\} = \{L\text{-exp}\} + \{L_i\}$
10. $\{A'\} = \emptyset$
11. $\{L'\} = \emptyset$
12. PARA CADA frase EM $\{R\}$:
13. $A' = \text{aplicar}(R3.1) + \text{aplicar}(R3.2)$
14. $L' = \text{aplicar}(R2.1) + \text{aplicar}(R2.2)$
15. FIM DO PARA
16. $\{A_i\} = \{A_i\} + \{A'\}$
17. $\{L_i\} = \{L_i\} + \{L'\}$
18. $\{A\} = \{A\} + \{A'\}$
19. $\{L\text{-exp}\} = \{L\text{-exp}\} + \{L'\}$
20. voltar à linha 02 até que $\{A_i\} = \emptyset$ e $\{L_i\} = \emptyset$

Saída: Termos aspectos identificados $\{A\}$, Léxico de sentimento expandido $\{L\text{-exp}\}$

O funcionamento do algoritmo dá-se da seguinte forma: são dados como entrada um conjunto de palavras opinativas semente (léxico de sentimento) e os textos, a partir dos quais serão extraídos os termos de aspectos e novas palavras opinativas. As resenhas devem passar por um processo de segmentação de frases, *parsing* e *post-tagging*, visto que o processo de extração se dá analisando frase a frase, e a extração propriamente dita se dá a partir das relações de dependência das frases e das classes gramaticais das palavras nelas presentes.

A cada iteração, todas as frases do corpus são percorridas duas vezes. Na primeira verificação, é realizada a ETA a partir de palavras opinativas e da extração de novas palavras opinativas a partir de palavras opinativas já conhecidas. Isso porque *a priori*, como só conhecemos as palavras opinativas, só conseguimos realizar a extração a partir destas. Por sua vez, os termos de aspecto e palavras opinativas ainda não conhecidos extraídos na primeira passagem pelas frases do corpus a cada iteração

são adicionados a um conjunto $\{A_i\}$ e $\{L_i\}$, respectivamente. Que, por sua vez, são adicionados ao conjunto final de termos de aspectos $\{A\}$ e palavras opinativas $\{L\}$.

A seguir, dentro da mesma iteração, mais uma vez as frases são verificadas, dessa vez com o objetivo de se extraírem palavras opinativas a partir de termos já conhecidos, bem como de extrair novos termos de aspecto. Os novos termos de aspectos e palavras opinativas identificados são adicionados a um conjunto auxiliar $\{A'\}$ e $\{L'\}$, respectivamente, que são adicionados a $\{A_i\}$ e $\{L_i\}$, que, por sua vez, são adicionados ao conjunto final $\{A\}$ e $\{L\}$.

Quando os conjuntos $\{A_i\}$ e $\{L_i\}$ estiverem vazios ao final de uma iteração, o processo iterativo termina. Caso contrário, o processo iterativo de extração continua até que nenhum novo termo de aspecto ou nova palavra opinativa sejam identificados.

Além do processo iterativo de extração, ainda são realizadas algumas tarefas chamadas de poda. Os métodos de poda podem ser tanto de agregação como de eliminação de termos de aspecto. Qiu et al. (2011) utilizam dois métodos de podagem para eliminação de termos de aspecto e um híbrido, que agrega e elimina termos ao final.

O primeiro método de poda, o qual chamaremos de Poda Baseada em Oração (PBO), verifica se dois termos de aspectos distintos aparecem em uma mesma oração e não estão conectados por uma conjunção. Caso isso aconteça, dentre os dois, é eliminado o termo de aspecto menos frequente no *corpus*.

O segundo método de podagem, o qual chamaremos de Podagem de Outros Produtos (POP), consiste em identificar nomes de outros produtos nas resenhas referentes a um determinado produto. Qiu et al. fazem isso utilizando padrões pré-determinados que indicam comparações entre produtos.

Seguimos esses mesmos padrões pré-definidos por Qiu et al., os quais precisaram apenas serem adaptados para o português, foram eles: “melhor que”, “pior que”. Por exemplo, quando se está analisando produtos sobre o iPhone, e há uma frase “ele é melhor que o Nokia”, o termo Nokia há de ser extraído pelas regras de extração. No entanto, por não se estar analisando esse produto, é desejável que esse termo seja eliminado da lista final de aspectos extraídos, portanto é utilizado o POP.

O terceiro método de poda, o qual chamaremos de Poda de Termos Compostos de Aspectos (PTCA), consiste em descobrir, a partir dos termos que só contém uma palavra (termos simples) extraídos no processo iterativo, termos compostos (que apresentam mais de uma palavra) que representam aspectos. O PTCA percorre o *corpus* analisando a ocorrência de cada termo de aspecto, analisando em uma janela de palavras de tamanho q se existem k adjetivos antes dos termos. Caso isso aconteça, o termo de aspecto composto é adicionado ao conjunto de termos de aspectos. Qiu et

al. utilizam q igual a 2, e k igual a 1. Ou seja, é verificada a presença de substantivos em uma janela de duas palavras do termo de aspecto, e a ocorrência de apenas um adjetivo antes do termo composto. Ao final, são eliminados os termos de aspectos que ocorrem apenas uma vez no corpus.

Na próxima seção, discutiremos algumas dificuldades encontradas ao se replicar o *Double Propagation*.

4.1.1 Dificuldades na Replicação do *Double Propagation*

Marrese-Taylor e Matsuo (2017) realizaram um estudo sobre as dificuldades em se reproduzir técnicas baseadas em análise sintática para a extração de termos de aspectos para a mineração de opinião. Dentre esses métodos, encontra-se o DP. Esse estudo foi motivado pois os autores identificaram uma falta de clareza de outros autores na literatura ao reportar os seus experimentos e resultados, o que acaba tornando difícil a replicação dos mesmos, conseqüentemente dificultando a evolução da pesquisa na área.

Além do fator mencionado acima, vale mencionar que a primeira dificuldade encontrada por Marrese-Taylor e Matsuo (2017) em reproduzir o trabalho de Qiu et al. (2011) foi que Qiu et al. utilizaram o *parser* Minipar para extrair as relações de dependências das frases. No entanto, essa ferramenta se tornou obsoleta, o site que mantinha essa ferramenta já não está mais no ar, e os binários da ferramenta encontrados na *web* estão corrompidos. Logo, Marrese-Taylor e Matsuo tiveram que recorrer ao *StanfordCoreNLP* (MANNING et al., 2014) para extrair as relações de dependência dos textos. Como haviam inconsistências entre as nomenclaturas das relações de dependência das duas ferramentas, os autores tiveram que recorrer ao trabalho de (LIU et al., 2015) para mapear as relações correspondentes entre as duas ferramentas.

Ainda que o *Minipar* tivesse sido projetado apenas para o inglês, o que não impactaria o presente trabalho, foi importante mencionar esse fato, pois neste trabalho foi adotada as mesmas relações de dependência mapeadas em (LIU et al., 2015).

Mais dificuldades foram encontradas nas etapas de poda dos termos de aspectos. Primeiro, na etapa de PBO, Qiu et al. (2011) afirmam que identificaram os limites das orações com o *Minipar*, no entanto não explicam como os aspectos são conectados por uma conjunção. Então, Marrese-Taylor e Matsuo (2017) consideraram como orações um conjunto de sub-árvores sintáticas rotuladas como “S”, que representa a palavra *sentence*, do inglês “frase”. Por fim, para verificar se os aspectos estão conectados por uma conjunção em uma cláusula, os autores consideraram apenas se há conjunção entre os aspectos dentro da cláusula.

Além das dificuldades relatadas em (MARRESE-TAYLOR; MATSUO, 2017),

outro problema encontrado no trabalho de (QIU et al., 2011) é em relação às técnicas de poda, pois não fica claro se essas são realizadas durante ou após o processo de extração. Ainda que seja preferível a eliminação de não-aspectos identificados como aspectos durante a extração para evitar desse não-aspecto ser responsável pela extração de mais ruídos, os autores não deixam claro quando são realizados esses métodos.

Por fim, Marrese-Taylor e Matsuo (2017) detalha a discrepância ao comparar os resultados dos seus experimentos com os resultados dos experimentos relatados por Qiu et al., mesmo utilizando a mesma base de dados utilizada no trabalho original.

4.2 DOUBLE PROPAGATION PTBR

Veremos aqui detalhes do processo para a ETA, baseada no algoritmo DP¹.

4.2.1 Visão Geral

Para realizar a adaptação do DP para a língua portuguesa, foi necessário criar um *parser* sintático capaz de extrair relações de dependência de textos escritos na língua portuguesa. Processo que será descrito na seção 4.2.2.

Após a criação do *parser*, será visto o processo proposto para a ETA baseada no DP. Propomos a adição de uma etapa de normalização de textos antes do processo de extração propriamente dito. Além disso, adicionamos mais regras para a ETA e mais etapas de poda para a eliminação de não aspectos. Isso será detalhado na seção 4.3.

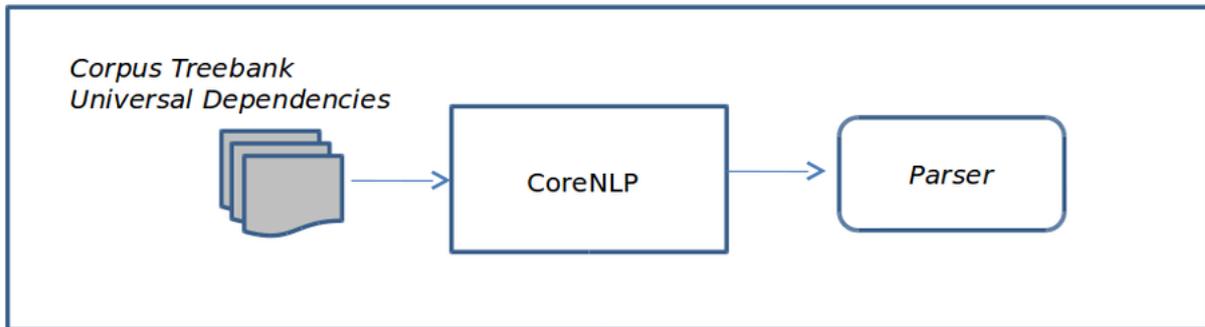
4.2.2 Criação do *Parser* Sintático

Como discutido previamente, o DP realiza a ETA a partir das relações de dependência sintática presentes em frases. Esse tipo de informação é adquirida a partir de *parsers* sintáticos. Como no momento do desenvolvimento do projeto deste trabalho não dispúnhamos de um *parser* para a língua portuguesa adequado para esse propósito, foi necessário criar um.

O processo realizado para a criação do *parser* pode ser visto na figura 2.

¹ Código, parser e dados disponíveis em: <https://github.com/arthurcaique/dppt>

Figura 2 – Processo de criação do *parser* sintático



Criado pelo autor

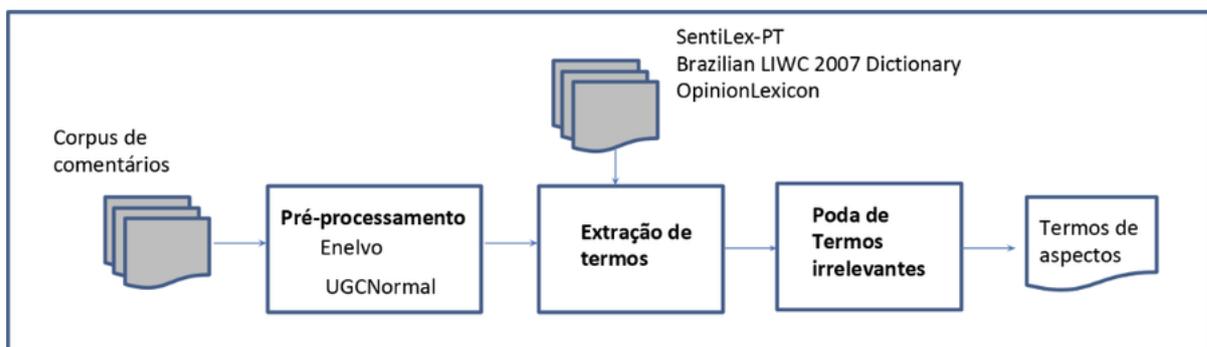
A *Stanford CoreNLP* (MANNING et al., 2014), conjunto de diversas ferramentas relacionadas ao PLN, possui um módulo que é capaz de treinar *parsers*, sendo necessário um *treebank* e um *word embedding* pré-treinado. Um *treebank* é um corpus cujos textos são anotados com as suas estruturas sintáticas e semânticas. O *treebank* que utilizamos para a criação do *parser* foi o *Brazilian Portuguese Universal Dependencies* (MARNEFFE; MACCARTNEY; MANNING, 2006) (MCDONALD et al., 2013). Já os *embeddings* utilizados foram os treinados com o Word2Vec por Hartmann et al. (2017).

4.2.3 Processo de Extração de Termos de Aspectos

Como pode ser visto na figura 3, o processo de extração Double Propagation PTBR conta com três etapas distintas: (1) pré-processamento dos textos de entrada; (2) extração de termos de aspecto usando o algoritmo DP adaptado para o Português brasileiro; e (3) pós-processamento da lista de termos extraídos.

Veremos aqui mais detalhes sobre a solução proposta neste trabalho. Cada etapa será detalhadamente descrita.

Figura 3 – Processo de Extração de Termos de Aspectos



Criado pelo autor

Pré-processamento dos textos de entrada

O pré-processamento realizado consistiu de duas etapas: (1) normalização dos textos; (2) e extração de informações sintáticas a partir de um *pos-tagger* e um *parser* sintático.

Na etapa de pré-processamento do algoritmo original, os textos eram processados apenas por *pos-taggers* e *parsers* sintáticos, para que fossem realizadas as segmentações de frases, etapa na qual são identificadas o início e o fim de cada frase nos textos, e a extração de informações sintáticas como as classes gramaticais de cada palavra e as relações de dependência das frases.

Os textos utilizados neste trabalho para avaliar a adaptação do DP são provenientes da web. E é bem sabido que esses textos, advindos de fóruns da web e afins, possuem muito ruído, pois não há cuidado com a sua sintaxe, nem com a ortografia, o que pode causar erros de etiquetagem das classes gramaticais das palavras, que pode causar erros na etiquetagem das relações de dependência das frases, consequentemente podendo prejudicar o processo de ETA, pois geralmente, ferramentas de PLN tendem a ter uma queda no desempenho quando são apresentadas a esse tipo de ruído, justamente por terem sido treinadas com textos bem escritos, dentro de uma norma gramatical. O método DP foi projetado dessa forma, para trabalhar com textos bem escritos, dentro de uma norma gramatical. Logo, quando se trata de textos provenientes da web, é muito difícil assegurar que se tenha somente textos dentro da norma. Dessa forma, torna-se desejável o uso de um normalizador de textos capaz de corrigir esses erros.

Neste trabalho, foi investigado a utilização do normalizador de texto *UGCNormal* (DURAN; AVANÇO; NUNES, 2015) no processo de ETA. As etapas de normalização do *UGCNormal* são: detecção de limites de frases, capitalização de siglas, capitalização da primeira letra de palavras que iniciam frases e correção de palavras baseada em fonética. A normalização dos textos foi realizada antes do pré-processamento realizado pelo *parser* sintático, justamente para minimizar os erros ortográficos ou falta de pontuações apresentados ao *parser*.

Extração de termos de aspectos

Para a extração dos termos que representam aspectos, propomos mais regras complementares àquelas propostas por (QIU et al., 2009) Qiu et al. (2011). Isso porque, no *corpus* de teste criado a partir do Buscapé², o qual será descrito na seção 5.1, foram observados alguns padrões de relações que envolviam termos de aspectos, que as regras criadas por Qiu et al. não eram capazes de extrair. Além disso, as palavras

² buscape.com

opinativas dos léxicos de sentimento englobam palavras cujas classes gramaticais vão além do adjetivo, é o caso das classes gramaticais verbo e advérbio.

Diante desse cenário, se fez possível e necessário para se alcançar uma maior cobertura de termos de aspectos extraídos, a inclusão de mais uma regra para a ETA, a qual pode ser vista na tabela 6.

As regras propostas por Qiu et al. apenas compreendiam a extração de termos a partir de opiniões racionais, que vem do tangível, como discutido na seção 2.2.1. A adição da R5, a qual pode ser vista na tabela 6 torna possível a extração de termos a partir de opiniões emotivas, como foi discutido na seção 2.2.1 e como pode ser observado pela própria tabela com o exemplo “Eu amei o iPhone”.

Tabela 6 – Novas regras adicionadas ao Double Propagation

IDs das regras	Regras	Saídas	Exemplos
R5	A -> A-Dep -> O, CG(O) {VB}, O {L}, A-Dep {MR}, CG(A) {NN}	a = A	Eu amei o “iPhone”. (iPhone -> DOBJ -> amei)

Foi adicionado um novo método de agregação de termos de aspectos compostos que também considera preposições. Esse método é similar ao PTCA, distinguindo apenas que verifica-se se há algum substantivo separado de um termo de aspecto apenas por uma preposição, caso sim, o substantivo, a preposição e o aspecto são agregados como um só termo. Por exemplo, no termo de aspecto “fone de ouvido”, a palavra “fone” seria extraída a partir das regras de extração, então ao aplicar-se essa regra, seria identificado o substantivo “ouvido” sendo separado de “fone” apenas pela preposição “de”, sendo extraído o termo composto “fone de ouvido” como termo de aspecto.

A próxima etapa do processo é a ETA a partir do método *Double Propagation* adaptado.

Pós-processamento da lista de termos extraídos (Poda)

Também foram considerados outros métodos de poda após o processo de extração. Foram esses: a poda baseada em *p-support*, e a poda baseada na frequência do termo em um corpus geral do produto.

Em relação ao processo de *pruning*, não foi possível realizar o método de PBO,

visto que o módulo utilizado para carregar o modelo de *parsing* treinado, não oferece a opção de segmentação de orações de frases. Ou seja, sem a segmentação de orações, não é possível realizar o PBO. Vale mencionar que a não utilização desse método não acarreta em uma queda no desempenho da cobertura do algoritmo, visto que o PBO foi projetado para eliminar termos de aspectos não relacionados, ou seja, termos de aspectos não deixarão de ser descobertos pela não inclusão do PBO.

O *p-support* foi utilizado em (HU; LIU, 2004) para a remoção de termos que são “subtermos” de aspectos. Por exemplo, quando tem o termo “fone de ouvido”, e tanto os termos “fone” como “fone de ouvido” são identificados como aspectos, é desejável que o termo fone seja eliminado, afinal já temos o termo “fone de ouvido”. No entanto, “fone” pode aparecer sozinho em muitos comentários, significando que a eliminação desse termo pode acarretar em perda de informações importantes. O *p-support* ajuda a resolver esse impasse por ser uma métrica que leva em o número de frases que o termo ocorre sem ser “subtermo” de nenhum outro aspecto identificado. Caso esse número esteja abaixo de um limiar, o termo é eliminado.

Por último, também foi avaliado um método baseado em frequência para a eliminação de termos de aspectos. Visto que há muitas resenhas de muitos artigos de um mesmo produto, que compartilham os mesmos aspectos, e conseqüentemente os mesmos termos de aspecto, possíveis termos extraídos de um artigo são buscados na base externa de artigos do mesmo produto, caso a frequência de um termo esteja abaixo de um limiar, o mesmo é eliminado. Esse método foi chamado de Eliminação Baseada na Frequência Externa do Termo de Aspecto.

4.3 CONSIDERAÇÕES FINAIS

Neste capítulo, foi detalhado o algoritmo *Double Propagation* original, bem como foi detalhada a criação do *parser* sintático para a extração de relações de dependência sintática das frases, assim como o processo de ETA baseado no DP proposto nesse trabalho. Apresentamos mais regras para a extração de termos de aspectos, bem como mais métodos para agregação de termos compostos de aspectos, assim como mais métodos para eliminação de não aspectos.

No próximo capítulo, serão detalhados os experimentos e resultados obtidos pelo processo proposto.

5 EXPERIMENTOS E RESULTADOS

Neste capítulo, serão detalhados: a criação do corpus de teste, os experimentos realizados e os resultados obtidos.

5.1 CRIAÇÃO DO CORPUS DE TESTE

Como visto anteriormente, em geral, mesmo quando se alcançam bons resultados, raramente recursos como as base de dados rotuladas utilizadas em outros trabalhos na área de mineração de opinião para o português são disponibilizados. Dessa forma, dada a escassez desse tipo de dado, fez-se necessária a construção de um corpus de teste, que incluiu duas etapas: aquisição automática dos dados e etiquetagem manual.

Inicialmente, foi implementado um *crawler* para coletar comentários avaliativos de produtos eletrônicos a partir do site Buscapé¹, por ser um dos sites mais populares do Brasil para comparação de preços de produtos em *e-commerces*, além de ter uma estrutura de organização de seções de produtos bem definida.

A implementação do *crawler* foi realizado em Python 3, utilizando o pacote *BeautifulSoup*². Para a coleta, selecionamos produtos de três seções: smartphones, smartvts e notebooks. Essas produtos foram selecionadas por serem as mais populares do site.

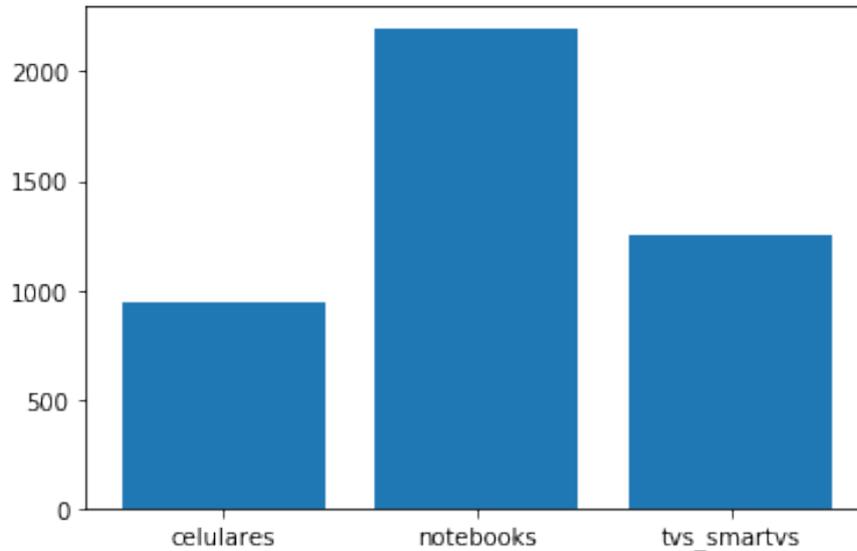
O processo de captura deu-se da seguinte forma: a partir das URLs referentes às seções de cada produto, o *crawler* capturou as URLs de todos os artigos referentes aos produtos. A partir das URLs dos produtos, o *crawler* foi capaz de percorrer as páginas dos respectivos produtos, realizando a coleta dos comentários avaliativos.

Como as páginas do Buscapé dispunham de um padrão, mudando apenas a numeração da página de um produto, bastou um processo iterativo para varrer todas as páginas de comentários dos artigos de um produto.

Na figura 4, podemos ver a distribuição do número de produtos por seção. A seção de notebooks é a que possui mais produtos, contando com mais de dois mil artigos de produtos, seguida pela seção de TVs e Smartvts que possui pouco mais de mil artigos cadastrados, enquanto que a de celulares tem pouco menos que mil, sendo a seção com o menor número de artigos dentre as três seções selecionadas.

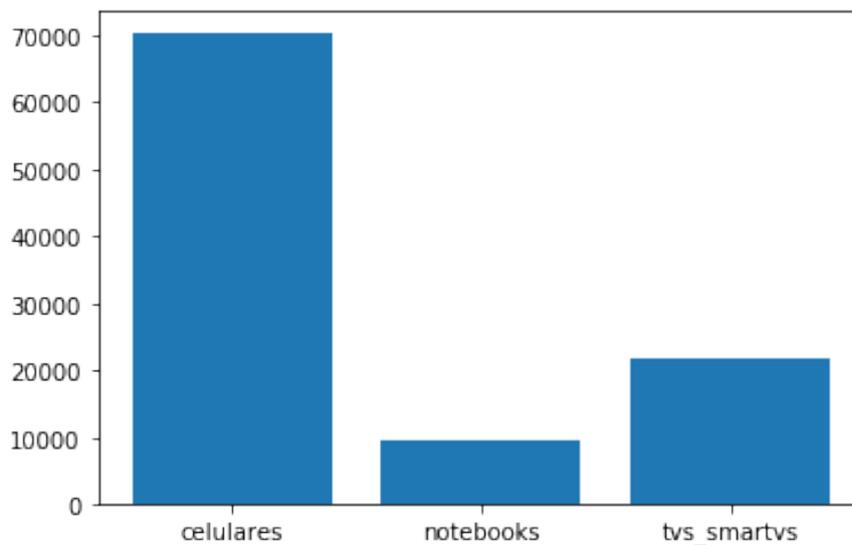
¹ www.buscapede.com.br

² <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

Figura 4 – Números de produtos por tipo de produto

Criada pelo autor

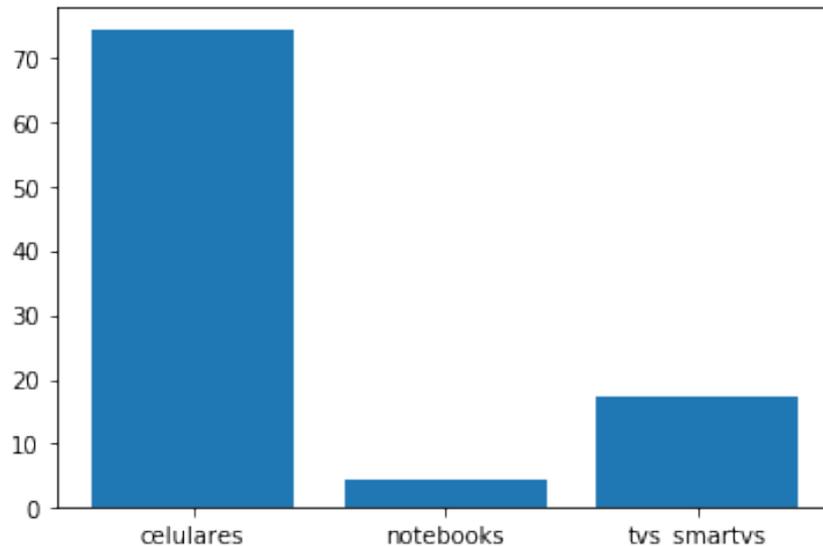
Por outro lado, quando avaliamos o número de comentários por tipo de produto, podemos observar que existem muito mais comentários para os celulares do que para os outros dois produtos, como pode ser visto na figura 5. A seção de celulares possui cerca de 70 mil comentários, enquanto que a seção de TVs e Smartvs possui cerca de 20 mil, e a de notebooks tem números próximos a 10 mil comentários.

Figura 5 – Números de comentários por tipo de produto

Por fim, para se ter uma ideia do impacto dos dados da figura 5, é necessário olhar para a média de comentários por tipo de produto, que é exibido na figura 6. A

seção de notebooks, a qual apresenta o maior número de produtos, é a que possui a menor média de número de comentários por produto, enquanto que a seção de celulares, a qual possui o menor número de produtos, é a que possui a maior média de número de comentários dentre as seções, com um número quase 20 vezes maior que a seção de notebooks, e 4 vezes maior que a seção de TVs e Smartvts.

Figura 6 – Média do número de comentários por seção



Criada pelo autor

5.1.1 Etiquetagem do Corpus de Teste

A etiquetagem do corpus de teste, indicando os termos de aspecto, é necessária para se realizar uma avaliação quantitativa do desempenho do processo de extração. A partir dessa marcação, é possível comparar a precisão e a cobertura do extrator de aspectos com a extração manual.

Como pode-se observar, o elevado número de comentários, 101.624 no total, torna impraticável uma anotação manual dos termos de aspectos que estão sendo avaliados em cada comentário. Dessa forma, foi selecionado um artigo de cada produto para a anotação de informações de aspectos. A seleção de cada artigo foi feita de forma aleatória. Na seleção automática aleatória, foram considerados apenas os produtos com mais de 50 comentários avaliativos.

Após a seleção dos produtos, foi realizada a anotação manual dos termos de aspectos avaliados nas resenhas dos respectivos produtos. Foram considerados apenas os termos de aspectos mencionados explicitamente nas resenhas. Foi desenvolvida uma ferramenta para auxiliar na tarefa de rotulação, cuja a interface pode ser vista na figura 7. Além dos termos de aspectos, foi rotulado o aspecto ao qual o termo se referia, a qual entidade o aspecto pertence, e o sentimento associado com o respectivo termo.

Figura 7 – Exemplo do sistema auxiliar para anotação de termos de aspectos

Anotador de Aspectos

Possuo este **aparelho** a quase 2 anos e não me arrependo sequer um pouquinho e tê-lo adquirido.

-No quesito **armazenamento** há um zero pecado da parte da motorola por optar ela exclusividade da memória interna, sendo posteriormente lançado um modelo com a opção micro SD e 4G.

-A **câmera** (5mp e 2mp) fica na categoria satisfatória, com mtas distorções no ambiente escuro, mas **fotos** mto boas com a iluminação adequada. Além da **personalização** limitada oferecida pelo software nativo da fabricante. Po

-**Bateria**, nunca me decepcionou, com o uso razoável, GPS, wi-fi e 3G ligados, é possível terminar o dia com por volta de 30%.

-Sua **memória Ram** de 1 gb e **processador** quad-core são perfeitos para o usuário médio, mas após anos de uso e vários aplicativos instalados observa-se engasgadas no **desempenho**.

-Seu **display** HD de 4,5" satisfaz, no dia a dia, jogos e vídeos no YouTube a 720p

-O **android** Puro oferecido pela **marca** é ótimo e pouquíssimos aplicativos obrigatórios de fábrica além dos impostos pela google. Destaque ao seu suporte q msm após 2 anos de **aparelho**, meu dispositivo esta com a última at

Considerações finais: Até msm após 2 anos de lançamento, há a vantagem em adquirir este **aparelho** desde que encontre-o por volta de 450 a 650 reais, esta , claro, minha opinião de consumidor satisfeito e fã da **marca**.

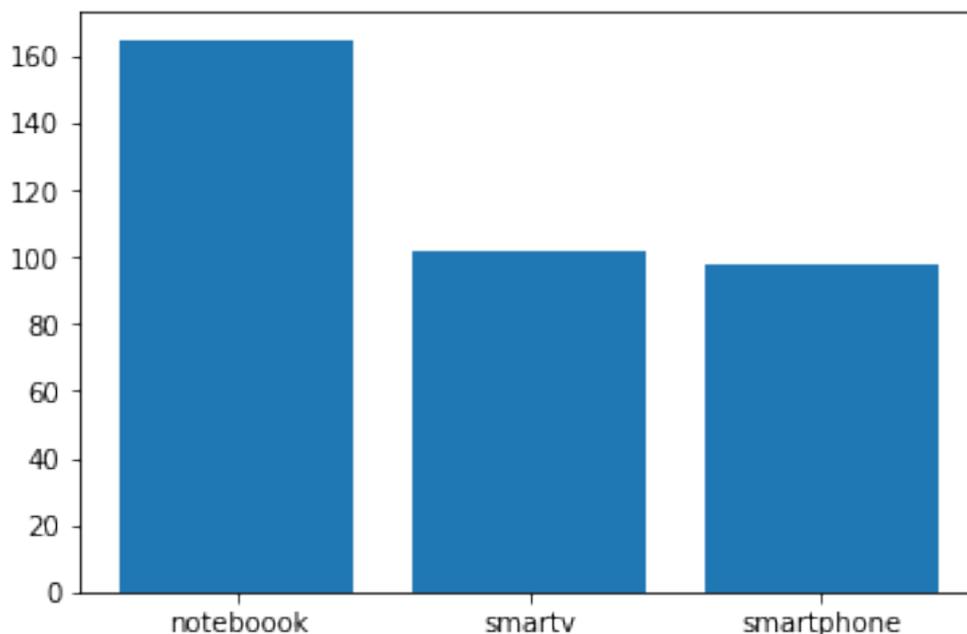
Expressão	Entidade	Aspecto	Sentimento
aparelho	XT1032 8 GB	aparelho	POSITIVO
armazenamento	XT1032 8 GB	armazenamento	NEGATIVO
câmera	XT1032 8 GB	câmera	POSITIVO
fotos	XT1032 8 GB	fotos	POSITIVO
personalização	XT1032 8 GB	personalização	NEGATIVO
preço	XT1032 8 GB	preço	POSITIVO
Bateria	XT1032 8 GB	Bateria	POSITIVO
memória Ram	XT1032 8 GB	memória Ram	POSITIVO
processador	XT1032 8 GB	processador	POSITIVO
desempenho	XT1032 8 GB	desempenho	NEGATIVO
display	XT1032 8 GB	display	POSITIVO
android	XT1032 8 GB	android	POSITIVO
marca	XT1032 8 GB	marca	POSITIVO

Anterior Próxima

Criada pelo autor

Ao fim desse processo, o corpus de teste para a realização dos experimentos contava com 365 resenhas de três produtos, cuja distribuição do número de resenhas por produto pode ser vista na figura 8, e a distribuição de termos de termos aspectos por produto na figura 9.

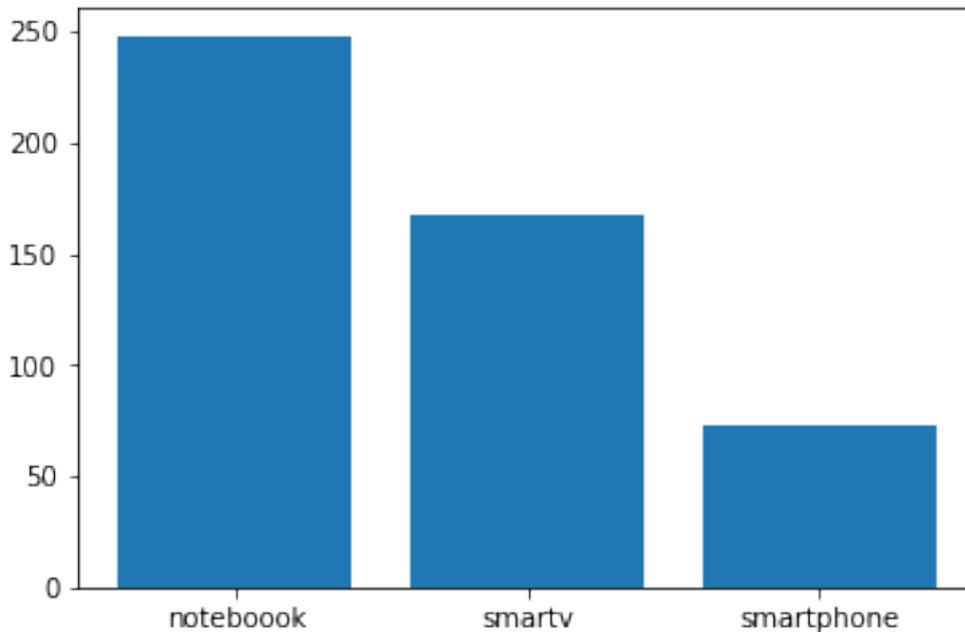
Figura 8 – Número de resenhas por produto no corpus de teste



Criada pelo autor

Como pode ser observado, o número de resenhas por produto no corpus de teste é maior que a média de resenhas por artigo de cada produto no corpus coletado do Buscapé. Nas Smartv's, esse número chega a ser cerca de doze vezes mais alto, enquanto que para o Smartphone, esse número chega a cinco vezes maior que a média de comentários por artigo de Smartphone.

Figura 9 – Número de termos de aspectos por produto no corpus de teste



Criada pelo autor

Em relação à quantidade de termos de aspectos anotados manualmente para cada produto, a distribuição é parecida com a de número de comentários por produto, como pode ser visto nas figuras 8 e 9.

Na próxima seção, serão detalhados os experimentos e resultados obtidos.

5.2 EXPERIMENTOS E RESULTADOS

Nesta seção, serão detalhados os experimentos e resultados obtidos.

5.2.1 Métricas de Avaliação

Como o problema tratado neste trabalho é um problema de recuperação de informação, as avaliações realizadas serão feitas a partir das métricas utilizadas para avaliar soluções desse tipo de problema, que são: precisão, cobertura e *f-measure* (BAEZA-YATES; RIBEIRO-NETO, 2011).

A precisão é calculada pela razão do número de termos corretamente extraídos (ttec) pelo número total de termos extraídos (tte) conforme a fórmula a seguir:

$$precisao = ttec/tte$$

A cobertura é obtida pela razão do número de termos de aspectos corretamente extraídos (ttec) pelo número total de termos de aspecto existentes no corpus (ttc) conforme a fórmula a seguir:

$$cobertura = ttec/ttc$$

E a *f-measure* é a média harmônica da precisão e cobertura conforme a fórmula a seguir:

$$f\ score = 2 * ((precisao * cobertura) / (precisao + cobertura))$$

5.2.2 Experimentos

Esta seção apresenta os experimentos realizados para avaliar o desempenho do protótipo construído.

5.2.2.1 Experimentos DP PTBR

Para os experimentos, avaliamos o desempenho médio geral de cada parâmetro do processo, foram eles: léxicos de sentimentos, uso ou não de normalizadores de texto, limiares dos métodos de poda, e da efetividade da nova regra proposta para a ETA. Os resultados serão detalhados a seguir. A melhor configuração de cada parâmetro foi utilizada para a comparação da abordagem proposta com outras abordagens.

5.2.2.1.1 Avaliação da Efetividade dos Léxicos de Sentimento

Nesta seção, apresentamos os resultados da efetividade de cada léxico de sentimento adotado para o processo de extração, foram eles: O SentiLex Lem, SentiLex Flex, LIWC 2007 PT-BR e o OPLex, os quais foram apresentados na seção 2.3.1.1.

Como pode ser observado na tabela 8, todos os léxicos tiveram uma precisão parecida, com exceção do OPLex que esteve bem abaixo. Em relação à cobertura, o OPLex esteve bem acima dos outros. No entanto, com relação à *f-measure*, os resultados apresentados são equiparáveis, com o OPLex um pouco mais abaixo dos outros, sendo o LIWC PT-BR o que apresentou essa taxa um pouco mais alta. Dessa forma, para a comparação dos resultados entre abordagens na seção 5.2.2.2, o DP PTBR será executado com o léxico de sentimento LIWC PT-BR como a lista de palavras opinativas semente.

No entanto, com base nos resultados apresentados na tabela 8, não é possível afirmar se há um léxico mais adequado para ser usado como a lista de palavras na tarefa de ETA do DP PTBR.

Os resultados apresentados na tabela 11 são baseados na média dos resultados de todas as configurações observando-se os resultados com a variação dos léxicos de sentimento.

Tabela 7 – Resultados da Efetividade dos Léxicos de Sentimento no Processo de Extração

Léxico de Sentimento	Precisão	Cobertura	F-Measure
SentiLex Lem	0.378	0.504	0.427
SentiLex Flex	0.376	0.509	0.427
LIWC 2007 PT-BR	0.376	0.514	0.429
OPLex	0.318	0.617	0.413

Desempenho dos Léxicos de Sentimento Utilizando Todos os Métodos de Poda

5.2.2.1.2 Avaliação da Efetividade do Uso de Normalizadores de Texto

Em relação ao uso ou não de um normalizador de texto, o UGCNormal obteve, na média geral dos resultados, um melhor desempenho do que a não utilização de um normalizador antes do processo de extração, o que pode corroborar a hipótese de que erros de *parsing* causados por ruídos como erros de pontuação ou ortografia podem prejudicar o desempenho de um sistema que dependem de informações sintáticas de textos.

Os resultados apresentados na tabela 8 são baseados na média dos resultados de todas as configurações observando-se os resultados com a variação do uso ou não de um normalizador de texto.

Tabela 8 – Avaliação dos normalizadores de texto

Normalizador	Precisão	Cobertura	F-Measure
Sem Normalizador	0.353	0.527	0.415
UGCNormal	0.372	0.545	0.433

Desempenho dos Léxicos de Sentimento Utilizando Todos os Métodos de Poda

5.2.2.1.3 Avaliação da Efetividade dos Métodos de Poda

Foram avaliadas as efetividades do *P-support* e da poda baseada na frequência do termo nos comentários considerando todos os artigos de um produto.

No *P-support*, como já era de se esperar, conforme aumenta-se o limiar de corte de termos, é melhorada a precisão do processo de extração, mas isso só ocorre até um certo ponto, como pode ser observado na tabela 10. Por sua vez, quando menor o limiar, maior a cobertura alcançada. A melhor configuração alcançada foi com o *p-support* em 0.5, o que significa que a melhor configuração do *P-support* é quando eliminam-se os termos cujo o número de frases que ele aparece sem ser “subtermo” de outro aspecto é menor que a metade do número total de vezes que ele ocorre no corpus.

Os resultados apresentados nas tabelas 10 e 10 são baseados na média dos resultados de todas as configurações observando-se os resultados com a variação do *p-support* e frequência dos termos, respectivamente.

Tabela 9 – Poda Baseada no P-support dos Termos

P-Support	Precisão	Cobertura	F-Measure
0	0.346	0.568	0.422
0.1	0.363	0.542	0.427
0.3	0.364	0.542	0.427
0.5	0.366	0.541	0.428
0.7	0.366	0.534	0.426
0.9	0.358	0.509	0.412

Da mesma forma, como apresentado na tabela 10, a precisão do processo de extração tem um maior desempenho quando eleva-se o limiar do método de eliminação baseada na frequência dos termos no corpus geral com todos os artigos de um produto. Por outro lado, a cobertura é maior quanto menor o limiar. O melhor ponto de corte encontrado foi o 2. O que significa que termos identificados como aspectos que ocorrem menos de duas vezes no corpus com todos os artigos do produto são eliminados.

Tabela 10 – Poda Baseada na Frequência de Termos No Corpus com Comentários sobre todos os Artigos de um Produto

Limiar	Precisão	Cobertura	F-Measure
0	0.303	0.612	0.398
2	0.36	0.56	0.431
4	0.363	0.531	0.424
6	0.368	0.525	0.425
8	0.371	0.52	0.425
10	0.371	0.517	0.425

5.2.2.1.4 Avaliação da Nova Regra Proposta para a Extração de Termos

Nesta seção, avaliamos a nova regra proposta para a extração de termos de aspectos, a qual foi detalhada em 4.3.

Tabela 11 – Resultados da Nova Regra Proposta Para o Processo de Extração

Uso da Nova Regra	Precisão	Cobertura	F-Measure
Sim	0.356	0.543	0.423
Não	0.387	0.498	0.43

Criado pelo autor

Como pode ser observado na tabela 11, a nova regra fez com que o processo alcançasse uma cobertura 5 pontos percentuais maior do que a versão sem a utilização do mesmo. No entanto, em relação à precisão, a inclusão da nova regra apresentou um desempenho 3% mais baixa nos experimentos realizados. Por sua vez, a F-Measure ficou balanceada entre as duas configurações, porém a não inclusão da regra alcançou um desempenho ligeiramente superior.

5.2.2.2 Comparação com outras abordagens

Para a comparação com outros métodos de ETA, foram considerados os resultados das melhores configurações dos parâmetros obtidas na seção anterior para o DP PTBR. Para se ter uma avaliação justa, foram realizados a normalização dos textos também quando realizou-se os experimentos para os outros métodos.

Na comparação de resultados, foram considerados métodos não supervisionados. Foram esses: o *DP*, o *DP* sem iterações, o *WhatMatter*, método baseado em substantivos e sintagmas nominais mais frequentes no corpus de (SIQUEIRA; BARROS, 2010), e um método baseado em ontologias. Para esse último, só foi possível realizar as comparações para as resenhas de *smartphones*, visto que no domínio dos produtos selecionados, só foi encontrada disponível, para a língua portuguesa, uma ontologia de termos de aspectos para *Smartphones*.

Vale mencionar que, para o *WhatMatter*, foi utilizada a configuração proposta por Siqueira (2010), que identifica como termo de aspecto, os 3% substantivos e sintagmas nominais mais frequentes no corpus. No entanto, também foi verificado, para cada produto, qual limiar alcançou a maior *f-measure*, e os resultados foram apresentados nas respectivas tabelas de cada avaliação de resultados a seguir.

Na resenhas de notebook sem normalização, como pode ser observado na tabela 17, o DB PTBR alcançou a maior cobertura e F-Measure. Em relação à F-Measure, é possível observar que todos os métodos, excluindo o *WhatMatter* original, que ficou bem abaixo, obtiveram um resultado muito parecido.

Tabela 12 – Resultados obtidos para as resenhas de notebook sem normalização

Método	Precisão	Cobertura	F-measure
DP PTBR	0.366	0.592	0.452
DP PTBR (Sem iterações)	0.362	0.506	0.422
Double Propagation	0.368	0.498	0.423
Double Propagation (Sem iterações)	0.368	0.498	0.423
WhatMatter	0.428	0.18	0.253
WhatMatter (30%)	0.383	0.467	0.421

Na resenhas de notebook utilizando o UGCNormal, como pode ser observado na tabela 17, os resultados foram parecidos àqueles apresentados na tabela 17, com o *WhatMatter* obtendo a maior precisão, e o DP PTBR alcançando a maior cobertura e F-Measure. É importante observar que o DB PTBR sem iterações obteve o segundo

melhor desempenho nessa configuração.

Tabela 13 – Resultados obtidos para as resenhas de notebook com o UGC Normal

Método	Precisão	Cobertura	F-measure
DP PTBR	0.358	0.595	0.447
DP PTBR (Sem iterações)	0.36	0.536	0.43
Double Propagation	0.327	0.464	0.384
Double Propagation (Sem iterações)	0.328	0.441	0.376
WhatMatter	0.43	0.18	0.253
WhatMatter (30%)	0.357	0.45	0.398

Nas resenhas de Smartvs, o mesmo panorama apresentado anteriormente se repetiu, como pode ser observado na tabela 17. No entanto, a precisão nos dados de Smartvs foram maiores, e a cobertura foi menor para esses dados.

Tabela 14 – Resultados obtidos para as resenhas de Smartvs sem normalização

Método	Precisão	Cobertura	F-measure
DP PTBR	0.422	0.497	0.456
DP PTBR (Sem iterações)	0.419	0.458	0.438
Double Propagation	0.44	0.431	0.436
Double Propagation (Sem iterações)	0.44	0.431	0.436
WhatMatter	0.5	0.137	0.215
WhatMatter (30%)	0.427	0.405	0.416

Os resultados alcançados com a utilização do UGCNormal nas resenhas de Smartvs foram um pouco melhores do que sem o uso do normalizador para o DP PTBR, como pode ser observado na tabela 17. Por sua vez, as duas versões do DP original tiveram uma queda no desempenho nessa configuração.

Tabela 15 – Resultados obtidos para as resenhas de Smartvts com o UGC Normal

Método	Precisão	Cobertura	F-measure
DP PTBR	0.429	0.503	0.463
DP PTBR (Sem iterações)	0.436	0.477	0.456
Double Propagation	0.4	0.397	0.399
Double Propagation (Sem iterações)	0.408	0.384	0.396
WhatMatter	0.6	0.119	0.198
WhatMatter (32%)	0.441	0.423	0.432

Como observado anteriormente, nas resenhas de Smartphones, foi possível comparar o método de extração de termos de aspectos a partir de ontologias. Como era de se esperar, por ser uma lista pronta com aspectos desse tipo de produto, esse método alcançou uma precisão superior aos demais métodos, como pode ser observado na tabela 17, que apresenta os resultados dos métodos comparados para as resenhas de smartphones sem normalizadores de texto. No entanto, a ontologia teve um baixo desempenho na cobertura dos termos de aspectos de smartphones, isso pode ser explicado por causa da data de criação da ontologia. Ela foi criada em 2007, e alguns termos de aspectos que se tornaram mais populares posteriormente não foram incluídos na ontologia. Por sua vez, o DP PTBR, mais uma vez, o DP PTBR apresentou a melhor cobertura nos resultados. No entanto, as versões originais do DP (com e sem iterações) apresentaram as melhores taxas de f-measure.

Tabela 16 – Resultados obtidos para as resenhas de Smartphones sem normalizador de textos

Método	Precisão	Cobertura	F-measure
DP PTBR	0.292	0.553	0.382
DP PTBR (Sem iterações)	0.29	0.5	0.367
Double Propagation	0.392	0.526	0.449
Double Propagation (Sem iterações)	0.392	0.526	0.449
WhatMatter	0.5	0.065	0.116
WhatMatter (18%)	0.416	0.394	0.405
Ontologia	0.55	0.144	0.229

Em geral, nas resenhas de Smartphones normalizadas pelo UGCNormal, houve

uma melhora no desempenho dos métodos. O DP original apresentou as melhores taxas de F-Measure para esse produto, o DP PTBR apresentou a melhor cobertura, e por sua vez, o *WhatMatter* obteve uma melhora significativa na precisão, obtendo a maior taxa de precisão nessa métrica.

Tabela 17 – Resultados obtidos para as resenhas de Smartphones com o UGC Normal

Método	Precisão	Cobertura	F-measure
DP PTBR	0.336	0.563	0.421
DP PTBR (Sem iterações)	0.345	0.549	0.424
Double Propagation	0.382	0.549	0.451
Double Propagation (Sem iterações)	0.376	0.535	0.442
WhatMatter	0.666	0.084	0.15
WhatMatter (26%)	0.391	0.507	0.441
Ontologia	0.55	0.144	0.229

5.2.3 Discussão dos resultados

Esta seção apresenta uma discussão final sobre os resultados gerais obtidos, exibidos na tabela 18. É válido mencionar que não consideramos os resultados baseado em ontologia para esta discussão, pois o método para a ETA baseada em ontologia só foi realizada nas resenhas de smartphone como explicitado na seção anterior.

Tabela 18 – Avaliação final

Método	Precisão (Média)	Precisão (Desvio padrão)	Cobertura (Média)	Cobertura (Desvio Padrão)	F-measure (Média)	F-Measure (Desvio padrão)
DP PTBR	0.362	0.06	0.536	0.076	0.424	0.032
DP PTBR (Sem iterações)	0.362	0.061	0.518	0.082	0.417	0.034
Double Propagation	0.362	0.055	0.48	0.057	0.407	0.032
Double Propagation (Sem iterações)	0.362	0.056	0.472	0.061	0.404	0.033
WhatMatter (Original)	0.52	0.094	0.127	0.048	0.197	0.55
WhatMatter (Melhores configurações)	0.402	0.03	0.44	0.042	0.418	0.016

Como pode ser visto na tabela de avaliações gerais 18, e nas tabelas de avaliações por produtos da seção anterior, o WhatMatter apresenta as maiores taxas de precisão. Isso se dá pelo fato desse método considerar apenas os n% substantivos mais frequentes, que acabam sendo realmente termos relacionados aos aspectos e entidades avaliadas.

No entanto, como discutido na seção 3.1.2, considerando apenas uma relativamente pequena parte de substantivos mais frequentes dificilmente consegue-se uma boa cobertura. Esse raciocínio é corroborado pelos resultados apresentados até aqui, e como pode ser observado na tabela 18.

Como pode ser observado, tanto o DP PTBR como o DP, assim como as suas versões sem iterações, alcançaram a mesma precisão média geral. Dessa forma, o fator responsável por mostrar um melhor desempenho entre essas técnicas ficar por conta da taxa de cobertura.

As maiores virtudes do *Double Propagation* são justamente em relação às taxas de cobertura, que acabam também por melhorar a *f-measure*, visto que esta é uma medida que leva em consideração a média harmônica entre a precisão e cobertura.

Com a adição de mais uma regra para extração de termos de aspectos, proposta neste trabalho, o DP PTBR apresenta as maiores taxas de cobertura, ainda que o DP original adaptado para a língua portuguesa apresente taxas comparáveis. Como

observado anteriormente, isso é refletido na *f-measure*.

Ainda que a nova regra de extração pudesse ter apresentado novos ruídos no conjunto de termos de aspectos extraídos, isso não prejudicou o método proposto, visto que as taxas de precisão permaneceram iguais. Em alguns casos, como pode ser visto na seção anterior, as taxas de precisão do DP PTBR até alcançou melhores resultados, isso devido aos métodos de poda existentes que foram adicionados ao método original.

5.3 CONSIDERAÇÕES FINAIS

Neste capítulo, foi apresentada a criação do corpus de teste, desde a sua aquisição, seleção dos comentários, até a rotulação manual de termos de aspectos realizada.

Também, foram apresentados os experimentos realizados, a fim de se descobrir qual o melhor valor para cada parâmetro do DP PTBR.

E, por fim, foi realizada uma comparação com outras abordagens não supervisionadas. Foi possível verificar que para o cenário apresentado, em resenhas de produtos eletrônicos, o DP PTBR obteve resultados satisfatórios, principalmente com relação à taxa de cobertura, tendo um desempenho até um pouco melhor do que o DP original.

6 CONCLUSÃO

Neste capítulo, serão apresentadas as principais contribuições deste trabalho de mestrado, bem como as direções futuras que pretende-se seguir a partir dele.

6.1 PRINCIPAIS CONTRIBUIÇÕES

Neste trabalho de mestrado, foi realizada uma discussão sobre o atual estado da arte da AS aplicada à língua portuguesa do Brasil, bem como foi apresentada uma visão das dificuldades e oportunidades de pesquisa possibilitadas por essas dificuldades.

O objetivo específico deste trabalho, foi a realização de uma adaptação do algoritmo *Double Propagation* para a língua portuguesa do Brasil, bem como a realização de propostas de melhorias ao algoritmo adaptado. Apesar das dificuldade reportadas na literatura sobre a replicação desse algoritmo, foi possível implementá-lo e adaptá-lo para a língua portuguesa. Dessa forma, fica como contribuição principal deste trabalho mais um processo bem sucedido de ETA para a língua portuguesa.

Como o DP PTBR utiliza relações de dependência para realizar a ETA, também fica como contribuição a investigação do uso dessas relações sintáticas para a ETA aplicada à língua portuguesa. Como observado, o uso dessas relações alcançaram altas taxas de cobertura na extração, o que significa que o uso de relações de dependência sintática é adequada para essa tarefa.

O *parser* criado que pode valer como alternativa aos *parsers* existentes para a língua portuguesa.

Também fomos capazes de investigar a efetividade na variação de léxicos de sentimento disponíveis para a língua portuguesa no processo de ETA. Também contribuímos com a investigação sobre a inclusão de mais regras de extração ao algoritmo, as quais tornaram possível a melhora no desempenho do algoritmo original.

Investigamos ainda se o uso de um normalizador de texto aplicado às resenhas antes do processo de extração seria capaz de melhorar o desempenho do algoritmo, o que de fato aconteceu.

Por fim, apresentamos propostas de trabalhos futuros a serem desenvolvidos a partir deste trabalho de mestrado.

6.2 TRABALHOS FUTUROS

Como trabalhos futuros, propomos:

- Realizar a investigação de mais métodos de poda
 - Como o processo de extração do DP PTBR apresentou baixas taxas de precisão, propomos a investigação de mais métodos de poda de termos.
- Investigar o *framework* de Aprendizado Perpétuo (*Lifelong Learning*) para a ETA na língua portuguesa
 - Visto que há muitos produtos que compartilham os mesmos aspectos, propomos a investigação desse *framework* para esta tarefa na língua portuguesa.
- Testar a efetividade do *DB PTBR* em outros domínios
 - Neste trabalho, investigamos o *DB PTBR* apenas ao domínio de resenhas de produtos eletrônicos. Então, para trabalhos futuros, propomos a investigação da efetividade desse processo em outros domínios.
- Avaliar a expansão dos léxicos de sentimento
 - Durante a execução deste trabalho de mestrado, não fomos capazes de realizar uma avaliação da expansão dos léxicos de sentimento. Dessa forma, essa tarefa ficará como trabalho futuro.
- Realizar o agrupamento de termos de aspectos que se referem ao mesmo aspecto
 - Visto que vários termos de aspectos podem se referir ao mesmo aspecto, como sinônimos, ficará como trabalho futuro a implementação de um método que realize o agrupamento de termos de aspectos.

REFERÊNCIAS

- ALMEIDA, M. S. C. et al. Aligning Opinions: Cross-Lingual Opinion Mining with Dependencies. In: **Annual Meeting of the Association for Computational Linguistics (ACL)**. [S.l.: s.n.], 2015.
- AVANÇO, L. V.; BRUM, H. B.; NUNES, M. G. V. Improving opinion classifiers by combining different methods and resources. **XIII Encontro Nacional de Inteligência Artificial e Computacional**, 2016.
- AVANÇO, L. V.; NUNES, M. das G. V. Lexicon-Based Sentiment Analysis for Reviews of Products in Brazilian Portuguese. In: **2014 Brazilian Conference on Intelligent Systems (BRACIS)**. [S.l.: s.n.], 2014. v. 1, n. 1, p. 277 – 281.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval: The Concepts and Technology Behind Search**. 2. ed. [S.l.]: Addison Wesley, 2011.
- BALAGE FILHO, P. P. **Aspect extraction in sentiment analysis for Portuguese**. 2017. 74 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos - SP.
- BALANGE FILHO, P. P.; PARDO, T. A. S.; ALUÍSIO, S. M. An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In: STIL 2013 – THE 9TH BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 2013. **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**. [S.l.], 2013. p. 215 – 219.
- BRUCKSCHEN, M. et al. **Anotação Lingüística em XML do Corpus PLN-BR**. [S.l.], 2008.
- CARVALHO, P.; SILVA, M. J. Sentilex-PT: Principais Características e Potencialidades. In: _____. **Linguística, Informática e Tradução: Mundo que se cruzam**. 1. ed. Oslo: [s.n.], 2015. v. 7, p. 425 – 438. ISBN 978-82-91398-12-9.
- CHAVES, M. S. et al. PIRPO: An Algorithm to Deal with Polarity in Portuguese Online Reviews from the Accommodation Sector. p. 296 – 301, 2012.
- CHAVES, M. S.; TROJAHN, C. Towards a Multilingual Ontology for Ontology-driven Content Mining in Social Web Sites. 2010.
- CIRQUEIRA, D. et al. Performance Evaluation of Sentiment Analysis Methods for Brazilian Portuguese. 2017.
- DURAN, M. S.; AVANÇO, L. V.; NUNES, M. das G. V. A Normalizer for UGC in Brazilian Portuguese. In: **Workshop on Noisy User-generated Text. Association for Computational Linguistics-ACL**. [S.l.: s.n.], 2015.
- FREITAS, C. et al. Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. In: INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO DA USP, 2012. **Anais do XI Encontro de Linguística de Corpus - ELC 2012**. [S.l.], 2012.

- FREITAS, L. A. D. **Feature-Level Sentiment Analysis Applied to Brazilian Portuguese Reviews**. 2015. 94 p. Tese (Doutorado em Ciência da Computação) — Pontifícia Universidade Católica do Rio Grande do Sul.
- FREITAS, L. A. D.; VIEIRA, R. Comparing Portuguese Opinion Lexicons in Feature-Based Sentiment Analysis. p. 147 – 158, 2013.
- FREITAS, L. A. D.; VIEIRA, R. Ontology-based Feature Level Opinion Mining for Portuguese Reviews. p. 367 – 370, 2013.
- FREITAS, L. A. D.; VIEIRA, R. Exploring Resources for Sentiment Analysis in Portuguese Language. In: **2015 Brazilian Conference on Intelligent Systems**. [S.l.: s.n.], 2015. p. 152 – 156.
- HARTMANN, N. et al. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In: **STIL**. [S.l.: s.n.], 2017. v. 1, p. 122 – 131.
- HU, M.; LIU, B. Mining and summarizing customer reviews. In: **Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04**. Seattle: [s.n.], 2004.
- LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: ICML, 2001, Williamstown. **Eighteenth International Conference on Machine Learning**. Williamstown, 2001. p. 282 – 289.
- LI, J.; HOVY, E. Reflections on Sentiment/Opinion Analysis. In: _____. **A Practical Guide to Sentiment Analysis**. [S.l.]: Springer International Publishing AG, 2017. v. 5, cap. 3. ISBN 978-3-319-55392-4.
- LIU, B. **Sentiment analysis**: Mining opinions, sentiments, and emotions. 1. ed. [S.l.]: Cambridge University Press, 2015. v. 1. ISBN 9781139084789, 9781107017894.
- LIU, Q. et al. Automated Rule Selection for Aspect Extraction in Opinion Mining. **Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15**, p. 1291 – 1297, 2015.
- LIWC. 2017. Disponível em: <<http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>>. Acesso em: 26/08/2017.
- MANNING, C. D. et al. The Stanford CoreNLP Natural Language Processing Toolkit. **Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations**, p. 55 – 60, 2014.
- MÁRIO J. SILVA et al. Automatic Expansion of a Social Judgment Lexicon for Sentiment Analysis. In: **Computational Processing of the Portuguese Language**. [S.l.: s.n.], 2010.
- MARNEFFE, M. de; MACCARTNEY, B.; MANNING, C. D. Generating Typed Dependency Parses from Phrase Structure Parses. **Proceedings of LREC**, v. 6, 2006.
- MARRESE-TAYLOR, E.; MATSUO, Y. Replication issues in syntax-based aspect extraction for opinion mining. In: **EACL 2017 Student Research Workshop (SRW)**. [S.l.: s.n.], 2017.

- MCDONALD, R. et al. Universal Dependency Annotation for Multilingual Parsing. **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics**, p. 92 – 97, 2013.
- MICHAELIS. 2017. Disponível em: <<http://michaelis.uol.com.br/>>. Acesso em: 21/08/2017.
- MORAES, S. M. et al. Comparing Approaches to Subjectivity Classification: A Study on Portuguese Tweets. In: PROPOR, 2016, Tomar. **Computational Processing of the Portuguese Language**. Tomar, 2016. p. 86 – 94.
- PASQUALOTTI, P. R.; VIEIRA, R. WordnetAffectBR: uma base lexical de palavras de emoções para a língua portuguesa.
- PENNEBAKER, J. W. et al. **The Development and Psychometric Properties of LIWC2007**. Austin, 2007.
- PENNEBAKER, J. W.; FRANCIS, M. E.; BOOTH, R. J. **Linguistic inquiry and word count: Liwc 2001**. [S.l.], 2001.
- QIU, G. et al. Expanding Domain Sentiment Lexicon through Double Propagation. 2009.
- QIU, G. et al. Opinion Word Expansion and Target Extraction through Double Propagation. **Computational Linguistics**, v. 37, n. 1, p. 9 – 27, Março 2011.
- RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. **Knowledge Based Systems**, v. 89, p. 14 – 46, Novembro 2015.
- SÁPIRAS, L. A. **Mineração de Opiniões Baseada em Aspectos em Fontes de Opiniões Fracamente Estruturadas**. 2015. 75 p. Dissertação (Programa de Pós-Graduação em Ciência da Computação) — Universidade Federal do Rio Grande do Sul.
- SHU, L.; XU, H.; LIU, B. Lifelong Learning CRF for Supervised Aspect Extraction. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2017. **Annual Meeting of the Association for Computational Linguistics**. [S.l.], 2017.
- SILVA, M.; CARVALHO, P.; SARMENTO, L. Building a sentiment lexicon for social judgement mining. **Computational Processing of the Portuguese Language**, p. 218 – 228, 2012.
- SIQUEIRA, H.; BARROS, F. A Feature Extraction Process for Sentiment Analysis of Opinions on Services. **Proceedings of International Workshop on Web and Text Intelligence**, p. 404 – 413, Outubro 2010.
- SIQUEIRA, H. B. A. **WhatMatter**: Extração e visualização de características em opiniões sobre serviços. 2010. 109 p. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Pernambuco.
- SOUZA, E. et al. Characterizing Opinion Mining: A Systematic Mapping Study of the Portuguese Language. p. 122 – 127, 2016.

SOUZA, M. et al. Construction of a Portuguese Opinion Lexicon from Multiple Resources. In: **8th Brazilian Symposium in Information and Human Language Technology**. [S.l.: s.n.], 2012.

TABOADA, M. et al. Lexicon-Based Methods for Sentiment Analysis. **Computational Linguistics**, v. 37, n. 2, p. 267 – 307, Junho 2011.

TURNEY, P. D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: ANAIS, 2002, Philadelphia. **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**. Philadelphia, 2002. p. 417 – 424.

VARGAS, F. A.; PARDO, T. A. S. **Estudo empirico sobre agrupamento e organização hierárquica de aspectos para mineração de opinião**. [S.l.], 2017.