UNIVERSIDADE DE SÃO PAULO
FFCLRP-DEPARTAMENTO DE COMPUTAÇÃO E MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

“Estudo e avaliação de métodos de análise de sentimentos
baseada em aspectos para textos opinativos em português”

Mateus Tarcinalli Machado

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Mestre em Ciências, Área: Computação Aplicada.

Ribeirão Preto-SP

2018

Mateus Tarcinalli Machado

“Estudo e avaliação de métodos de análise de sentimentos
baseada em aspectos para textos opinativos em português”

Versão Corrigida

(Versão original encontra-se na Faculdade de Filosofia, Ciências e
Letras de Ribeirão Preto da Universidade de São Paulo que aloja
o Programa de Pós-graduação em Computação Aplicada)

Dissertação apresentada à Facul-
dade de Filosofia, Ciências e Le-
tras de Ribeirão Preto da USP,
como parte das exigências para
a obtenção do título de Mestre
em Ciências, Área: Computação
Aplicada.

Orientador: Prof. Dr. Evandro Eduardo Seron Ruiz

Ribeirão Preto–SP

2018

Mateus Tarcinalli Machado

Estudo e avaliação de métodos de análise de sentimentos baseada em aspectos para textos opinativos em português. Ribeirão Preto–SP, 2018.

88p. : il.; 30 cm.

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Mestre em Ciências,
Área: Computação Aplicada.

Orientador: Prof. Dr. Evandro Eduardo Seron Ruiz

1. Mineração de Opinião. 2. Identificação de Polaridade. 3. Identificação de Aspectos. 4. Análise não Supervisionada de Texto.

Mateus Tarcinalli Machado

“Estudo e avaliação de métodos de análise de sentimentos
baseada em aspectos para textos opinativos em português”

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Mestre em Ciências, Área: Computação Aplicada.

Trabalho aprovado. Ribeirão Preto–SP, 05 de setembro de 2018:

**Prof. Dr. Evandro Eduardo Seron
Ruiz**
Orientador

**Prof. Dr. José Eduardo Santarem
Segundo**
Examinador 1

**Profa. Dra. Helena de Medeiros
Caseli**
Examinador 2

**Prof. Dr. Tiago Agostinho de
Almeida**
Examinador 3

Ribeirão Preto–SP
2018

*Este trabalho é dedicado aos meus filhos, Miguel e Ana Helena, e a minha esposa
Vanessa, fontes da minha inspiração.*

Resumo

Esta dissertação tem como objeto de estudo a análise de sentimentos baseada em aspectos, aplicação derivada da análise de sentimentos e da área de processamento de linguagem natural. A análise de sentimentos baseada em aspectos é focada em analisar textos avaliativos (textos contendo opiniões) buscando identificar e relacionar sentimentos e aspectos de uma determinada entidade (produtos, serviços entre outros).

As principais etapas do desenvolvimento deste trabalho são a identificação de aspectos, que busca identificar as características de determinada entidade no texto e a identificação de sentimentos que procura encontrar o sentimento expresso pelo autor com relação ao aspecto mencionado.

O objetivo deste trabalho é implementar, analisar, melhorar e criar métodos não supervisionados de análise de sentimentos baseada em aspectos para textos em português. Essa exploração se dará pela implementação de métodos para identificação de aspectos e sentimentos, criação e combinação de léxicos de sentimentos.

Para alcançar esse objetivo realizamos experimentos com conjunto de dados anotado, ou seja, já com os aspectos e sentimentos relacionados marcados em seu texto. Para o processamento, além de técnicas de processamento de língua natural, como a análise gramatical, foram utilizados métodos de análise estatística dos textos e resultados.

Palavras-chave: mineração de opinião. identificação de polaridade. identificação de aspectos. análise não supervisionada de texto.

Abstract

This dissertation has as object of study the aspect based sentiment analysis, application derived from sentiment analysis and the area of natural language processing. The aspect based sentiment analysis focuses on analyzing evaluative texts (texts containing opinions) seeking to identify and relate feelings and aspects of a particular entity (products, services among others).

The main stages of the development of this work are the identification of aspects, which seeks to identify the characteristics of a certain entity in texts, and sentiment identification that aims to identify the feelings expressed by the author concerned about the mentioned aspects.

The purpose of this work is to implement, analyze, improve and create unsupervised methods of aspect based sentiment analysis applying them in portuguese language texts. This exploration will be through the implementation of methods for identifying aspects and sentiments, creation and combination of sentiment lexicons.

To achieve this goal we performed experiments with annotated data set, that is, texts with the related aspects and sentiments already marked. For processing, in addition to natural language processing techniques, such as grammatical analysis, methods of statistical analysis of texts and results were used.

Keywords: opinion mining. polarity identification. identification of aspects. unsupervised analysis of text.

Lista de figuras

Figura 1 – Resultados FREQ Baseline corpus ReLi.	46
Figura 2 – Resultados FREQ Baseline corpus ReLi estendido.	48
Figura 3 – Resultados Hu & Liu corpus ReLi sem sintagmas.	50
Figura 4 – Resultados FREQ + <i>Word2Vec</i> corpus ReLi.	51
Figura 5 – Resultados Hu & Liu + <i>Word2Vec</i> corpus ReLi.	53
Figura 6 – Resultados Hu & Liu + Infrequentes <i>Word2Vec</i> corpus ReLi.	55
Figura 7 – Resultados corpus ReLi.	57
Figura 8 – Resultados corpus ReLi Estendido.	58
Figura 9 – Resultados dos experimentos realizados com léxicos individuais.	73
Figura 10 – Resultados dos experimentos realizados com léxicos combinados.	74
Figura 11 – Resultados Hu & Liu corpus ReLi com sintagmas.	83
Figura 12 – Resultados Hu & Liu corpus ReLi estendido com sintagmas.	84
Figura 13 – Resultados Hu & Liu corpus ReLi estendido sem sintagmas.	84

Lista de tabelas

Tabela 1 – Conteúdo abordado na fundamentação teórica.	34
Tabela 2 – Composição do ReLi.	36
Tabela 3 – Distribuição de polaridades no ReLi.	37
Tabela 4 – Ocorrências de aspectos por frase.	37
Tabela 5 – Ocorrências de aspectos simples e compostos.	37
Tabela 6 – Composição do ReLi estendido.	38
Tabela 7 – Termos mais frequentes dos corpus analisados.	42
Tabela 8 – Tabela de contingência:	44
Tabela 9 – Resultados FREQ Baseline corpus ReLi.	46
Tabela 10 – Resultados FREQ Baseline corpus ReLi estendido.	48
Tabela 11 – Resultados Hu & Liu corpus ReLi.	49
Tabela 12 – Resultados Hu & Liu corpus ReLi sem sintagmas	49
Tabela 13 – Resultados FREQ + <i>Word2Vec</i> corpus ReLi.	51
Tabela 14 – Resultados FREQ + <i>Word2Vec</i> corpus ReLi estendido.	52
Tabela 15 – Resultados Hu & Liu + <i>Word2Vec</i> corpus ReLi.	53
Tabela 16 – Resultados Hu & Liu + <i>Word2Vec</i> corpus ReLi estendido.	54
Tabela 17 – Resultados Hu & Liu + <i>Word2Vec</i> nos Infrequentes corpus ReLi.	55
Tabela 18 – Resultados Hu & Liu + <i>Word2Vec</i> nos Infrequentes corpus ReLi.	56
Tabela 19 – Distribuição de polaridades relacionadas a aspectos no ReLi.	60
Tabela 20 – Distribuição de polaridades encontradas nos léxicos.	63
Tabela 21 – Resultados Polaridade das Palavras.	67
Tabela 22 – Resultados Somente Adjetivos	68
Tabela 23 – Resultados Somente Adjetivos <i>Word2Vec</i> Versão 1	69
Tabela 24 – Resultados Somente Adjetivos <i>Word2Vec</i> Versão 2.	69
Tabela 25 – Resultados Preferência aos Adjetivos.	70
Tabela 26 – Resultados Aspectos + Polaridade das Palavras.	71
Tabela 27 – Resultados Aspectos + Somente Adjetivos.	72
Tabela 28 – Resultados Aspectos + Preferência aos Adjetivos	72
Tabela 29 – Resultados Hu & Liu corpus ReLi estendido com sintagmas	83
Tabela 30 – Resultados Hu & Liu corpus ReLi estendido sem sintagmas	85
Tabela 31 – Resultados Preferência aos Adjetivos <i>Word2Vec</i> Versão 1	85
Tabela 32 – Resultados Preferência aos Adjetivos <i>Word2Vec</i> Versão 2	86
Tabela 33 – Resultados Aspectos + Somente Adjetivos <i>Word2Vec</i> Versão 1	86
Tabela 34 – Resultados Aspectos + Somente Adjetivos <i>Word2Vec</i> Versão 2	87
Tabela 35 – Resultados Aspectos + Preferência aos Adjetivos <i>Word2Vec</i> Versão 1	87
Tabela 36 – Resultados Aspectos + Preferência aos Adjetivos <i>Word2Vec</i> Versão 2	88

Sumário

1	INTRODUÇÃO	19
1.1	Objetivos	21
1.2	Materiais e Métodos	21
1.3	Contribuições	22
1.4	Organização do Trabalho	22
2	FUNDAMENTOS TEÓRICOS	23
3	TRABALHOS RELACIONADOS	27
4	IDENTIFICAÇÃO DE ASPECTOS	35
4.1	Introdução	35
4.2	Dados Analisados	36
4.3	Métodos para Identificação de Aspectos	38
4.4	Métricas de Avaliação	43
4.5	Resultados dos Experimentos	44
4.6	Conclusão	56
5	IDENTIFICAÇÃO DE SENTIMENTOS	59
5.1	Introdução	59
5.2	Dados Analisados	60
5.3	Métodos para Criação de Léxicos de Sentimentos	60
5.4	Métodos para Identificação de Sentimentos	63
5.5	Métricas de Avaliação	66
5.6	Resultados dos Experimentos	66
5.7	Conclusão	73
6	CONCLUSÕES	75
6.1	Dificuldades e Limitações	76
6.2	Contribuições	76
6.3	Trabalhos Futuros	76
	Referências	79
	Anexos	83

Introdução

A análise de sentimentos, aqui abreviada por SA, do Inglês *sentiment analysis*, é a área de pesquisa em computação aplicada, mais especificamente em processamento de linguagem natural (PLN), que busca extrair opiniões, emoções, avaliações e atitudes relacionadas a produtos, serviços, organizações, pessoas, eventos e assuntos expressos em um texto. Esta área de pesquisa também pode ser referenciada na literatura como mineração de opinião, extração de opinião, mineração de sentimentos, análise de subjetividade, análise de efeitos, análise de emoções, mineração de avaliações, entre outros termos (LIU, 2012).

Enquanto a pesquisa em processamento de linguagem natural tem uma longa história na ciência da computação, a qual muitos dizem iniciar no antológico artigo de Alan Turing “*Computing Machinery and Intelligence*” (TURING, 1950), poucas pesquisas foram realizadas sobre o tema de análise textual automatizada de opiniões e sentimentos pessoais antes do ano 2000. Desde então, a análise de sentimentos se tornou uma área muito ativa de pesquisa por vários motivos, os quais entre eles destacamos:

1. Trata-se de área com grande amplitude de aplicações, abrangendo praticamente todos os domínios do conhecimento. Mais especificamente, na avaliação de produtos houve uma grande proliferação de aplicações comerciais, o que levou ao recente impulsionamento das pesquisas nesta área;
2. A análise de sentimentos oferece muitos problemas de pesquisa desafiadores e ainda abertos em computação que basicamente cobrem tópicos específicos da compreensão computacional de textos livres. Muitos trabalhos foram publicados desde então, como por exemplo, o trabalho de Yang Liu (LIU et al., 2007) e sua equipe que propuseram um modelo para prever o desempenho de vendas por meio de blogs; Hong e Skiena (HONG; SKIENA, 2010) que analisaram as relações entre as apostas da NFL (Liga Nacional de Futebol Americano dos Estados Unidos) e opiniões públicas postadas em blogs e no *Twitter*; Tumasjan (TUMASJAN et al., 2010) e sua equipe utilizaram análise de sentimentos no *Twitter* para prever resultados de

eleições; Asur e Huberman (ASUR; HUBERMAN, 2010) utilizaram dados do *Twitter*, resenhas de filmes e blogs para prever bilheteria para filmes e também, Filho e Pardo (FILHO; PARDO, 2014) que desenvolveram um sistema de busca e análise de opiniões sobre produtos;

3. Hoje temos um grande volume de dados opinativos nas chamadas mídias sociais da web e, não surpreendentemente, o início do rápido crescimento da análise de sentimentos ocorreu junto aos meios de comunicação social.

Atualmente a análise de sentimentos, além de ter revigorado as pesquisas em processamento de linguagem natural, também influenciaram positivamente as Ciências de Gestão, as Ciências Políticas, a Economia, as Ciências Sociais Aplicadas, assim como outras áreas afetadas pela opinião pública, como veremos neste capítulo.

Uma das primeiras áreas de aplicação da análise de sentimentos foi a da avaliação de produtos que ocorreu com o advento da Web 2.0 seguindo o crescimento do comércio eletrônico e da participação mais ativa dos consumidores e internautas na web. Desde o artigo seminal de Pang & Lee em 2008 (PANG; LEE, 2008), os trabalhos científicos mostraram que a análise de sentimentos pode ser uma tarefa bastante complexa. Ela pode se tornar ainda mais desafiadora se o texto faz menções a mais de uma característica de uma mesma entidade avaliada. Neste caso, o avaliador poderá qualificar o item com sentimentos diferentes, e determinadas classificações podem ser positivas ou negativas dependendo da característica do produto (*aspecto*) analisada. Por exemplo, o fato de um produto ser considerado barato leva a um sentimento positivo, enquanto esse mesmo adjetivo relacionado ao material utilizado na construção do produto nos remete a um sentimento negativo. Dessa necessidade de qualificar características específicas de cada item ou objeto analisado nasceu a área conhecida como análise de sentimentos baseada em aspectos (ABSA, *aspect based sentiment analysis*). A ABSA realiza análises mais profundas em textos avaliativos, identificando e relacionando sentimentos e aspectos (THET; NA; KHOO, 2010) de uma determinada entidade.

Neste mesmo movimento advindo das tecnologias chamadas de Web 2.0 que propiciaram aos usuários a oportunidade de serem geradores de conteúdo com mais facilidade, as redes sociais conquistaram uma posição inovadora como espelhos da opinião pública. Além das redes sociais, temos hoje uma grande diversidade de sites especializados na avaliação de produtos e serviços, onde consumidores podem consultar ou compartilhar opiniões. Como exemplos, podemos citar os sites Buscapé¹ e Zoom² nos quais o consumidor pode opinar sobre produtos adquiridos, Adoro Cinema³ opiniões sobre filmes, Skoob⁴

¹ <<https://www.buscapi.com.br/>>

² <<https://www.zoom.com.br/>>

³ <<http://www.adorocinema.com/>>

⁴ <<https://www.skoob.com.br/>>

sobre livros, entre muitos outros. Por consequência, esses sites se tornaram uma grande fonte de dados para estudos relacionados a ABSA.

Apesar de termos, hoje, uma grande quantidade de trabalhos de pesquisa, recursos e ferramentas relacionados à área de análise de sentimentos baseada em aspectos para a língua inglesa, ainda não possuímos tal abundância com relação a língua portuguesa.

1.1 Objetivos

Dada essa carência relativa de trabalhos na língua portuguesa, o objetivo deste trabalho é implementar, analisar, melhorar e criar métodos de análise de sentimentos baseada em aspectos para textos em português, focado em métodos não supervisionados, isto é, métodos que não necessitam da classificação manual de textos para seu funcionamento. Essa exploração se dará através da implementação e avaliação de métodos para análise de sentimentos baseada em aspectos. A avaliação será realizada com a comparação dos resultados obtidos com dados previamente classificados.

1.2 Materiais e Métodos

Dividimos o processo de análise de sentimentos baseada em aspectos em duas etapas: a) identificação de aspectos, etapa com objetivo de identificar as características do produto avaliado mencionadas no texto, e; b) identificação de sentimentos, a etapa cujo foco é identificar a polaridade (positiva ou negativa) relacionada a cada aspecto do produto.

Na identificação de aspectos buscamos analisar diferentes técnicas para identificação de aspectos, como métodos frequenciais, métodos baseados em substantivos e sintagmas nominais e também o uso do modelo de semântica distribucional *Word2Vec* com objetivo de remover aspectos erroneamente identificados.

Na etapa de identificação de sentimentos, além de utilizarmos léxicos de sentimentos existentes, testamos técnicas para combinação e criação de novos dicionários. Avaliamos os léxicos existentes e os criados com métodos de identificação de polaridade.

Aplicamos os métodos de identificação de aspectos e sentimentos sobre o corpus ReLi, que possui marcação dos aspectos e sentimentos relacionados. Assim foi possível avaliar a eficiência dos métodos e léxicos implementados.

1.3 Contribuições

Este projeto contribui aplicando, avaliando e incrementando métodos de análise de sentimentos baseada em aspectos em textos na língua portuguesa. Os resultados apresentados podem servir como base para justificar o uso de determinados métodos ou léxicos, e apontar pontos fortes e fracos de cada um. Especificamente para análise de sentimentos, nossa principal contribuição é o método para criação de léxicos de sentimentos sensíveis ao contexto, que pode ser facilmente aplicado na criação de outros léxicos relacionados a diferentes temas. Os métodos de criação e combinação de léxicos são importantes dada a escassez de recursos na língua portuguesa, e por serem formas eficazes na obtenção de melhores resultados.

1.4 Organização do Trabalho

Este documento está estruturado da seguinte forma: no Capítulo 2 apresentamos os fundamentos teóricos sobre análise de sentimentos baseada em aspectos. No Capítulo 3 os trabalhos relacionados. Já no Capítulo 4 abordamos os métodos de identificação de aspectos, os experimentos e as conclusões relacionadas. No Capítulo 5 apresentamos os métodos de identificação de sentimentos, os experimentos e as conclusões relacionadas a criação de léxicos para este fim. Por fim, no Capítulo 6, apresentamos as conclusões gerais do trabalho e comentamos sobre eventuais trabalhos futuros.

Fundamentos Teóricos

Análise de Sentimentos Baseada em Aspectos

Como mencionado no capítulo anterior, Introdução, a análise de sentimentos estuda principalmente textos opinativos que expressam ou implicam sentimentos positivos ou negativos. Ao contrário da informação factual, opiniões e sentimentos são formados por textos subjetivos. Devido a essa característica, é importante que sejam examinados conjuntos de opiniões de muitas pessoas, caso contrário temos apenas a visão subjetiva de uma pessoa sobre algo o que geralmente não é esperado de uma aplicação que envolva análise de sentimentos. Como lidamos com conjuntos de opiniões, faz-se necessário alguma forma de sumarização dessas informações (HU; LIU, 2004a).

Inicialmente precisamos estabelecer uma definição de opinião. Observe a seguinte avaliação de uma câmera Canon, adaptada de (LIU, 2012):

Postado por: John Smith

Data: 10 de setembro de 2011

“(1) Comprei uma câmera Canon G12 seis meses atrás. (2) Simplesmente amei. (3) A qualidade das fotos é incrível. (4) A duração da bateria também é longa. (5) Entretanto, minha esposa achou a câmera muito pesada.”

A partir dessa opinião podemos notar alguns pontos importantes: o texto traz várias opiniões positivas e uma negativa sobre a câmera. A frase (2) traz um sentimento positivo sobre a câmera no geral. A frase (3) traz uma opinião positiva sobre a qualidade das fotos. A frase (4) traz uma opinião positiva sobre a duração da bateria. A frase (5) expressa uma opinião negativa sobre o peso da câmera. Podemos concluir que uma opinião consiste em dois componentes chaves, formando o par (g, s) : um alvo g e um sentimento s relacionado a esse alvo.

O alvo g pode se referir a qualquer entidade ou aspecto da entidade sobre a qual a opinião se refere. O sentimento s pode ser positivo, negativo ou neutro, ou ainda um

valor numérico que expressa a intensidade do sentimento como, por exemplo, uma escala entre 1 e 5 estrelas. Esse tipo de classificação em positivo, negativo ou neutro é chamado de polaridade do sentimento ou da opinião.

Uma entidade pode ser definida como produto, serviço, tópico, questão, pessoa, organização ou evento que está sendo avaliado. Uma entidade pode ser decomposta em uma hierarquia de partes, subpartes e assim por diante cada qual com seus conjuntos de atributos.

Voltando ao nosso exemplo, temos a câmera Canon G12 como entidade raiz e outros atributos ou partes mencionadas nas frases. Temos na frase (2) a própria entidade como aspecto, na frase (3) o atributo qualidade das fotos, na frase (4) a bateria e na frase (5) o peso da câmera.

As opiniões expressas no texto se referem a duas pessoas, o autor do texto ("John Smith") e sua esposa, mencionada na frase (5). Outra característica que aparece no texto é sua data. Datas são características importantes pois permitem que seja avaliado como as opiniões sobre determinado produto mudam com o passar do tempo.

Dessa forma podemos definir opinião como uma quintupla (LIU, 2010; HU; LIU, 2004a):

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

tal que:

- e_i como o nome da entidade;
- a_{ij} como um aspecto da entidade e_i ;
- h_k como o emissor da opinião;
- t_l como o tempo em que a opinião foi emitida por h_k ;
- s_{ijkl} como um sentimento relacionado ao aspecto a_{ij} da entidade e_i , dada por um emissor h_k em um tempo t_l ;

Assim temos e_i e a_{ij} representando juntos o alvo da opinião. Os subscritos da definição servem para enfatizar o fato de que as cinco partes da quintupla devem corresponder-se entre si. Por exemplo, um sentimento s_{ijkl} deve ter sido expresso por um emissor h_k sobre um aspecto a_{ij} de uma entidade e_i em um tempo t_l .

O tipo de opinião que temos definido aqui é a chamada opinião regular. Outro tipo seria a opinião comparativa, que necessita de uma definição diferente, mas não será tratada neste trabalho.

Com base na definição anterior, podemos concluir que o objetivo da análise de sentimentos baseada em aspectos é dado um documento opinativo d , descobrir todas as quintuplas de opinião $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ presentes em d .

As principais tarefas relacionadas a análise de sentimentos derivam diretamente dos componentes dessa quintupla. Do primeiro componente, entidade, temos as tarefas de extração e classificação de entidades. O processo de extração é similar a tarefa de reconhecimento de entidades nomeadas (*Named Entity Recognition, NER*) da área de extração de informação. Após a extração é necessária realizar um agrupamento dessas entidades, dado que, pessoas podem se referir a uma mesma entidade de diversas maneiras. Por exemplo, Motorola pode ser escrito como “Mot”, “Moto” ou “Motorola”. É necessário então reconhecer que todas essas formas se referem à mesma entidade.

Como segundo componente temos os aspectos a_{ij} . O problema é basicamente o mesmo encontrado em relação às entidades, sendo necessária uma etapa de extração e outra de agrupamento. Por exemplo, “foto”, “fotografia” e “imagem” remetem a um mesmo aspecto relacionado a câmeras. Cada aspecto deve aparecer com um único nome dentro de uma aplicação. O processo de agrupar essas expressões que representam os aspectos em categorias é chamado de categorização de aspectos.

Os aspectos geralmente aparecem na forma de substantivos e sintagmas nominais, mas também podem aparecer como verbos, sintagmas verbais, adjetivos e advérbios. Às expressões encontradas nas formas de substantivos e sintagmas nominais damos o nome de aspectos explícitos. Por exemplo, “qualidade da imagem” na frase “A qualidade da imagem dessa câmera é ótima” é uma expressão de aspecto explícito. Expressões encontradas em outras formas classificamos como expressões de aspectos implícitos. Por exemplo, a expressão “cara” na frase “Esta câmera é cara” remete a um aspecto preço.

O terceiro componente é o sentimento. A tarefa de ABSA implica a identificação do sentimento relacionado ao aspecto mencionado. Esse sentimento pode ser classificado como positivo, negativo ou neutro, ou ser atribuído na forma de uma escala numérica.

Algumas palavras são fortes indicadores de sentimentos e são chamadas de palavras de sentimentos ou palavras de opinião. Elas estão geralmente relacionadas com sentimentos positivos ou negativos. Por exemplo, *bom*, *maravilhoso* e *excelente* são palavras positivas e *ruim*, *péssimo* e *terrível* são palavras de sentimento negativas.

Além de palavras individualmente, também encontramos frases ou expressões que indicam sentimentos, como por exemplo, *custou os olhos da cara*. Palavras de opinião e expressões indicativas de sentimentos são importantes ferramentas para análise de sentimentos. Listas contendo essas palavras e suas respectivas polaridades são chamadas de léxicos de sentimentos (ou léxicos de opinião) e são muito utilizadas nas tarefas de identificação de sentimentos.

Por último, o quarto e o quinto componentes são o emissor da opinião e o tempo. Esses também necessitam ser extraídos e categorizados de acordo com sua relação para com a entidade ou aspecto. O emissor pode ser a pessoa ou organização que expressou tal opinião. Para sites de avaliação de produtos e blogs os emissores geralmente são os autores dos textos.

Aplicações de análise de sentimentos baseadas nessas tarefas são chamadas de análise de sentimentos baseada em aspectos ou mineração de opinião (HU; LIU, 2004a; LIU; HU; CHENG, 2005). Apresentaremos, no próximo capítulo, trabalhos relacionados a essas tarefas que serviram como base para essa dissertação.

Trabalhos Relacionados

Dado que a análise de sentimentos baseada em aspectos é uma área que envolve várias tarefas de processamento de linguagem natural, selecionamos trabalhos que abordam tópicos específicos na grande área de processamento de textos, como também alguns que abordam o tema em sua totalidade.

Análise de Sentimentos Baseada em Aspectos

Nessa seção apresentamos alguns trabalhos que tratam do tema análise de sentimentos baseada em aspectos.

Ioannis Pavlopoulos e Ion Androutsopoulos (PAVLOPOULOS, 2014; PAVLOPOULOS; ANDROUTSOPOULOS, 2014) apresentam, de forma original e bastante didática, as tarefas envolvidas na análise de sentimentos baseada em aspectos. O processamento total foi dividido em três tarefas principais: extração de termos aspectos, agregação de termos aspectos e estimação da polaridade dos termos aspecto. A primeira etapa trata da detecção de termos simples ou compostos que identifiquem os aspectos das entidades discutidas no texto. A segunda etapa realiza a agregação dos aspectos identificados, que é o que entendemos pelo processo de junção de aspectos semanticamente iguais, mas que muitas vezes aparecem como termos diferentes nos textos. Na última etapa, estima-se o sentimento por aspecto ou grupos de aspectos.

Os autores apresentaram em seu trabalho novas bases de dados com textos críticos sobre restaurantes, hotéis e laptops devidamente classificados contendo marcação dos aspectos e sentimentos encontrados. Eles também discutiram métricas de avaliação de resultados, propondo outras medidas que acreditam ser mais apropriadas para trabalhar com identificação de aspectos.

Trata-se de importante trabalho da área, dado que além de apresentar essa nova proposta de divisão de tarefas, foram propostos novos métodos (ou melhorias em relação aos métodos anteriores), mostrando experimentalmente nos conjuntos de dados criados

que os novos métodos (ou suas versões melhoradas) são melhores ou, pelo menos, comparáveis aos do estado da arte.

Não tão recente, mas reconhecido como importante trabalho em mineração de opiniões, foi o trabalho desenvolvido por Minqing Hu e Bing Liu (HU; LIU, 2004b). Eles desenvolveram um sistema de sumarização de análises de produtos através da identificação dos aspectos mencionados do produto e contabilização das análises positivas e negativas para cada aspecto.

O primeiro passo do trabalho foi a etiquetagem morfológica objetivando encontrar substantivos, dado ser difícil encontrar características de produtos classificadas em outras categorias morfológicas. Em uma segunda etapa, foi realizada a identificação de expressões candidatas a aspectos compostos, formadas por dois ou três termos, através da verificação de frequência da ocorrência conjunta desses termos. A terceira etapa efetuou a exclusão de termos compostos que aparentam não possuir sentido e remoção de aspectos redundantes. A quarta etapa tratou da identificação daquilo que o autor chamou de palavras de opinião (*opinion words*) através da identificação dos adjetivos localizados próximos aos aspectos encontrados nas etapas anteriores. Por fim, para a identificação do sentimento envolvido, foi feita uma análise semântica das palavras de cada sentença utilizando para isso o léxico *WordNet*¹. Como experimento, os autores trabalharam com análises de cinco produtos eletrônicos coletados através do site da Amazon.com² e c|net³, para os quais obtiveram uma média para a revocação de 80% e cerca de 72% na precisão final do processo de sumarização. Os autores concluíram que as técnicas são promissoras e poderiam ser utilizadas em situações reais.

Hridoy *et al.* (HRIDOY et al., 2015) desenvolveram trabalho analisando *tweets* sobre o celular *iPhone 6* com o objetivo de descobrir a opinião de públicos específicos sobre o produto. A coleta dos *tweets* foi realizada utilizando-se o termo “*iPhone 6*” combinado com alguns aspectos como: “camera”, “iOS”, “iTunes”, “screen”, “sound” e “touch”. Para cada *tweet* também foi armazenado o nome do usuário e a localização de onde a mensagem foi postada. Após a coleta, foram removidas mensagens que não fossem relevantes para o estudo, ou seja, textos não opinativos. Para essa tarefa utilizaram a ferramenta *Stanford Natural Language Processing*, que possibilita, entre outras tarefas, a detecção de diversas relações gramaticais entre as palavras dentro de uma frase. Através da análise dessas relações foi possível selecionar somente os *tweets* opinativos. Para a identificação da polaridade, foi utilizado método baseado em léxico de sentimentos aplicado com o dicionário *SentiWordNet*⁴. Para obter o gênero dos usuários foi utilizada uma ferramenta

¹ <https://wordnet.princeton.edu/>

² <https://www.amazon.com/>

³ <http://www.cnet.com/>

⁴ <http://sentiwordnet.isti.cnr.it/>

para mineração de dados *NamSor*⁵. Finalizando o mapeamento das informações, os dados foram armazenados em banco de dados possibilitando o cruzamento dos dados e a visualização, por exemplo, das médias de sentimentos por característica, por cidade, por gênero, por idade, entre outros. O estudo demonstrou ser possível determinar a popularidade/opinião/sentimento sobre um produto em diferentes localizações e entre homens e mulheres através da rede social *Twitter*.

Uma abordagem para análise de sentimentos baseada na criação de ontologias foi apresentada por Kontopoulos *et al.* (KONTOPOULOS *et al.*, 2013). Uma ontologia pode ser definida com uma “explícita, legível por máquina, especificação de uma conceitualização compartilhada” (STUDER; BENJAMINS; FENSEL, 1998). Ontologias podem ser utilizadas para modelar termos e suas relações dentro de um domínio de interesse. O objetivo dos autores foi criar um sistema que tenha como entrada um conjunto de tweets sobre determinado assunto e, a partir destes calcular pontuações de sentimentos para todos aspectos mencionados.

A primeira etapa do processamento é a construção da ontologia, podendo essa ser desenvolvida de duas maneiras: a) *Formal Concept Analysis* (FCA), e; b) *Ontology learning*. FCA é uma teoria para análise matemática de dados, normalmente utilizada em Representação do Conhecimento e Gestão da Informação. Sua principal característica é que ela apresenta uma metodologia passo a passo e dirigida pelo usuário para a criação de modelos de domínios. O algoritmo desenvolvido recebe como parâmetro um conceito inicial determinado pelo usuário, realiza uma coleta de *tweets* e a partir desses dados faz um mapeamento dos objetos e seus respectivos atributos. O método *Ontology learning* foi baseado na ferramenta *OntoGen*⁶, que é um editor de ontologias semiautomático e orientado a dados que por meio de técnicas de mineração de dados e de sua interface com o usuário possibilita a redução do tempo de desenvolvimento e da complexidade da criação de ontologias. A ontologia passou então por um processo de enriquecimento semântico, no qual, por meio da base léxica *WordNet*⁷, foram adicionados sinônimos e hipônimos dos termos encontrados.

Por fim, para a análise de sentimentos, os autores utilizaram o serviço web *OpenDover*⁸, uma ferramenta com arquitetura orientada a ontologias que localiza em textos de um mesmo domínio opiniões e sentimentos mensurando sua intensidade. Como experimento foi realizada a análise de *tweets* relacionados a quatro modelos de *smartphones*. O resultado final foi a atribuição de uma pontuação de sentimento para cada aspecto encontrado. Trata-se de abordagem diferente da comumente utilizada, tendo a criação de ontologias a base para a detecção dos aspectos e sentimentos.

⁵ <https://github.com/namsor/namsor-api>

⁶ <http://ontogen.ijs.si/>

⁷ <https://wordnet.princeton.edu/>

⁸ <http://opendover.nl/>

Babour e Khan (BABOUR; KHAN, 2014) trabalharam no tema de análise de sentimentos com foco na plataforma *Twitter*. Eles desenvolveram método de análise de sentimentos que considera o contexto no qual as palavras estão inseridas, o que muitas vezes gera diferentes interpretações na análise usual de sentimentos que não considera o contexto. Muitos dos processos de análise de sentimentos são implementados com base em léxicos de polaridade, no entanto muitas palavras-chaves podem ser bipolares. Por exemplo, a palavra ‘barato’, se estiver se referindo ao preço apresenta uma polaridade positiva, enquanto, se estiver referindo a qualidade de um material passa a ter uma polaridade negativa. Utilizando um conjunto de treinamento, o sistema localiza palavras que tenham sido classificadas erroneamente pelo classificador léxico quando se referem a um determinado aspecto, criando assim novas regras de classificação para essa palavra quando associada ao aspecto identificado. Em experimento realizado observou-se uma melhoria considerável nos resultados, comprovando-se assim a importância da análise dessas palavras-chaves dentro do contexto a qual pertencem.

Identificação de Sentimentos

Nessa seção apresentamos trabalhos relacionados apenas a identificação dos sentimentos encontrados em textos.

Vosoughi, Zhou e Roy (VOSOUGHI; ZHOU; ROY, 2015) desenvolveram trabalho científico utilizando técnica chamada de supervisão distante. Neste trabalho eles coletaram cerca de 18 milhões de *tweets* contendo *emoticons* (conjuntos de símbolos que representam sentimentos) relacionados a sentimentos positivos e negativos, possibilitando assim a identificação de suas polaridades e criação de um grande conjunto de treinamento para aplicação de técnicas de aprendizado de máquina supervisionado. Através desses conjuntos, criaram um classificador que obteve uma precisão de 86,2% na análise de sentimentos. Apesar de tratar de método supervisionado de análise, é um trabalho interessante pois utiliza técnicas que dispensam a classificação manual de textos, principal dificuldade na utilização de métodos supervisionados.

O grupo formado por Liu, Li e Guo (LIU; LI; GUO, 2012) desenvolveu trabalho utilizando aprendizado de máquina, integrando no conjunto de treinamento textos manualmente classificados e dados analisados através de *emoticons*. Os autores trabalharam com modelos de linguagens probabilísticos. Um conjunto de *tweets* manualmente classificados como positivos foram reunidos em um único documento e o mesmo foi feito com os negativos, assim gerando dois modelos de linguagem. Para a classificação, cada *tweet* foi comparado com os dois modelos, gerando uma probabilidade de acordo com a semelhança entre o *tweet* e o modelo, a maior probabilidade calculada identifica a classe à qual o texto pertence. Trabalho similar foi feito com os *emoticons*, em que *tweets* contendo o *emoticon*

‘:’) formaram um modelo de linguagem positivo e *tweets* contento ‘:(’ formaram um modelo negativo. Além da classificação com relação a polaridade, foi realizada uma classificação entre *tweets* subjetivos, mensagens que contém algum sentimento, e objetivos, mensagens nas quais não foram encontrados sentimentos. Os experimentos realizados mostraram um aumento na precisão dos resultados obtidos com o novo método tanto na análise de polaridade quanto na análise de subjetividade.

Kolchyna *et al.* (KOLCHYNA et al., 2015) desenvolveram trabalho utilizando duas abordagens para análise de sentimentos: uma baseada em léxico de sentimentos e outra utilizando aprendizado de máquina. Eles realizaram estudo comparativo com várias combinações de léxicos e técnicas de enriquecimento semântico desses léxicos através do uso de *emoticons*, abreviaturas e gírias comuns nas redes sociais, buscando assim melhorar a precisão na classificação de textos obtidos através da rede social *Twitter*. As abordagens baseadas em léxicos calculam o sentimento de um texto através da polaridade de suas palavras ou frases, utilizando para isso dicionário contendo listas de palavras e respectivas polaridades. O cálculo da chamada pontuação de sentimento é realizado através da soma das polaridades de cada palavra encontrada no texto. Na classificação textual utilizando aprendizado de máquina, o algoritmo analisa um conjunto de dados previamente classificado como positivo, negativo ou neutro, extrai as características que modelam as diferenças entre as diferentes classes e assim infere uma função que será utilizada para classificar novos textos. Trata-se de trabalho interessante dado que compara as abordagens supervisionadas com não supervisionadas nos experimentos realizados.

Um sistema híbrido para análise de sentimentos em *tweets* foi proposto por Filho e Pardo (FILHO; PARDO, 2013). O sistema trabalha com uma sequência de classificadores, em que os textos que não atingem uma pontuação pré-determinada na classificação são analisados pelo próximo classificador. Em uma primeira etapa, foi feita a normalização dos *tweets* e uma classificação através de regras relacionadas à presença de *emoticons* no texto. O segundo analisador aplicado foi um classificador léxico, que fez a classificação dos textos através da soma dos valores de polaridade de suas palavras, obtidos através de um léxico de sentimentos. O último classificador aplicado foi um baseado em aprendizado de máquina que, usando exemplos manualmente classificados, aprende como classificar novos textos. Os experimentos realizados demonstraram melhorias nos resultados quando comparados com os obtidos por sistema anteriormente desenvolvido pelo mesmo grupo, indicando ser uma boa opção a utilização de sistemas híbridos.

Kouloumpis, Wilson e Moore (KOULOUMPIS; WILSON; MOORE, 2011) investigaram a utilização de características linguísticas específicas da Internet na detecção de sentimentos em mensagens do *Twitter*. Para isso utilizaram abordagem baseada em aprendizado de máquina supervisionado. Na criação dos conjuntos de treinamento os autores

utilizaram *tweets* com *hashtags* extraídos do *Edinburgh Twitter Corpus*⁹, esse conjunto foi denominado HASH. Outro conjunto foi criado com base em *emoticons*, com dados que foram obtidos através do site *twitter sentiment*¹⁰, e foi chamado de EMOT.

Em um primeiro experimento foi comparado a utilização de um conjunto HASH com outro formado pela união dos conjuntos HASH e EMOT, comprovando uma melhoria nos resultados quando utilizado o segundo conjunto. Os autores também realizaram experimentos utilizando recursos léxicos obtidos a partir do léxico de subjetividade MPQA¹¹, etiquetagem morfológica e características de linguagem da Internet obtidas através de dicionários de gírias como o *Intent Lingo Dictionary*¹², sendo que a combinação do léxico de subjetividade com a linguagem da Internet obteve os melhores resultados. Apesar de utilizar método supervisionado, os conjuntos de treinamento foram criados com bases de dados já existentes, dispensando assim a necessidade de classificação manual de textos e tornando a abordagem mais atraente.

Rossi, Lopes, Faleiros e Rezende (ROSSI et al., 2014) propuseram um algoritmo para classificação de textos baseado em Redes Complexas. Os textos foram representados na forma de uma rede heterogênea bipartida, composta de um lado pelos documentos e do outro conectado a esses documentos, os termos encontrados. Esse tipo de rede é relativamente fácil de ser construída, podendo ser utilizada para modelar qualquer tipo de coleção de textos. O objetivo do algoritmo é induzir pesos dos termos para as classes, utilizando para isso documentos já classificados, e utilizar esses pesos como modelo de classificação. Os resultados foram superiores a algoritmos comumente utilizados na classificação de texto e a outros algoritmos indutivos considerados como estado da arte.

Vários métodos para análise de sentimentos no *Twitter* foram comparados por Psomakelis et al. (PSOMAKELIS et al., 2014). Através de experimentos utilizando combinação das abordagens *bag-of-words*, n-gramas e grafos de n-gramas, em conjunto com metodologias baseadas em léxicos de sentimentos e algoritmos de aprendizado de máquina, o estudo concluiu que existe uma superioridade na combinação de métodos com grafos de n-gramas para análise de *tweets*, em relação aos métodos baseados em dicionários. Isso pode ser explicado pelo fato dos usuários do *Twitter* utilizarem uma quantidade significativa de abreviações e gírias da Internet em seus *posts*. O estudo também mostrou que a combinação de métodos de processamento de linguagem natural e algoritmos de máquina podem melhorar muito os resultados na análise de sentimentos.

Outro trabalho importante neste contexto foi o de Saif, He e Alani (SAIF; HE; ALANI, 2012). A limitação de 140 caracteres do *Twitter* leva a uma grande variedade de termos escritos em formas curtas e irregulares. Para contornar esse problema de disper-

⁹ <http://demeter.inf.ed.ac.uk>

¹⁰ <http://twittersentiment.appspot.com>

¹¹ http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

¹² <http://www.netlingo.com/>

são nos dados, os autores propuseram um método de interpolação semântica que extrai conceitos semânticos ocultos dos *tweets* e os incorpora junto ao conjunto de treinamento, diminuindo assim o tamanho do vocabulário a ser processado e melhorando os resultados na detecção de sentimentos. A escrita irregular também é problema comum em textos opinativos e o método proposto lidou muito bem com isso.

Léxicos de Sentimentos

Uma importante referência que utiliza léxicos de sentimentos é o trabalho de Taboada *et al.* (TABOADA *et al.*, 2011). Os autores apresentaram um sistema chamado de SO-CAL (Semantic Orientation CALculator), que calcula a polaridade de um texto utilizando dicionários de palavras com suas respectivas orientações semânticas (polaridade e força). O sistema também identifica palavras chamadas de modificadores, que intensificam, invertem ou neutralizam essas polaridades. A principal conclusão do trabalho foi que os métodos baseados em léxicos para análise de sentimentos são robustos, conseguindo bons resultados independentemente do domínio. Também podem ser facilmente aprimorados com outras fontes de conhecimentos, sem a necessidade de alterações no sistema ou classificação manual de textos.

No trabalho de Souza *et al.* (SOUZA *et al.*, 2011) os autores apresentam e avaliam diversos recursos e técnicas linguísticas para a criação de um léxico de sentimentos para a língua portuguesa. Um primeiro léxico foi criado utilizando a análise de um corpus anotado. Outro foi criado através da busca de antônimos e sinônimos de uma lista inicial de palavras. Por fim, um último léxico foi criado através da tradução automática do *Lius English Lexicon* (HU; LIU, 2004b). Esses três léxicos foram combinados dando origem ao *OpLexicon*. Experimentos foram realizados comparando os léxicos individualmente, o léxico combinado e outros léxicos existentes. Os resultados indicaram que a técnica de combinação de léxicos foi realmente efetiva.

Silva *et al.* (SILVA *et al.*, 2010) apresentaram proposta de método para expansão de um léxico especializado na mineração de julgamentos sociais em textos. Inicialmente os autores criaram um dicionário de adjetivos com suas respectivas polaridades. Com base nesse dicionário, criaram manualmente um conjunto de regras léxico sintáticas. Essas regras foram aplicadas em uma grande coleção de textos em português obtidos na internet, com o objetivo de aumentar essa lista de adjetivos. Por último essa lista foi novamente expandida com o uso de grafos de sinônimos. Ao novo léxico criado foi dado o nome de *SentiLex*. Os autores acreditam, que apesar de serem necessários mais testes, o método deve ser efetivo se aplicado para construção de outros tipos de léxicos ou idiomas.

Filho, Pardo e Aluísio (FILHO; PARDO; ALUÍSIO, 2013) desenvolveram trabalho avaliando a versão brasileira do dicionário para análise de sentimentos *LIWC* (*Linguistic*

Inquiry and Word Count). Essa versão foi criada por três equipes de tradução a partir do dicionário original em inglês. O estudo avaliou o léxico obtido comparando-o com os léxicos *SentiLex* e *OpLexicon*, alcançando resultados superiores na análise de sentimentos realizada em frases.

Na Tabela 1 elencamos os principais assuntos abordados nos artigos pesquisados nesta revisão bibliográfica.

Tabela 1 – Conteúdo abordado na fundamentação teórica.

Artigos	Assuntos Abordados		
	Aspectos	Sentimentos	Léxicos
(PAVLOPOULOS, 2014)	X	X	
(PAVLOPOULOS; ANDROUTSOPOULOS, 2014)	X	X	
(HU; LIU, 2004b)	X	X	X
(HRIDOY et al., 2015)	X	X	X
(KONTOPOULOS et al., 2013)	X	X	
(BABOUR; KHAN, 2014)	X	X	X
(VOSOUGHI; ZHOU; ROY, 2015)		X	
(LIU; LI; GUO, 2012)		X	
(KOLCHYNA et al., 2015)		X	X
(FILHO; PARDO, 2013)		X	X
(KOULOUMPIS; WILSON; MOORE, 2011)		X	X
(ROSSI et al., 2014)		X	
(PSOMAKELIS et al., 2014)		X	X
(SAIF; HE; ALANI, 2012)		X	
(TABOADA et al., 2011)		X	X
(SOUZA et al., 2011)			X
(SILVA et al., 2010)			X
(FILHO; PARDO; ALUÍSIO, 2013)			X

Concluindo, temos como principais referências relacionadas a essa dissertação, os trabalhos de Pavlopoulos e Androutsopoulos (PAVLOPOULOS, 2014; PAVLOPOULOS; ANDROUTSOPOULOS, 2014) por apresentarem a análise de sentimentos baseada em aspectos como um todo, ou seja, envolvendo todas as suas fases. Seguimos a mencionada divisão de etapas apresentada por estes autores e os métodos relacionados a identificação de aspectos. Outro trabalho referencial importante foi o de Taboada (TABOADA et al., 2011). Este trabalho está relacionado a utilização de léxicos de sentimentos na etapa de identificação de sentimentos. Seguimos no próximo capítulo com a explicação dos métodos e experimentos da etapa de identificação de aspectos.

Identificação de Aspectos

4.1 Introdução

Conforme apresentado anteriormente e seguindo a divisão apresentada por Pavlopoulos (PAVLOPOULOS, 2014), umas das tarefas da análise de sentimentos baseada em aspectos é a identificação de aspectos. Essa tarefa realiza a identificação das características de uma entidade que pode ser um produto, serviço, tópico, questão, pessoa, organização ou evento expressos em um texto.

Os métodos para identificação de aspectos, geralmente, são aplicados em análises de textos relacionados a opiniões sobre produtos ou serviços. Para avaliar esse métodos são utilizados conjuntos de textos organizados na forma de corpus anotado, ou seja, além das sentenças que compõem as opiniões possuem a marcação dos aspectos mencionados e sentimentos relacionados, facilitam o desenvolvimento e avaliação de métodos para extração de aspectos. Temos vários trabalhos que apresentaram e examinaram diferentes conjuntos de dados referentes a produtos eletrônicos (HU; LIU, 2004a; POPESCU; ETZIONI, 2007; DING; LIU; YU, 2008), restaurantes (GANU; ELHADAD; MARIAN, 2009), hotéis (PAVLOPOULOS; ANDROUTSOPOULOS, 2014) entre outros. Os trabalhos citados têm em comum o fato de utilizarem textos em inglês.

Em português, os trabalhos são menos abundantes. Temos algumas análises baseadas em corpus anotado obtido a partir do site de avaliações Buscapé¹ (HARTMANN et al., 2014) e mais recentemente, sobre avaliações de livros através do corpus ReLi (REsenhas de Livros) (FREITAS et al., 2012).

Apresentamos no 17º Workshop de Informática Médica (WIM 2017) um experimento inicial sobre identificação de aspectos em *tweets* relacionados a dengue, chikungunya, zika e microcefalia (MACHADO et al., 2017) e no 14º Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2017) trabalho referente ao experimento

¹ <https://www.buscapi.com.br/>

aqui apresentado (MACHADO; RUIZ; PARDO, 2017).

4.2 Dados Analisados

O corpus ReLi (FREITAS et al., 2012) foi criado através do projeto Anotadores Semânticos, coordenado pelo Professor Ruy Milidiú do Departamento de Informática da PUC do Rio de Janeiro². O ReLi consiste em 1.600 resenhas de livros anotadas manualmente quanto à presença de opinião, identificação do aspecto e sua respectiva polaridade. Essas resenhas foram obtidas a partir do site Skoob³, uma rede social de leitores. Foram coletados dados referentes a 13 livros de 7 autores diferentes, totalizando cerca de 260 mil palavras e 12 mil frases, conforme apresentado na Tabela 2.

Tabela 2 – Composição do ReLi.

Autor	Título	Resenhas	Frases	Palavras
Stephenie Meyer	Crepúsculo	409	3.266	62.268
Thalita Rebouças	Fala sério, amiga!; Fala sério, amor!; Fala sério, mãe!; Fala sério, pai!; Fala sério, professor!	161	910	16.864
Sidney Sheldon	O Outro lado da meia-noite; O Reverso da Medalha; Se houver Amanhã	230	1.569	31.712
Jorge Amado	Capitães da Areia	187	1.348	32.117
George Orwell	1984	202	2.228	51.320
José Saramago	Ensaio sobre a Cegueira	271	1.991	42.152
J.D. Salinger	O Apanhador nos Campos de Centeio	140	1.158	23.725
Total		1.600	12.470	259.978

Fonte: (FREITAS et al., 2012).

Na marcação do corpus foram consideradas apenas opiniões relativas ao livro ou parte dele, como personagens, capítulos, linguagem, entre outros. As opiniões foram anotadas no nível de frase e do sintagma. A anotação foi realizada em 4 etapas: (i) identificação e anotação das frases que contêm opinião; (ii) identificação do alvo da opinião (aspecto); (iii) identificação e anotação da polaridade do trecho (palavras ou expressões) relacionado ao aspecto, sendo consideradas apenas opiniões positivas ou negativas, as opiniões neutras não foram consideradas, e; (iv) identificação e anotação da polaridade das frases que contêm opinião. É importante citar que estas anotações e marcações foram feitas por linguistas, pessoas especializadas que eventualmente podem ter opiniões mais acuradas sobre os termos marcados.

² <http://www.inf.puc-rio.br/>

³ <https://www.skoob.com.br/>

A Tabela 3 demonstra a distribuição de polaridade das opiniões encontradas no ReLi. Das 12.470 frases, cerca de 25% continham opiniões. Dessas, 77% eram positivas, 17% negativas e 5% continham opiniões mistas, isto é, aspectos positivos e negativos em uma mesma frase.

Tabela 3 – Distribuição de polaridades no ReLi.

Dados do ReLi	Quantidade	Porcentagem
Frases opinativas positivas	2.433	77,34 %
Frases opinativas negativas	544	17,29 %
Frases opinativas mistas	169	5,37 %
Total de frases opinativas	3.146	100,00 %
Opiniões positivas sobre aspectos	4.268	80,67 %
Opiniões negativas sobre aspectos	1.023	19,33 %
Total opiniões sobre aspectos	5.291	100,00 %

Fonte: Adaptado (FREITAS et al., 2012)

Com relação aos aspectos, a Tabela 4 apresenta a frequência da ocorrência de aspectos nas frases. A primeira linha mostra a existência muitas frases sem aspectos. A segunda linha demonstra que a maior parte das frases restantes possui apenas um aspecto. A última linha mostra o total de frases que possuem aspectos.

Tabela 4 – Ocorrências de aspectos por frase.

Ocorrência Aspectos	Quantidade	Porcentagem
Frases sem aspectos	10.237	82,16 %
Frases com aspectos	2.233	17,91 %
Total de Frases	12.470	100,00 %
com apenas 1 aspecto	1.818	14,58 %
com 2 ou mais aspectos	415	3,33 %

A Tabela 5 mostra a composição dos aspectos encontrados no corpus. Em sua maioria são aspectos simples, formados por apenas um termo. Temos alguns aspectos compostos com dois ou três termos e uma quantidade significativa de aspectos maiores, composta principalmente pelos títulos dos livros.

Tabela 5 – Ocorrências de aspectos simples e compostos.

Composição Aspectos	Quantidade	Porcentagem
1 termo	2.347	84,73 %
2 termos	102	3,68 %
3 termos	71	2,56 %
mais que 4 termos	250	9,03 %
Total de Aspectos	2.770	100,00 %

Alguns dos nossos experimentos utilizaram outras resenhas dos mesmos livros que compuseram o corpus ReLi, com o objetivo de testar os algoritmos em conjuntos maiores

de dados. Para isso coletamos novas resenhas a partir do site Skoob⁴. Esse conjunto se caracterizou como um corpus adicional, ao qual atribuímos o nome de ReLi estendido, e sua composição é demonstrada na Tabela 6.

Tabela 6 – Composição do ReLi estendido.

Autor	Títulos	Resenhas	Frases	Palavras
Stephenie Meyer	Crepúsculo	3.733	24.163	432.875
Thalita Rebouças	Fala sério, amiga!; Fala sério, amor!; Fala sério, mãe!; Fala sério, pai!; Fala sério, professor!	298	1.536	26.643
Sidney Sheldon	O Outro lado da meia-noite; O Reverso da Medalha; Se houver Amanhã	519	3.763	65.539
Jorge Amado	Capitães da Areia	496	4.342	101.704
George Orwell	1984	646	7.872	163.780
José Saramago	Ensaio sobre a Cegueira	506	4.675	96.047
J.D. Salinger	O Apanhador nos Campos de Centeio	500	4.797	100.052
Total		6.698	51.148	980.640

4.3 Métodos para Identificação de Aspectos

A seguir são descritos os cinco métodos para identificação de aspectos implementados neste trabalho.

FREQ Baseline

Este é, provavelmente, um dos métodos mais simples utilizados para identificação de aspectos. É baseado no cálculo da frequência das ocorrências de substantivos e sintagmas nominais. Após encontrar os termos mais frequentes de um texto, é executado um mecanismo de poda que mantém apenas aqueles substantivos ou sintagmas nominais que tenham uma frequência superior a um valor pré-estipulado, evitando assim que todos os substantivos e sintagmas sejam identificados como aspectos. Apesar de simples, é um método razoavelmente efetivo e muito popular como base de comparação (HU; LIU, 2004a; WEI et al., 2010; LIU, 2012).

⁴ <https://www.skoob.com.br/>

Hu & Liu

O algoritmo de Hu & Liu pode ser considerado um método híbrido de detecção de aspectos. Ele combina análise frequencial de substantivos e sintagmas nominais, com regras para detecção de termos compostos e remoção de termos candidatos não relacionados.

Apresenta as seguintes vantagens sobre o método **FREQ**:

- É capaz de detectar termos mais longos, ou seja, compostos por mais de uma palavra;
- Pode identificar termos pouco frequentes, isto é, que apareçam poucas vezes no texto, e;
- Possui mecanismos de correção que buscam eliminar termos que tenham sido erroneamente identificados como aspectos.

O método de Hu & Liu pode ser resumido nos seguintes oito passos:

1. Detecção de termos frequentes: após análise morfológica dos termos que compõem as frases, os substantivos e sintagmas nominais são ordenados de acordo com sua frequência, formando uma lista de termos candidatos. Dessa lista, termos que possuam uma frequência inferior a um valor pré-determinado são removidos. Em nossa implementação variamos essa frequência de corte entre 1 e 10% buscando encontrar um valor de poda mais adequado;
2. Identificação de termos compostos: bigramas e triplas de substantivos devem ser extraídas na mesma ordem em que aparecerem em uma frase. Esses termos são incluídos na lista de termos candidatos. Por exemplo, a palavra ‘campo’ pode aparecer muitas vezes no texto, mas ela faz parte da expressão “O apanhador no campo de centeio”. Dessa forma, além da palavra ‘campo’, que já aparece na lista devido a sua frequência, a tripla ‘apanhador campo centeio’ também é incluída como termo candidato;
3. Cálculo do *p*-suporte (*pure support*): o *p*-suporte de um termo candidato *t* se refere a quantidade de sentenças que contenham *t*. Nesse cálculo não são consideradas frases que contenham um candidato *t'*, em que o candidato *t* seja um subconjunto de *t'*. Seguindo o exemplo anterior, teríamos na lista de termos candidatos: ‘apanhador’, ‘campo’, ‘centeio’, ‘apanhador campo’, ‘campo centeio’ e ‘apanhador campo centeio’. Nas frases onde a expressão “O apanhador no campo de centeio” aparecer, somente o *p*-suporte da tripla ‘apanhador campo centeio’ seria incrementado, dado que os outros termos são subconjuntos deste;

4. Cálculo do índice de compacidade: em uma sentença que contenha termos candidatos compostos é verificada a distância entre cada par de termos do candidato. Se essa distância for maior do que três palavras esse índice é incrementado;
5. Poda por compacidade: termos que possuam um índice de compacidade maior do que 1, ou seja, que tenham aparecido mais de uma vez em uma forma não compacta, são removidos;
6. Poda por p -suporte: candidatos que tenham um índice de p -suporte menor do que 3, o que é um indicativo de que outra forma que contém o candidato seja a mais frequente, também são removidos;
7. Extração de palavras opinativas: adjetivos encontrados próximos a um termo candidato são adicionados a uma lista de palavras opinativas;
8. Detecção de termos infrequentes: as frases onde não foram encontrados termos candidatos são reanalisadas em busca de palavras opinativas. Se alguma for encontrada, o substantivo mais próximo a esse adjetivo é adicionado à lista de termos candidatos.

FREQ Baseline + *Word2Vec*

Dado que o FREQ Baseline é um método simplesmente baseado em frequência, muitas vezes termos não relacionados ao assunto em estudo, mas que apareçam muitas vezes no texto, são identificados como aspectos. Para contornar esse problema seguimos o exemplo de Pavlopoulos (PAVLOPOULOS; ANDROUTSOPOULOS, 2014) e implementamos função de detecção de contexto, baseada na representação dos termos em espaço vetorial contínuo, segundo o modelo *Word2Vec* (MIKOLOV et al., 2013). Com essa função foi possível implementar uma etapa de poda, eliminando assim aspectos que não fossem relacionados ao contexto de análise de livros.

Modelo *Word2Vec*

O modelo *Word2Vec* foi desenvolvido por Mikolov e seus companheiros do Google (MIKOLOV et al., 2013), buscando ser um método eficiente para criação de modelos de representação de palavras através de vetores, a partir de grandes conjuntos de dados. O método foi baseado em modelos anteriores de representação de palavras, tais como o modelo de espaço vetorial, em que palavras são representadas por densos vetores em um contínuo espaço vetorial. No modelo *Word2Vec*, os vetores são representações semânticas das palavras e são tratados como variáveis latentes, cada um representando uma dimensão do significado da palavra. Nessa representação, palavras semanticamente similares possuem representações vetoriais próximas.

Antes do modelo *Word2Vec*, a criação dos vetores baseava-se nas estatísticas de coocorrência das palavras com suas vizinhas em um grande corpus de texto. De uma maneira diferente, o *Word2Vec* utiliza um modelo de linguagem preditivo, obtido através de uma arquitetura de redes neurais, mas construído de forma a ser computacionalmente muito eficiente. Modelos preditivos tentam prever uma palavra a partir de seus vizinhos baseados nos vetores aprendidos.

Formalmente, esses modelos probabilísticos de linguagem são treinados utilizando o princípio da máxima verossimilhança com objetivo de maximizar a probabilidade da próxima palavra, w_{t+j} , em uma vizinhança $|c|$, dadas as palavras anteriores w_j . Dada certa sequência de palavras para treinamento $(w_1, w_2, w_3, \dots, w_T)$ em um corpus com T palavras, o objetivo do modelo *Word2Vec* é maximizar a seguinte média do log de probabilidade:

$$J_{ML} = \frac{1}{T} \sum_{t=1}^T \sum_{\substack{-c < j < c, \\ j \neq 0}} \log p(w_{t+j} | w_j)$$

Na fórmula acima, c se refere ao tamanho do contexto de treinamento para a palavra central w_t . Valores altos de c resultam em mais amostras de treinamento, e como consequência, tendem a criar um modelo mais preciso. Em outras palavras o objetivo desse método é tentar correlacionar palavras em um vocabulário medindo sua compatibilidade através valores de probabilidades.

O vetor de cada termo é construído através do treinamento de um modelo de linguagem probabilístico obtido através de um corpus. Para textos em português, utilizamos um modelo treinado por Hartmann *et al.* (HARTMANN et al., 2017). Foram usados na construção desse modelo textos obtidos a partir da Wikipédia, artigos científicos, textos jornalísticos, palestras, legendas de filmes, textos técnicos e outros documentos variados.

Identificação de Contexto com *Word2Vec*

O objetivo dessa ferramenta é verificar se determinada palavra é mais próxima de um contexto específico ou de um contexto comum. A intenção é filtrar palavras que pertençam a esse contexto comum, ou seja, que não estejam relacionadas ao contexto em estudo, no nosso caso as análises literárias.

Para isso selecionamos algumas palavras para representar cada contexto. Para representar o universo literário, escolhemos os 10 substantivos mais frequentes do ReLi. A alta frequência desses termos remete a ideia de que eles sejam os mais representativos desse contexto.

O próximo passo foi encontrar termos que representem um contexto mais comum. Com esse objetivo, buscamos os substantivos mais frequentes em um corpus que representasse a língua portuguesa como um todo. Efetuamos alguns testes com os corpora de

textos jornalísticos Floresta e Mac Morpho, e outro com as obras completas de Machado de Assis. O corpus de Machado de Assis mostrou um desempenho melhor, dado que as palavras identificadas eram visivelmente mais próximas ao vocabulário utilizado no cotidiano do que as obtidas dos conjuntos jornalísticos. Para representar o contexto comum selecionamos os 20 substantivos mais frequentes das obras de Machado de Assis. Essa diferença na quantidade se deve à intuição de que o contexto comum seja mais abrangente que o específico, necessitando assim de mais termos para representá-lo. A Tabela 7 mostra os termos mais frequentes encontrados em cada corpus.

Tabela 7 – Termos mais frequentes dos corpus analisados.

Corpus	Termos encontrados
ReLi	livro, história, leitura, personagens, romance, personagem, livros, crepúsculo, final, vida
Machado de Assis	casa, tempo, coisa, olhos, dois, homem, vida, duas, anos, moça, amor, dias, noite, coração, verdade, coisas, vezes, mulher, palavras, idéia
Floresta	nacional, passado, política, internacional, passada, mundial, brasileira, social, público, portuguesa, português, político, especial, anterior, local, eleitoral, pública, total, final, comercial
Mac Morpho	editoria, página, Paulo, Brasil, anos, governo, presidente, milhões, dois, país, três, mercado, pessoas, folha, tempo, dias, parte, programa, empresa, trabalho

Cada palavra foi, então, convertida em seu vetor correspondente através do modelo de linguagem *Word2Vec* já mencionado (HARTMANN et al., 2017). Calculamos a centroide dos 10 vetores relacionados ao contexto especializado e dos 20 vetores do contexto comum. Como resultado obtivemos dois vetores: um de centroide comum e outro de centroide especializado, ou específico.

Para saber se determinado termo pertence ao contexto literário, inicialmente recupera-se seu vetor no modelo de linguagem. O vetor encontrado deve ser comparado com ambos vetores de centroide, o comum e o especializado. Para isso utilizamos a tradicional medida de distância por cosseno. Se o menor resultado encontrado para essa medida for em relação ao centroide do vetor específico, o termo é mantido, pois isso é indicativo que ele é pertencente ao contexto literário. Por outro lado, se a menor medida for em relação ao centroide do vetor comum, o termo é descartado.

Hu & Liu + *Word2Vec*

Apesar do Hu & Liu ser um método mais elaborado que o **FREQ**, esse ainda retorna uma grande quantidade de termos não relacionados ao contexto em estudo. Utilizamos então

a mesma técnica de detecção do contexto aplicada no **FREQ** com objetivo de remover esses termos não relacionados.

A detecção de contexto foi utilizada logo no primeiro passo do método, após a realização da poda frequencial, como foi sugerido por (PAVLOPOULOS; ANDROUTSOPOULOS, 2014). Termos que não foram cortados por sua frequência, foram então analisados com a ferramenta de detecção de contexto baseada em *Word2Vec* apresentada anteriormente. Termos não pertencentes ao contexto em estudo, no caso, relacionados ao universo literário, foram então removidos.

Hu & Liu + *Word2Vec* nos Termos Infrequentes

Observamos que apesar da poda utilizando *Word2Vec* no método anterior, alguns termos não relacionados ainda estavam aparecendo na lista final de aspectos detectados. A maior parte desses termos resultaram do último passo do algoritmo, a detecção de termos infrequentes. Modificamos então essa etapa para que nela também fosse realizada a análise de contexto em seus resultados.

Cada termo substantivo localizado próximo a uma palavra opinativa, teve seu contexto analisado com o *Word2Vec*. Caso o termo pertencesse ao contexto literário, ele seria adicionado à lista de termos candidatos. Já se o termo fosse relacionado ao contexto comum, o algoritmo seguiria em busca do próximo substantivo mais próximo da palavra opinativa em análise. Essa análise de contexto foi realizada novamente e o processo repetido até que fosse encontrado um termo que pertencesse ao contexto literário. Se nenhum termo fosse encontrado nessas condições, o algoritmo voltaria ao termo mais próximo da palavra opinativa e o adicionaria à lista de termos candidatos.

4.4 Métricas de Avaliação

A avaliação dos métodos foi realizada comparando-se os aspectos identificados com os aspectos de referência anotados no corpus ReLi. Essa metodologia é semelhante à utilizada nos workshops do SemEval. Veja, por exemplo, (PONTIKI et al., 2015). Os aspectos foram extraídos com suas respectivas posições de início e fim na sentença. Foram calculadas a precisão e revocação pelo método da micro média (SEBASTIANI, 2002), além da medida *f*, que é a média harmônica entre precisão e revocação.

Para o cálculo dessas medidas foi necessário obtermos a quantidade de aspectos identificados corretamente (verdadeiros positivos), aqueles que não eram aspectos e dessa forma foram classificados (verdadeiros negativos), os classificados erroneamente como aspectos (falsos positivos) e os erroneamente não identificados como aspectos (falsos negati-

vos). Esses quatro valores são demonstrados na Tabela de contingência 8 (SOKOLOVA; LAPALME, 2009).

Tabela 8 – Tabela de contingência:

Classificação anotada:	Classificado como:	
	Aspecto	Não Aspecto
Aspecto	verdadeiro positivo (vp)	falso negativo (fn)
Não Aspecto	falso positivo (fp)	verdadeiro negativo (vn)

Dessa forma, a precisão (P^μ) e revocação por micro média (R^μ) podem ser encontradas pelo seguinte somatório:

$$P^\mu = \frac{\sum_{c=1}^C vp_i}{\sum_{c=1}^C (vp_i + fp_i)}, \quad R^\mu = \frac{\sum_{c=1}^C vp_i}{\sum_{c=1}^C (vp_i + fn_i)}.$$

Para C a quantidade de c classes a serem avaliadas e i que representa cada elemento individualmente nessas classes.

4.5 Resultados dos Experimentos

Conforme explicado anteriormente, as frases do corpus ReLi foram marcadas desta maneira: (i) com relação a presença de opinião; (ii) aspectos relacionados a essas opiniões; e (iii) identificação da polaridade relacionada aos aspectos identificados. Dado que o objetivo do experimento é a busca pelos aspectos, selecionamos no corpus ReLi apenas os textos opinativos, ou seja, somente aqueles que continham alguma opinião sobre algum aspecto do livro. Caso não fosse realizado esse filtro inicial, seria necessário mais uma etapa anterior a essa buscando identificar os textos opinativos, dado que, como demonstrado anteriormente, existem muitas frases que não expressam opinião dentro do conjunto total de dados.

Realizado esse filtro, restaram 2.333 frases opinativas. Como o conjunto de dados ficou relativamente pequeno, resolvemos testar os algoritmos também nas 51.149 frases do já mencionado ReLi estendido.

É importante ressaltar que, como fase preliminar, ambos conjuntos de dados passaram por etapas de pré-processamento. No caso do ReLi, os dados já vieram divididos em *tokens* e com a classificação morfológica realizada. Mesmo assim descartamos essa classificação e analisamos os textos novamente fazendo a marcação morfológica utilizando o analisador morfológico do módulo *nlpnet* (*Natural Language Processing with neural networks*⁵) (FONSECA; ROSA, 2013).

⁵ <http://nilc.icmc.usp.br/nlpnet/>

Para os dados do ReLi estendido foram necessárias mais etapas de pré-processamento, dado que foram textos coletados diretamente da web. O primeiro passo foi a remoção de tags HTML, processo que foi realizado utilizando o módulo *Beautiful Soup*⁶ da linguagem Python. Posterior a essa etapa foram necessários dois processos de segmentação, o primeiro dividindo os textos em frases e o segundo fazendo a divisão das frases em *tokens*, ambos utilizando o *NLTK* (*Natural Language Toolkit*)⁷. Por último, assim como no ReLi, foi realizada a análise morfológica também utilizando o módulo *nlpnet*.

Como última etapa, foi realizada a identificação dos sintagmas nominais. Para isso foi utilizada a ferramenta *chunker* do *NLTK* treinada com o corpus Floresta Sinta(c)tica. Esse treinamento foi feito utilizando o software *NLTK Trainer*⁸.

Os processamentos de ambos conjuntos de dados foram realizados de acordo com os métodos já apresentados. Segue-se os resultados para cada um.

FREQ Baseline

Conforme explicado anteriormente, o método analisa as frequências dos substantivos e sintagmas nominais e os adiciona a uma lista de termos candidatos. Dessa lista são retirados os termos menos frequentes, ou seja, que possuam frequência inferior a determinado valor, que chamaremos daqui em diante de frequência de corte.

Essa frequência de corte é definida de forma empírica. Hu & Liu, em um de seus trabalhos (HU; LIU, 2004a), utilizaram 1% como frequência de corte, valor que acabou sendo comumente utilizado em outros estudos. Neste trabalho optamos por investigar qual seria a melhor frequência de corte. Executamos o algoritmo com a frequência variando entre 1% e 10%, e analisamos os resultados com as métricas já mencionadas: precisão, revocação e medida f, calculadas por micro média.

Observando a Tabela 9 na página 46, podemos ver que caso utilizássemos a frequência de corte de 1%, nosso resultado seria muito ruim, com uma precisão de somente 20,07%. No outro extremo, com a frequência de corte em 9% e 10%, temos as Precisões mais altas do experimento (66,97%). Acreditamos que esse valor se deve a uma característica do corpus, em que alguns termos aparecem muitas vezes. Nesse caso específico, apenas 2 termos foram identificados como aspectos: ‘livro’ e ‘história’.

Devido a esses fatores, estamos considerando a medida f como uma melhor métrica para avaliar os experimentos, já que ela exprime um equilíbrio entre a precisão e revocação, exatamente por ser uma média harmônica das duas medidas. Nesse primeiro experimento, a maior medida f apareceu com a Frequência de Corte de 5% com um valor de 55,81%,

⁶ <https://www.crummy.com/software/BeautifulSoup/>

⁷ <http://www.nltk.org/>

⁸ <https://github.com/japerk/nltk-trainer>

além de 64,39% para a precisão e 49,25% para a revocação. A Tabela 9 apresenta os resultados desse experimento e o gráfico apresentado na Figura 1 demonstra a relação entre os resultados e respectivas frequências de corte.

Tabela 9 – Resultados FREQ Baseline corpus ReLi.

Frequência de Corte	Precisão	Revocação	Medida F
1%	20,07	39,89	26,71
2%	31,26	43,29	36,31
3%	41,42	48,92	44,86
4%	46,46	40,37	43,20
5%	64,39	49,25	55,81
6%	65,38	47,24	54,85
7%	66,21	44,13	52,96
8%	66,21	44,13	52,96
9%	66,97	40,04	50,11
10%	66,97	40,04	50,11

Valores de precisão, revocação e medida f dados em porcentagem.

Figura 1 – Resultados FREQ Baseline corpus ReLi.

Abaixo listamos alguns exemplos de aspectos identificados corretamente pelo método utilizando a frequência de corte de 5%. Os aspectos marcados no corpus aparecem sublinhados:

- ‘Um livro sensível, intenso e profundo.’: o algoritmo detectou corretamente o aspecto ‘livro’, termo que mais se repete no corpus;

- ‘Em “O outro lado de a meia-noite” o brilhante autor constrói uma história totalmente sólida de amor, ódio e vingança.’: termo ‘história’ detectado corretamente, outro termo muito frequente no corpus;

Alguns exemplos em que o algoritmo falhou:

- ‘Um livro brasileiro que tem crônicas excelentes pra se ler em alguns minutos e logo não querer parar de se ler.’: algoritmo identificou erroneamente o termo ‘livro’ ao invés de ‘crônicas’;
- ‘Tem muitas crônicas agradáveis, muita coisa bacana sobre a amizade, a mãe que faz queimar a nossa cara, amigos com problemas e qualidades.’: nesse caso, o algoritmo não identificou nenhum aspecto, enquanto deveria ter identificado o termo ‘crônicas’ como aspecto;

Realizamos o mesmo experimento utilizando o ReLi estendido. Com ele fizemos a identificação dos aspectos, e através do ReLi a validação, comparando o resultado encontrado com o marcado no corpus, e assim calculando as métricas. Na Tabela 10 observamos que com as frequências de corte 1% e 2% encontramos o mesmo problema observado no experimento anterior, poucos termos aparecendo muitas vezes. Nesse caso, apenas o termo ‘livro’ foi encontrado, e mesmo assim o valor da precisão foi relativamente alto (69,45%). Com frequências de corte maiores, não foram identificados aspectos. Dado o tamanho do corpus, seria necessário que um mesmo aspecto se repetisse muitas vezes para que pudesse ser identificado pelo algoritmo.

Para contornar esse problema mudamos o intervalo da frequência de corte para valores entre 0,1% e 4%. O melhor resultado encontrado, foi no corte realizado em 0,2%, para o qual tivemos uma medida f de 55,81%. Coincidentemente, os mesmos valores observados no melhor resultado do experimento realizado somente com o ReLi. Esses resultados estão apresentados na Tabela 10 e na Figura 2.

Hu & Liu

Como já explicamos, Hu & Liu é um método híbrido de análise, sendo bem mais complexo que o FREQ Baseline. Tem como principais vantagens uma melhor detecção de termos compostos e possui etapas de poda que buscam corrigir seus resultados.

No primeiro experimento os resultados ficaram pouco aquém do esperado, como pode ser observado na Tabela 11. Ainda assim, observando-se somente a Frequência de Corte de 1%, que é comumente utilizada, o resultado (46,44% na medida f) foi bem superior ao obtido com o FREQ para a mesma frequência (26,71%). Também pode-se

Tabela 10 – Resultados FREQ Baseline corpus ReLi estendido.

Frequência de Corte	Precisão	Revocação	Medida F
0,1%	33,15	41,35	36,80
0,2%	64,39	49,25	55,81
0,3%	66,21	44,13	52,96
0,4%	66,97	40,04	50,11
0,5%	66,97	40,04	50,11
0,6%	66,97	40,04	50,11
0,7%	69,45	32,50	44,28
0,8%	69,45	32,50	44,28
0,9%	69,45	32,50	44,28
1%	69,45	32,50	44,28
2%	69,45	32,50	44,28
3%	0,00	0,00	0,00
4%	0,00	0,00	0,00

Valores de precisão, revocação e medida f dados em Porcentagem

Figura 2 – Resultados FREQ Baseline corpus ReLi estendido.

notar que a diferença entre os maiores e menores valores obtidos em cada métrica foram menores que as obtidas no FREQ, o que é um indício de que esse método é menos suscetível às variações na Frequência de Corte.

Ao analisarmos os resultados através da lista de aspectos obtida, encontramos vários sintagmas detectados incorretamente. Testamos o algoritmo sem levar em consideração os sintagmas nominais, e o resultado apresentou uma melhora considerável, como ser observado na Tabela 12. O melhor resultado foi obtido com um corte frequencial de 5%, para o qual obtivemos uma medida f de 57,03%. O efeito nas diferenças entre maiores

Tabela 11 – Resultados Hu & Liu corpus ReLi.

Frequência de Corte	Precisão	Revocação	Medida F
1%	37,79	60,22	46,44
2%	40,35	59,52	48,09
3%	42,66	59,41	49,66
4%	45,38	53,02	48,90
5%	46,34	52,25	49,11
6%	46,07	51,01	48,41
7%	45,41	49,03	47,15
8%	45,41	49,03	47,15
9%	44,64	46,76	45,68
10%	44,64	46,76	45,68

Valores de precisão, revocação e medida f dados em porcentagem.

e menores valores de cada métrica também se repetiu, como pode ser notado através do gráfico apresentado na Figura 3.

Tabela 12 – Resultados Hu & Liu corpus ReLi sem sintagmas

Frequência de Corte	Precisão	Revocação	Medida F
1%	44,80	62,89	52,33
2%	49,81	62,12	55,29
3%	52,11	60,84	56,14
4%	57,73	54,88	56,27
5%	59,84	54,48	57,03
6%	59,93	53,42	56,49
7%	59,33	51,63	55,21
8%	59,33	51,63	55,21
9%	59,07	49,76	54,02
10%	59,07	49,76	54,02

Valores de precisão, revocação e medida f dados em Porcentagem

Segue-se abaixo alguns exemplos de aspectos não tão frequentes, agora identificados pelo método:

- ‘Mas gostei muito do final, compensa.’: o algoritmo detectou corretamente o termo ‘final’ como aspecto;
- ‘A narrativa fica mais divertida a partir daí, se é que é possível.’: a palavra ‘narrativa’ foi detectada corretamente como aspecto;

Exemplos de onde o algoritmo falhou:

- ‘A narrativa é fluida’: o algoritmo não detectou nenhum aspecto;

Figura 3 – Resultados Hu & Liu corpus ReLi sem sintagmas.

- ‘O tipo de história divertida, narrativa empolgante, e aquele jeitinho brasileiro que a gente deveria valorizar mais.’: o algoritmo identificou ‘história’ mas não detectou ‘narrativa’;

Nos testes realizados com o ReLi estendido, os resultados foram parecidos, repetindo-se o problema com os sintagmas. A medida f mais alta obtida foi de 45,68% para uma frequência de corte de 0,5%. Da mesma forma que no experimento anterior, executamos o algoritmo desconsiderando os sintagmas. Os resultados também tiveram significativa melhora. Ainda assim, como era esperado, ficaram um pouco abaixo dos resultados obtidos somente com o ReLi. A melhor medida f (54,02%) foi obtida com uma frequência de corte de 0,5%. Os resultados destes testes encontram-se na seção de anexos.

FREQ Baseline + *Word2Vec*

Buscando eliminar aspectos detectados erroneamente, testamos o método FREQ Baseline com uma etapa de poda baseada na rotina de detecção de contextos. Como explicado anteriormente, termos pertencentes ao contexto literário foram mantidos e aqueles não pertencentes eliminados do conjunto de aspectos.

Nas frequências de corte entre 1% e 3% houve uma significativa melhora nos resultados, quando em comparação com o FREQ, como mostrado na Tabela 13. No experimento anterior com um corte de 1% conseguimos uma precisão de 20,07%, com a poda utilizando

Word2Vec obtivemos 25,87%. A revocação caiu um pouco, mas essa queda foi compensada pela alta na precisão como podemos observar pelo aumento na medida f.

Tabela 13 – Resultados FREQ + *Word2Vec* corpus ReLi.

Frequência de Corte	Precisão	Revocação	Medida F
1%	25,87	37,62	30,66
2%	36,02	42,34	38,92
3%	47,33	48,01	47,67
4%	46,46	40,37	43,20
5%	64,39	49,25	55,81
6%	65,38	47,24	54,85
7%	66,21	44,13	52,96
8%	66,21	44,13	52,96
9%	66,97	40,04	50,11
10%	66,97	40,04	50,11

Valores de precisão, revocação e medida f dados em Porcentagem

Para as frequências de corte acima de 3%, obtivemos os mesmos resultados do FREQ. Isso indica que apesar da simplicidade, o método é realmente capaz de identificar aspectos relacionados somente de acordo com suas frequências. O melhor resultado foi uma medida f de 55,81% obtido com uma Frequência de corte de 5%. O gráfico apresentado na Figura 4 exibe os resultados obtidos.

Figura 4 – Resultados FREQ + *Word2Vec* corpus ReLi.

Nos testes realizados com o ReLi estendido, os resultados foram parecidos. Assim como no experimento do FREQ sem *Word2Vec*, executamos o algoritmo para frequências de corte entre 0,1% e 10%. A poda realizada com auxílio do *Word2Vec* só teve efeito

para o corte realizado em 0,1% aumentando de 36,80% para 39,37% a medida f. Para as outras frequências de corte foram obtidos os mesmos resultados do experimento realizado somente com corpus ReLi, alcançando uma medida f de 55,81% para uma frequência de corte de 0,2%. Os resultados estão apresentados na Tabela 14.

Tabela 14 – Resultados FREQ + *Word2Vec* corpus ReLi estendido.

Frequência de Corte	Precisão	Revocação	Medida F
0,1%	38,35	40,44	39,37
0,2%	64,39	49,25	55,81
0,3%	66,21	44,13	52,96
0,4%	66,97	40,04	50,11
0,5%	66,97	40,04	50,11
0,6%	66,97	40,04	50,11
0,7%	69,45	32,50	44,28
0,8%	69,45	32,50	44,28
0,9%	69,45	32,50	44,28
1%	69,45	32,50	44,28
2%	69,45	32,50	44,28
3%	0,00	0,00	0,00
4%	0,00	0,00	0,00

Valores de precisão, revocação e medida f dados em Porcentagem

Hu & Liu + *Word2Vec*

Em experimento similar ao anterior, adicionamos uma etapa de poda utilizando *Word2Vec* ao método de Hu & Liu. Aplicamos a detecção de contexto na lista de candidatos a aspectos obtida na primeira etapa do método, com o objetivo de eliminarmos termos não relacionados ao contexto literário.

A melhora nos resultados foi realmente significativa com relação ao método sem a poda *Word2Vec*. Obtivemos uma medida f de 59,80% com uma frequência de corte de 3%, como demonstra a Tabela 15. O máximo obtido no experimento anterior havia sido 57,03% para uma frequência e corte de 5%. A Figura 5 apresenta esses mesmos resultados, onde observa-se uma menor variação nos resultados quando comparado ao FREQ Baseline.

Seguem abaixo alguns exemplos de aspectos identificados corretamente:

- ‘Quando eu o li, me identifiquei 100 % com o personagem.’: o termo ‘personagem’ foi detectado como aspecto infrequente corretamente;
- ‘É simplesmente demais a narrativa, e o espírito jovial de o livro.’: o algoritmo foi capaz de detectar os dois aspectos: ‘narrativa’ e ‘livro’;

Exemplos de aspectos identificados erroneamente:

Tabela 15 – Resultados Hu & Liu + *Word2Vec* corpus ReLi.

Frequência de Corte	Precisão	Revocação	Medida F
1%	51,39	64,10	57,05
2%	57,11	61,79	59,36
3%	59,21	60,40	59,80
4%	57,73	54,88	56,27
5%	59,84	54,48	57,03
6%	59,93	53,42	56,49
7%	59,33	51,63	55,21
8%	59,33	51,63	55,21
9%	59,07	49,76	54,02
10%	59,07	49,76	54,02

Valores de precisão, revocação e medida f dados em Porcentagem

Figura 5 – Resultados Hu & Liu + *Word2Vec* corpus ReLi.

- ‘A história é leve, me identifiquei em alguns momentos com o mau humor do personagem, e achei muito interessante.’: o algoritmo identificou corretamente o aspecto ‘história’, mas apontou erroneamente o termo ‘personagem’;
- ‘Há episódios impagáveis do início ao fim.’: o algoritmo não identificou nenhum aspecto;

Os testes realizados com o ReLi estendido tiveram resultados semelhantes, como podemos observar na Tabela 16. Foram superiores aos resultados do experimento sem a poda *Word2Vec* e um pouco inferiores ao com poda realizado somente sobre o corpus ReLi. O melhor resultado foi uma medida f de 57,55% obtida com uma frequência de

corte de 2%. O experimento também demonstrou uma menor variação nos resultados quando comparado ao **FREQ** Baseline.

Tabela 16 – Resultados Hu & Liu + *Word2Vec* corpus ReLi estendido.

Frequência de Corte	Precisão	Revocação	Medida F
1%	52,27	60,51	56,09
2%	56,87	58,24	57,55
3%	57,90	51,59	54,56
4%	59,07	49,76	54,02
5%	59,07	49,76	54,02
6%	59,07	49,76	54,02
7%	59,07	49,76	54,02
8%	57,28	44,90	50,34
9%	57,28	44,90	50,34
10%	57,28	44,90	50,34

Valores de precisão, revocação e medida f dados em Porcentagem

Hu & Liu + *Word2Vec* nos Termos Infrequentes

Conforme explicamos anteriormente, encontramos alguns termos não relacionados na lista de aspectos obtida na etapa de identificação de termos infrequentes do método de Hu & Liu. Nesse experimento, mantivemos a poda utilizando detecção de contextos na primeira etapa, e também a aplicamos na detecção de aspectos infrequentes.

Os resultados não foram muito melhores do que o experimento anterior, o que já era esperado, dado que estamos lidando com termos infrequentes. As diferenças entre as medidas f do experimento anterior e o atual aumentaram proporcionalmente à frequência de corte, em 1% a diferença foi de 0,82% chegando a 2,45% no corte realizado com 10%. Essa alteração se explica devido ao fato que quanto maior a frequência de corte mais termos são processados como infrequentes, etapa onde essa alteração implementa sua melhoria.

Neste caso, a melhor medida f encontrada foi 61,07% e foi obtida com uma frequência de corte de 3%. Essa também foi a frequência de corte do melhor resultado do experimento anterior, que alcançou uma medida f de 59,80%. Os resultados desse experimento estão apresentados na Tabela 17 e na Figura 6.

Abaixo elencamos alguns exemplos de aspectos detectados corretamente:

- ‘Tem muitas crônicas agradáveis, muita coisa bacana sobre a amizade, a mãe que faz queimar a nossa cara, amigos com problemas e qualidades.’: o algoritmo detectou corretamente o termo ‘crônicas’, outros algoritmos detectaram erroneamente o termo ‘coisa’, que não está relacionado ao contexto literário;

Tabela 17 – Resultados Hu & Liu + *Word2Vec* nos Infrequentes corpus ReLi.

Frequência de Corte	Precisão	Revocação	Medida F
1%	52,53	64,42	57,87
2%	58,79	62,23	60,46
3%	61,23	60,91	61,07
4%	60,67	55,72	58,09
5%	63,31	55,39	59,09
6%	63,67	54,41	58,68
7%	63,34	52,69	57,52
8%	63,34	52,69	57,52
9%	63,60	50,79	56,47
10%	63,60	50,79	56,47

Valores de precisão, revocação e medida f dados em Porcentagem

Figura 6 – Resultados Hu & Liu + Infrequentes *Word2Vec* corpus ReLi.

- ‘No começo é um bocado confuso, e te deixa até um pouco deprimido, mas na real, é um troço que realmente vale a pena ler.’: outros algoritmos também detectaram incorretamente termos não relacionados ao contexto;

E alguns exemplos onde o algoritmo falhou na detecção:

- ‘Além disso, o estilo é extremamente autêntico.’: o algoritmo não detectou nenhum aspecto;
- ‘O enredo não é nada demais, só a rotina de um adolescente depois de ser expulso do internato.’: detectou incorretamente ‘internato’ como aspecto;

No teste realizado com o ReLi estendido, mais uma vez os resultados foram um pouco inferiores quando comparados aos testes feitos somente com o corpus ReLi. Como diferença, nesse experimento, o resultado da frequência de corte em 1% (57,22%) foi muito próximo ao resultado obtido somente com o ReLi (57,87%). O melhor resultado foi alcançado no corte realizado com 2% de frequência, com uma medida f de 58,95%, como apresentado na Tabela 18.

Tabela 18 – Resultados Hu & Liu + *Word2Vec* nos Infrequentes corpus ReLi.

Frequência de Corte	Precisão	Revocação	Medida F
1%	53,84	61,06	57,22
2%	58,99	58,90	58,95
3%	61,72	52,65	56,83
4%	63,60	50,79	56,47
5%	63,60	50,79	56,47
6%	63,60	50,79	56,47
7%	63,60	50,79	56,47
8%	63,25	46,69	53,72
9%	63,25	46,69	53,72
10%	63,25	46,69	53,72

Valores de precisão, revocação e medida f dados em Porcentagem

4.6 Conclusão

Nossos experimentos apresentaram resultados variados, principalmente devido a variação nas frequências de corte. Contudo eles foram condizentes com o que se esperava, nos seguintes termos: (i) Hu & Liu obteve melhores resultados que o **FREQ**, (ii) o *Word2Vec*, em geral, melhorou a detecção de aspectos; e (iii) foram poucas as diferenças nos resultados obtidos com o corpus estendido.

O **FREQ** Baseline, apesar de simples, alcançou bons resultados na detecção de aspectos. Tem como limitação uma grande dependência no parâmetro da frequência de corte. No nosso estudo, como trabalhamos com corpus anotado, podemos encontrar a melhor frequência de corte. Em um cenário de uso real do algoritmo não é possível fazer essa validação da frequência utilizada. Em nosso experimento com o corpus ReLi encontramos como pior resultado uma medida f de 26,71% e como melhor 55,81%, uma diferença muito grande com a alteração de apenas um parâmetro.

Nossa implementação do método de Hu & Liu alcançou resultados mais equilibrados entre precisão e revocação e também menos dependentes da frequência de corte. A diferença entre os melhores e piores resultados foi bem menor. Essa maior independência com relação à frequência de corte é importante para um método não supervisionado,

dado que como explicamos, em caso de um cenário real de uso do algoritmo é mais difícil verificar se a frequência de corte escolhida foi realmente a melhor.

Com o *Word2Vec* foi possível fazer a detecção do contexto dos aspectos candidatos, verificando se pertenciam a um contexto mais cotidiano ou ao contexto literário. A eliminação de candidatos não relacionados melhorou os resultados obtidos. No caso específico do **FREQ** essa melhora só apareceu com frequências de corte menores, dado que a poda pela frequência acabou eliminando termos não relacionados.

No Hu & Liu com o *Word2Vec*, o benefício foi mais evidente. Houve uma melhora nos resultados e uma diminuição na diferença entre o pior e melhor resultado obtido, indicando uma diminuição ainda maior da dependência do método com relação ao parâmetro da frequência de corte.

Com a mudança implementada na detecção dos infrequentes e uso do *Word2Vec* também nessa etapa, houve uma pequena melhora nos resultados. De qualquer forma, os resultados foram interessantes dado que a diferença entre pior e melhor resultado foi ainda menor. A Figura 7 apresenta os melhores resultados obtidos com cada método.

Figura 7 – Resultados corpus ReLi.

Nos testes com o corpus estendido, como mostrado na Figura 8, os algoritmos alcançaram desempenhos próximos. Os resultados foram um pouco inferiores aos dos experimentos realizados somente com o ReLi. Dada a considerável diferença no tamanho dos corpora, pode-se concluir que os métodos são válidos tanto para conjuntos menores ou maiores que os usados aqui.

Figura 8 – Resultados corpus ReLi Estendido.

Identificação de Sentimentos

5.1 Introdução

Como explicamos, a análise de sentimentos utiliza técnicas de processamento de linguagem natural com objetivo de extrair e classificar opiniões, emoções, avaliações e atitudes relacionadas a produtos, organizações, pessoas, eventos, entre outros, expressos em textos livres. A identificação de sentimentos tratada neste capítulo refere-se a detecção da polaridade, ou seja, a detecção de sentimentos positivos, negativos ou neutros, relacionados aos aspectos identificados em um texto.

Existem duas abordagens principais para resolução do problema de identificação de sentimentos (TABOADA et al., 2011). Uma abordagem baseada em léxicos que envolve o cálculo da polaridade do documento a partir da orientação semântica de suas palavras ou frases, sendo essas orientações obtidas a partir de algum dicionário. Outra forma é a classificação de textos criando-se classificadores a partir de textos já classificados, ou seja, utilizando método supervisionado de aprendizado de máquina.

Em nosso trabalho optamos pela primeira abordagem por se tratar de método não supervisionado, assim como os métodos utilizados na identificação de aspectos. Apesar de encontrarmos grande quantidade de recursos relacionados a identificação de sentimentos para língua inglesa, o mesmo não ocorre para a língua portuguesa. Devido a esse fato, neste trabalho desenvolvemos um método para criação de um léxico sensível ao contexto, ou seja, especializado na detecção de sentimentos para um tipo específico de textos. Também exploramos métodos para a combinação de léxicos de sentimentos, com objetivo de obtermos melhores resultados na análise.

Os léxicos criados foram avaliados utilizando métodos para detecção de sentimentos, aplicados nos textos do já mencionado corpus ReLi.

5.2 Dados Analisados

O corpus ReLi é formado por textos opinativos sobre livros, e possui marcações dos aspectos e respectivas polaridades, encontrados em cada frase. Para dar sequência ao trabalho, selecionamos apenas os textos que continham algum aspecto no intuito de identificar a polaridade relacionada a esse aspecto. Similar a metodologia adotada no congresso *SemEval*, levamos em consideração apenas os aspectos marcados no corpus, e não os aspectos encontrados no capítulo anterior.

A Tabela 19 apresenta o total dos dados que foram analisados nessa etapa, 2.770 aspectos com suas respectivas polaridades. Observa-se um grande desbalanceamento na distribuição das polaridades, ou seja, uma grande quantidade de aspectos classificados como positivos com relação aos negativos ou desconhecidos.

Tabela 19 – Distribuição de polaridades relacionadas a aspectos no ReLi.

Aspectos	Quantidade	Porcentagem
Positivos	2.089	75,42
Negativos	586	21,16
Desconhecidos	95	5,37
Total	2.770	100,00

O corpus ReLi estendido, utilizado na identificação de aspectos, foi utilizado também na criação de um léxico direcionado para a análise de textos opinativos relacionados a livros, como explicaremos na próxima seção.

5.3 Métodos para Criação de Léxicos de Sentimentos

A principal ferramenta utilizada na abordagem escolhida para identificação de sentimentos, foram os léxicos de sentimentos, dicionários de palavras opinativas com suas respectivas polaridades (positiva, negativa ou neutra). Selecionamos três léxicos frequentemente utilizados em outros trabalhos: OpLexicon, SentiLex e o LIWC (*Linguistic Inquiry and Word Count*).

O léxico OpLexicon (SOUZA et al., 2011; SOUZA; VIEIRA, 2012) foi criado através da combinação de diferentes técnicas. Um primeiro léxico intermediário foi criado utilizando a análise de corpus anotado. Outro foi criado utilizando busca de sinônimos e antônimos a partir de uma lista inicial de palavras. Um último léxico foi obtido através da tradução automática do *Liu's English Opinion Lexicon* (HU; LIU, 2004a). A combinação desses três léxicos deu origem ao OpLexicon, com um total de 32.191 entradas entre *emoticons* (66), termos com *hashtag* (471), adjetivos (24.475), verbos (6.889) e expressões (290) classificados em negativos (14.569), neutros (9002) e positivos (8.620).

Outro léxico utilizado, o Sentilex (SILVA et al., 2010), foi apresentado em trabalho que demonstrou um método para ampliação de léxicos de sentimentos. O desenvolvimento do léxico se deu em duas etapas: primeiro foi criado um dicionário de adjetivos com suas respectivas polaridades. A partir desses adjetivos, foi manualmente criado um conjunto de regras léxico-sintáticas, que foram aplicadas em uma grande coleção de n-gramas. As frequências dos adjetivos e regras encontradas foram utilizadas como entradas para um classificador estatístico. Em uma terceira etapa, foram atribuídas polaridades aos novos adjetivos encontrados e essa lista foi expandida por meio da exploração de grafos de sinônimos. Em sua composição final, o Sentilex conta com 82.347 entradas entre expressões idiomáticas (34.700), verbos (29.504), adjetivos (16.863) e substantivos (1.280) classificados em negativos (53.973), positivos (20.670) e neutros (7.704). A polaridade foi atribuída na forma de uma graduação numérica com valores entre -10 e 10, sendo zero considerado neutro.

O *Brazilian Portuguese LIWC 2007* (FILHO; PARDO; ALUÍSIO, 2013) foi criado via tradução do dicionário original em inglês (PENNEBAKER; FRANCIS; BOOTH, 2001). Foram três equipes de tradução que utilizaram vários dicionários bilíngues português-inglês. As conjugações foram inseridas automaticamente usando o dicionário Unitex-PB¹ do Núcleo Interinstitucional de Linguística Computacional². As categorias foram obtidas do dicionário em inglês. O LIWC é composto por 127.160 entradas, mas ao contrário dos outros léxicos elas não estão morfologicamente categorizadas. A classificação dessas entradas também é diferente. O léxico não possui uma classe para os neutros e além dos positivos (12.878 entradas) e negativos (15.115), encontramos categorias como palavras de negação (19), saúde (7.003), dinheiro (5.353), corpo (4.766), amizade (679), família (96), entre outras, totalizando 64 categorias.

Neste trabalho, criamos léxicos a partir da combinação dos três dicionários mencionados, utilizando duas abordagens para isso. Foi necessária uma padronização dos dicionários, dado que cada um possui um formato diferente. Somente foram consideradas as polaridades positivas, negativas e neutras e entradas repetidas e expressões idiomáticas foram removidas. Em um primeiro experimento apenas combinamos os dicionários ignorando eventuais discordâncias entre eles prevalecendo nesses casos, a classificação adotada pelo primeiro dicionário adicionado ao conjunto. Como nesse caso a ordem de combinação dos dicionários altera o resultado final, criamos léxicos com as seis combinações possíveis na ordem de inclusão dos três dicionários. No total foram gerados 6 léxicos com quantidades iguais de entradas (111.707), mas com ligeiras diferenças em suas polaridades, sendo 34.433 a média das entradas positivas, 64.638 as negativas e 12.636 as neutras.

Na segunda abordagem utilizada, realizamos um processo de conciliação quando

¹ <http://www.nilc.icmc.usp.br/nilc/projects/unitex-pb/web/dicionarios.html>

² <http://www.nilc.icmc.usp.br/>

encontrada discordância na polaridade de uma palavra entre os léxicos. Foram adicionadas ao léxico somente palavras que tenham sido encontradas em apenas um dos três léxicos, dado que, nesses casos, não existe possibilidade de discordância. Das palavras que apareceram em mais de um dicionário, somente foram incluídas aquelas classificadas com a mesma polaridade em pelo menos dois dicionários. A composição final desse léxico contém 106.374 entradas sendo 62.676 positivas, 32.543 negativas e 11.155 neutras.

Também construímos um léxico próprio, daqui em diante denominado Léxico ReLi (LexReLi), direcionado para identificação de polaridade de aspectos em textos opinativos sobre livros. Para a construção desse léxico combinamos métodos de identificação de aspectos com identificação de sentimentos, trabalhando com o conjunto de dados do ReLi Estendido. O objetivo foi criar um léxico somente com adjetivos, identificando suas polaridades através do contexto a que pertencem (polaridade das frases).

Ao observar os primeiros resultados das análises de polaridades realizadas com o corpus ReLi, percebemos que a polaridade de algumas palavras não estava correta dentro do contexto analisado, textos opinativos sobre livros. Essas palavras com polaridade incorreta na maioria das vezes foram encontradas em frases que foram classificadas com a polaridade que seria a correta para a palavra. Dado isso, chegamos a hipótese de que é possível descobrir a polaridade de uma palavra a partir do contexto a qual ela está inserida, ou seja, a polaridade dessa palavra na maior parte das vezes coincide com a polaridade encontrada para a frase. Se temos um texto falando bem de um determinado livro, os aspectos referenciados provavelmente serão em sua maioria positivos.

Dados esses conceitos, elaboramos método combinando de detecção de aspectos com detecção de polaridade das frases com o objetivo de criar nosso léxico sensível ao contexto. Os seguintes passos foram seguidos para a criação do LexReLi:

- Identificação de frases que possuem termos candidatos a aspectos (geralmente encontrados na forma de substantivos) próximos a palavras opinativas (geralmente adjetivos);
- Essas frases passaram então por um processo de identificação de polaridade. Para isso utilizamos o método Preferência aos Adjetivos com o léxico obtido da combinação do LIWC, Sentilex e OpLexicon, pois essa combinação obteve bons resultados nos experimentos que realizamos. O método, experimentos e resultados serão apresentados na próxima seção;
- Fizemos então o cálculo da frequência de cada adjetivo em frases positivas e negativas;
- A polaridade atribuída a cada adjetivo foi aquela em que o adjetivo foi mais frequente.

Portanto, como podemos observar, o LexReLi foi formado somente por adjetivos relacionados ao universo literário. A composição final do LexReLi contém 1.543 entradas, sendo 1.091 positivas e 452 negativos. Ao combinarmos o LexReLi com os outros léxicos, formamos mais 6 léxicos com 111.903 entradas cada um possuindo em média 64.685 entradas negativas, 37.301 positivas e 38.464 neutras.

A Tabela 20 apresenta a composição final (pós normalização) dos léxicos de sentimentos utilizados neste trabalho.

Tabela 20 – Distribuição de polaridades encontradas nos léxicos.

Léxico	Positivo	Neutro	Negativo	Total
OpLexicon	8.595	8.974	14.550	32.119
SentiLex	20.478	7.600	51.112	79.190
LIWC	12.376	0	14.612	26.988
LexReLi	1.091	0	452	1.543
Conciliado	32.543	11.155	62.676	106.374
LIWC+OpLexicon+SentiLex	34.241	11.697	65.769	111.707
LIWC+SentiLex+OpLexicon	34.267	11.365	66.075	111.707
OpLexicon+LIWC+SentiLex	33.795	12.208	65.704	111.707
OpLexicon+SentiLex+LIWC	34.655	13.824	63.228	111.707
SentiLex+LIWC+OpLexicon	34.926	13.249	63.532	111.707
SentiLex+OpLexicon+LIWC	34.716	13.474	63.517	111.707
LexReLi+LIWC+OpLexicon+SentiLex	34.762	11.318	65.823	111.903
LexReLi+LIWC+SentiLex+OpLexicon	34.786	11.019	66.098	111.903
LexReLi+OpLexicon+LIWC+SentiLex	34.353	11.772	65.778	111.903
LexReLi+OpLexicon+SentiLex+LIWC	35.212	13.388	63.303	111.903
LexReLi+SentiLex+LIWC+OpLexicon	35.463	12.884	63.556	111.903
LexReLi+SentiLex+OpLexicon+LIWC	35.270	13.082	63.551	111.903

5.4 Métodos para Identificação de Sentimentos

Implementamos quatro métodos para identificação de sentimentos, todos utilizando os léxicos de sentimentos mencionados:

- Polaridade das palavras: polaridade calculada com base em todas as palavras da frase, independentemente de classe gramatical;
- Somente Adjetivos: polaridade calculada somente com base em adjetivos;
- Preferência aos Adjetivos: calcula a polaridade utilizando os adjetivos, caso não isso seja possível, utiliza todas as palavras da frase;
- Baseada em Aspecto: busca identificar a polaridade relacionada especificamente a cada aspecto;

- Polaridade de termo utilizando *Word2Vec*: implementado em combinação com os métodos anteriores, utilizada quando determinado termo não é encontrado no léxico;

Os métodos foram implementados de forma incremental, cada um com objetivo de melhorar o resultado do método anterior.

Polaridade das Palavras

Este é um método simples de identificação de sentimentos utilizando léxicos. Cada palavra da frase é buscada no léxico de sentimentos independentemente de sua classe gramatical. O valor da polaridade encontrada para cada palavra é então somado. Se o valor final obtido nessa soma for maior que zero, indica uma polaridade positiva, igual a zero polaridade neutra e inferior a zero negativa. Assim como no método SO-Cal (TABOADA et al., 2011) a negação foi levada em consideração. Uma lista de termos negativos foi obtida a partir do léxico LIWC. Quando o algoritmo encontra uma palavra dessa lista, a polaridade da próxima palavra encontrada é invertida.

A polaridade obtida pelo método está relacionada a frase como um todo, e não a cada aspecto. Atribuímos então para cada aspecto encontrado na frase essa mesma polaridade, não sendo possível, portanto identificar polaridades distintas em uma mesma frase. Por sua simplicidade, esse método foi escolhido como base de comparação para os outros experimentos de identificação de sentimentos.

Somente Adjetivos

Similar ao método anterior, mas nesse caso somente foram somadas as polaridades dos adjetivos. A negação também foi considerada, invertendo a polaridade do adjetivo mais próximo ao termo negativo. A polaridade encontrada para a frase foi atribuída aos aspectos e frases onde não foram encontrados adjetivos foram marcadas como neutras.

Preferência aos Adjetivos

Trata-se de uma combinação dos métodos anteriores, inicialmente foram consideradas somente as polaridades dos adjetivos. Quando não encontrados adjetivos com polaridade, o algoritmo analisa as polaridades de todas as palavras, assim como no primeiro método. A negação foi tratada da mesma maneira que nos métodos anteriores, invertendo a polaridade da palavra ou adjetivo mais próximo ao termo negativo. Esse também é considerado um método de identificação de sentimento em nível de frase.

Baseada em Aspecto

A prioridade nesse método é localizar a polaridade relacionada a cada aspecto encontrado nas frases. Dado um aspecto o algoritmo procura o adjetivo mais próximo a ele e busca sua polaridade no léxico. Caso exista uma palavra de negação entre aspecto e adjetivo essa polaridade é invertida. Caso não seja encontrado um adjetivo próximo ou o adjetivo não apareça no léxico, um dos métodos anteriores de identificação de sentimento em nível de frases é utilizada.

Polaridade de Termo Utilizando *Word2Vec*

Para contornar a ausência de alguns termos nos léxicos, implementamos uma função auxiliar aos métodos anteriores com objetivo de identificar a polaridade dessas palavras. Para isso utilizamos o já estudado modelo *Word2Vec* (veja Capítulo 4). Realizamos duas implementações dessa função.

Com base nos léxicos, localizamos os quarenta adjetivos positivos e negativos mais frequentes no ReLi. Essa quantidade de adjetivos foi encontrada após a realização de alguns testes com outras quantidades e verificação da quantidade de adjetivos que tiveram sua polaridade identificada corretamente. Utilizando o modelo de linguagem treinado pela equipe de Hartmann (HARTMANN et al., 2017) cada palavra desses dois conjuntos foi convertida em seu respectivo vetor.

Nesse ponto as duas implementações se diferenciam. Na primeira implementação, calculamos o centroide dos quarenta vetores positivos e também centroide dos vetores negativos. Como resultado obtivemos dois vetores: um de centroide positiva e outro de centroide negativa.

Para identificar a polaridade de um termo, inicialmente recupera-se seu vetor no mesmo modelo de linguagem utilizado. O vetor encontrado é então comparado com ambos vetores de centroide, positivo e negativo. Essa comparação é realizada utilizando a medida de semelhança por cosseno. O vetor de centroide que apresentar maior semelhança para com o vetor do termo, indicará sua polaridade.

Em nossa segunda implementação, não realizamos o cálculo de centroides. Utilizando também a medida de semelhança por cosseno, comparamos o vetor do termo que se deseja identificar a polaridade, com os vetores das palavras contidas nas listas de positivos e negativos. Dessa forma encontramos o vetor de uma palavra positiva e o vetor de uma palavra negativa mais semelhantes ao vetor de interesse. Se a semelhança maior for com o vetor do termo positivo, essa polaridade é atribuída ao termo, caso contrário a polaridade negativa é retornada.

5.5 Métricas de Avaliação

A avaliação dos métodos foi realizada através da comparação dos sentimentos identificados para cada aspecto, com os sentimentos de referência anotados no corpus ReLi. Assim como na identificação de aspectos, utilizamos metodologia semelhante à utilizada nos workshops do SemEval (PONTIKI et al., 2015). Realizamos os cálculos da precisão, revocação e medida f individualmente para cada classe (positivos e negativos) e a acurácia como uma medida geral para cada experimento.

Para efetuarmos os cálculos foi necessário obtermos para cada classe: a quantidade de sentimentos identificados corretamente com a classe analisada (verdadeiros positivos, vp), os classificados erroneamente com a classe oposta (falsos negativos, fn), os classificados erroneamente com a classe analisada (falsos positivos, fp) e os corretamente classificados como a classe oposta (verdadeiros negativos, vn).

Dessa forma, as seguintes fórmulas calculam a precisão (P_c), revocação (R_c) e medida f (F_c) para cada classe e a acurácia (A):

$$P_c = \frac{vp}{vp + fp} \quad R_c = \frac{vp}{vp + fn} \quad F_c = 2 \cdot \frac{P_c \cdot R_c}{P_c + R_c} \quad A = \frac{\sum_{c=1}^C (vp_i + vn_i)}{\sum_{c=1}^C (vp_i + vn_i + fp_i + fn_i)}$$

Onde C é a quantidade de c classes a serem avaliadas e i representa cada elemento individualmente nessas classes.

5.6 Resultados dos Experimentos

Polaridade das Palavras

Conforme explicado anteriormente, este método busca encontrar a polaridade da frase, através das polaridades das palavras que a compõem independentemente de suas classes gramaticais. Testamos o método com os léxicos que foram apresentados. Observamos principalmente a métrica da acurácia pelo fato desta ser relacionada com o experimento como um todo e não individualmente com suas classes (positivos e negativos).

A Tabela 21 apresenta os resultados dos primeiros experimentos realizados. Na primeira parte da tabela mostramos os resultados dos léxicos individualmente. Nota-se que o LexReLi obteve um resultado positivo expressivo, 69,20% de acurácia, apesar de possuir uma quantidade menor de palavras com relação aos outros dicionários, trata-se de um léxico especializado em textos opinativos sobre livros.

A segunda parte da tabela apresenta os resultados para as combinações de léxicos. Podemos ver que o léxico Conciliado, que foi obtido pela combinação e conciliação da discordância entre os léxicos, teve o menor resultado entre os léxicos combinados com 60,82% de acurácia. A intenção de se repetir o experimento com diferentes combinações de léxico, foi verificar se a ordem de inclusão dos léxicos interferiria no resultado final. Pode-se observar que os melhores resultados foram obtidos nas combinações que foram iniciadas com a inclusão do léxico LIWC, provavelmente devido ao fato de que se excluindo o LexReLi, esse foi o léxico com melhor resultado nos testes individuais.

Em sua última seção, a tabela apresenta os resultados da combinação do LexReLi com os outros léxicos. Dado que nos experimentos realizados com os léxicos combinados, as combinações iniciadas com o LIWC obtiveram os melhores resultados e que este obteve o melhor resultado individual entre os três léxicos (LIWC, OpLexicon e SentiLex), optamos na combinação dos quatro léxicos iniciarmos com a inclusão do LexReLi, já que esse alcançou os melhores resultados individuais entre os quatro léxicos. Os melhores resultados desse experimento, 74,69% e 74,62% de acurácia, foram encontrados nas combinações de léxicos iniciadas pelo LIWC.

Tabela 21 – Resultados Polaridade das Palavras.

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	84,51	60,84	70,75	44,89	38,23	41,29	55,89
SentiLex	84,94	51,03	63,76	46,03	45,56	45,80	49,83
LIWC	79,35	80,56	79,95	41,29	14,16	21,09	66,02
LexReLi	79,68	84,49	82,02	62,32	14,68	23,76	69,20
Conciliado	81,94	68,41	74,56	43,42	33,79	38,00	60,82
LIWC+Op+Sent	80,18	83,48	81,80	51,31	23,38	32,12	70,32
LIWC+Sent+Op	80,90	83,53	82,20	53,48	24,91	33,99	70,69
Op+LIWC+Sent	80,28	81,09	80,69	47,97	24,23	32,20	68,64
Op+Sent+LIWC	81,43	71,80	76,32	42,96	30,72	35,82	62,80
Sent+LIWC+Op	81,70	70,75	75,83	46,67	33,45	38,97	62,58
Sent+Op+LIWC	82,54	69,03	75,18	44,94	36,35	40,19	61,87
LexReLi+LIWC+Op+Sent	78,58	91,86	84,71	59,85	13,48	22,01	74,69
LexReLi+LIWC+Sent+Op	78,76	91,77	84,77	59,40	13,48	21,97	74,62
LexReLi+Op+LIWC+Sent	78,57	91,62	84,60	57,45	13,82	22,28	74,58
LexReLi+Op+Sent+LIWC	79,23	86,55	82,73	52,79	17,75	26,56	71,48
LexReLi+Sent+LIWC+Op	79,67	86,64	83,01	54,45	17,75	26,77	71,55
LexReLi+Sent+Op+LIWC	79,64	86,31	82,84	53,03	17,92	26,79	71,33

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem.

Somente Adjetivos

Nesse experimento para a identificação da polaridade das frases somente foram analisados os adjetivos. Devido a essa limitação, realizamos testes utilizando o modelo distribucional

Word2Vec em tentativa de identificar a polaridade de uma quantidade maior de adjetivos.

Os resultados, como podemos observar na Tabela 22, foram inferiores aos do experimento anterior. Para os léxicos individuais, novamente o LexReLi obteve o melhor resultado com 59,14% de acurácia. Para os léxicos combinados sem o LexReLi o melhor resultado foi uma acurácia de 52,60%. Quando combinados com o LexReLi obtivemos 59,78% de acurácia.

Tabela 22 – Resultados Somente Adjetivos

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	85,09	53,28	65,53	57,76	34,30	43,04	49,12
SentiLex	87,00	46,15	60,31	59,29	39,76	47,60	44,75
LIWC	82,45	34,85	48,99	57,61	18,09	27,53	31,18
LexReLi	83,80	67,59	74,83	65,89	29,01	40,28	59,14
Conciliado	86,60	49,21	62,76	59,89	38,23	46,67	46,80
LIWC+Op+Sent	85,36	57,49	68,71	59,54	35,15	44,21	52,60
LIWC+Sent+Op	86,20	56,20	68,04	61,14	38,40	47,17	52,30
Op+LIWC+Sent	85,13	53,71	65,86	58,26	34,30	43,18	49,46
Op+Sent+LIWC	85,00	53,71	65,83	58,26	34,30	43,18	49,46
Sent+LIWC+Op	85,91	52,23	64,96	60,73	39,59	47,93	49,46
Sent+Op+LIWC	86,66	50,36	63,70	60,87	40,61	48,72	48,22
LexReLi+LIWC+Op+Sent	84,32	67,69	75,09	64,46	31,57	42,38	59,78
LexReLi+LIWC+Sent+Op	84,44	67,30	74,91	64,48	31,91	42,69	59,55
LexReLi+Op+LIWC+Sent	84,41	67,64	75,10	64,46	31,57	42,38	59,74
LexReLi+Op+Sent+LIWC	84,31	67,64	75,06	64,46	31,57	42,38	59,74
LexReLi+Sent+LIWC+Op	84,44	67,26	74,87	64,38	32,08	42,82	59,55
LexReLi+Sent+Op+LIWC	84,44	67,26	74,87	64,38	32,08	42,82	59,55

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Em nosso primeiro experimento utilizando o modelo *Word2Vec*, como já fora explicado, trabalhamos de forma similar ao método utilizado na identificação de aspectos. Calculamos centroides dos adjetivos positivos e negativos mais frequentes e as comparamos com os vetores dos adjetivos que não foram encontrados nos léxicos. Os resultados demonstrados na Tabela 23 foram quase que em suas totalidades inferiores ao experimento anterior, o que nos levou ao desenvolvimento de uma segunda versão de detecção de polaridade utilizando *Word2Vec*.

No segundo experimento, não fizemos o cálculo dos centroides. Utilizando a medida de semelhança por cosseno, comparamos o vetor do adjetivo que não foi encontrado no léxico com os vetores dos adjetivos mais frequentes positivos e negativos. Verificando qual possui a maior similaridade e adotando para o adjetivo sua polaridade. Entretanto os resultados dessa abordagem também não foram satisfatórios. Como podemos ver na Tabela 24 foram ligeiramente inferiores aos resultados obtidos no primeiro teste realizado com o modelo *Word2Vec*.

Tabela 23 – Resultados Somente Adjetivos *Word2Vec* Versão 1

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,46	57,25	67,92	47,07	37,03	41,45	52,82
SentiLex	84,92	53,37	65,55	45,52	43,34	44,41	51,18
LIWC	83,85	46,96	60,20	38,47	44,71	41,36	46,47
LexReLi	84,30	65,01	73,41	56,51	34,81	43,08	58,39
Conciliado	84,86	55,00	66,74	48,83	42,66	45,54	52,30
LIWC+Op+Sent	84,17	59,55	69,75	50,90	38,74	43,99	54,99
LIWC+Sent+Op	85,07	57,83	68,85	50,61	42,32	46,10	54,43
Op+LIWC+Sent	83,38	57,64	68,16	48,00	36,86	41,70	53,08
Op+Sent+LIWC	83,38	57,64	68,16	48,00	36,86	41,70	53,08
Sent+LIWC+Op	85,01	55,10	66,86	48,94	43,17	45,87	52,49
Sent+Op+LIWC	84,85	55,24	66,92	48,73	42,66	45,50	52,49
LexReLi+LIWC+Op+Sent	84,33	66,73	74,51	59,70	34,13	43,43	59,59
LexReLi+LIWC+Sent+Op	84,42	66,40	74,33	59,24	34,47	43,58	59,40
LexReLi+Op+LIWC+Sent	84,36	66,63	74,46	59,47	34,30	43,51	59,55
LexReLi+Op+Sent+LIWC	84,36	66,63	74,46	59,47	34,30	43,51	59,55
LexReLi+Sent+LIWC+Op	84,41	66,11	74,15	58,84	34,64	43,61	59,21
LexReLi+Sent+Op+LIWC	84,41	66,11	74,15	58,84	34,64	43,61	59,21

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Tabela 24 – Resultados Somente Adjetivos *Word2Vec* Versão 2.

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,56	59,36	69,41	47,25	36,69	41,31	54,39
SentiLex	85,03	56,82	68,12	48,91	41,98	45,18	53,57
LIWC	83,56	49,88	62,47	38,44	40,27	39,33	47,78
LexReLi	83,67	65,25	73,32	55,49	32,76	41,20	58,13
Conciliado	84,89	57,83	68,79	50,00	41,30	45,23	54,21
LIWC+Op+Sent	84,26	59,69	69,88	49,67	38,40	43,31	55,03
LIWC+Sent+Op	85,07	58,64	69,42	50,94	41,81	45,92	54,95
Op+LIWC+Sent	83,50	59,84	69,72	48,75	36,52	41,76	54,73
Op+Sent+LIWC	83,50	59,84	69,72	48,75	36,52	41,76	54,73
Sent+LIWC+Op	85,07	58,11	69,06	50,61	42,15	46,00	54,62
Sent+Op+LIWC	85,07	58,11	69,06	50,72	42,15	46,04	54,62
LexReLi+LIWC+Op+Sent	83,91	66,40	74,13	58,82	32,42	41,80	58,95
LexReLi+LIWC+Sent+Op	84,07	66,20	74,08	59,34	33,62	42,92	59,07
LexReLi+Op+LIWC+Sent	83,99	66,30	74,10	58,72	32,76	42,06	58,95
LexReLi+Op+Sent+LIWC	83,99	66,30	74,10	58,72	32,76	42,06	58,95
LexReLi+Sent+LIWC+Op	84,12	65,92	73,91	59,05	33,96	43,12	58,92
LexReLi+Sent+Op+LIWC	84,12	65,92	73,91	59,05	33,96	43,12	58,92

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem.

Preferência aos Adjetivos

Nesse experimento analisamos inicialmente os adjetivos, caso não sejam encontrados no léxico, o algoritmo passa a realizar a análise da frase como um todo, conforme já explicamos

Tabela 25 – Resultados Preferência aos Adjetivos.

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	84,07	65,44	73,59	47,31	45,05	46,15	60,97
SentiLex	85,11	56,92	68,22	48,86	51,02	49,92	55,63
LIWC	80,33	80,33	80,33	46,60	24,57	32,18	68,11
LexReLi	81,13	83,96	82,52	64,93	29,69	40,75	72,07
Conciliado	83,46	71,76	77,17	49,11	47,10	48,08	66,36
LIWC+Op+Sent	82,50	84,87	83,67	53,67	39,93	45,79	75,03
LIWC+Sent+Op	83,29	84,68	83,98	54,98	43,34	48,47	75,63
Op+LIWC+Sent	82,39	82,86	82,63	51,29	40,61	45,33	73,61
Op+Sent+LIWC	82,75	75,78	79,11	48,28	43,17	45,59	68,64
Sent+LIWC+Op	83,16	74,25	78,45	50,54	47,95	49,21	68,49
Sent+Op+LIWC	83,78	72,95	77,99	50,09	49,49	49,79	67,81
LexReLi+LIWC+Op+Sent	81,30	91,34	86,02	61,35	34,13	43,86	78,80
LexReLi+LIWC+Sent+Op	81,37	91,38	86,09	61,21	34,47	44,10	78,92
LexReLi+Op+LIWC+Sent	81,28	91,24	85,97	60,79	34,13	43,72	78,73
LexReLi+Op+Sent+LIWC	81,48	87,84	84,54	58,69	35,15	43,97	76,30
LexReLi+Sent+LIWC+Op	81,60	87,89	84,63	58,87	35,67	44,42	76,45
LexReLi+Sent+Op+LIWC	81,62	87,79	84,59	58,66	35,84	44,49	76,41

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem.

na Seção 5.4 Métodos para Identificação de Sentimentos.

Os resultados do experimento são apresentados na Tabela 25 e foram superiores aos experimentos anteriores. Com relação ao desempenho dos léxicos, os resultados foram similares, com o LexReLi alcançando o melhor resultado com uma acurácia de 72,07%. O conjunto de léxicos LIWC + SentiLex + OpLexicon obteve 75,63% de acurácia. Acrescentando o LexReLi a essa mesma combinação, alcançamos uma acurácia de 78,92%, melhor resultado entre os experimentos realizados.

Dados os resultados não satisfatórios obtidos com as abordagens que utilizaram o Modelo *Word2Vec*, optamos por incluir na seção de anexos as tabelas com os resultados dos experimentos com as 2 versões, com e sem uso de centroides.

Baseada em Aspectos

Conforme explicado na Seção 5.4 Métodos para Identificação de Sentimentos, até agora trabalhamos com as frases em sua totalidade, sem nos preocuparmos com a polaridade dos aspectos individualmente. Nesse experimento inicialmente localizamos o aspecto na frase e o adjetivo mais próximo a ele. Verificamos a polaridade desse adjetivo e a atribuímos ao aspecto. Caso não exista adjetivo próximo ou a polaridade do mesmo não seja encontrada no léxico, utilizamos um dos métodos anteriores para a identificação da polaridade do aspecto.

Tabela 26 – Resultados Aspectos + Polaridade das Palavras.

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	84,16	65,63	73,75	46,39	46,08	46,23	61,35
SentiLex	84,79	57,40	68,46	48,47	54,10	51,13	56,67
LIWC	80,79	79,70	80,24	45,45	27,30	34,12	68,22
LexReLi	82,11	81,71	81,91	56,85	38,23	45,71	72,19
Conciliado	83,83	70,99	76,88	47,29	49,15	48,20	66,21
LIWC+Op+Sent	83,20	82,96	83,08	50,98	44,37	47,45	74,50
LIWC+Sent+Op	83,83	82,62	83,22	51,58	47,27	49,33	74,88
Op+LIWC+Sent	82,64	81,33	81,98	48,71	41,98	45,10	72,71
Op+Sent+LIWC	83,04	74,53	78,56	45,99	44,03	44,99	67,85
Sent+LIWC+Op	83,74	72,71	77,84	48,61	50,85	49,71	67,93
Sent+Op+LIWC	84,02	71,47	77,24	48,10	51,88	49,92	67,18
LexReLi+LIWC+Op+Sent	82,24	87,98	85,01	53,93	42,15	47,32	77,94
LexReLi+LIWC+Sent+Op	82,39	87,79	85,01	53,62	43,00	47,73	77,98
LexReLi+Op+LIWC+Sent	82,21	87,84	84,93	53,46	42,15	47,14	77,83
LexReLi+Op+Sent+LIWC	82,28	84,92	83,58	52,23	41,98	46,55	75,51
LexReLi+Sent+LIWC+Op	82,44	84,73	83,57	52,17	43,00	47,15	75,59
LexReLi+Sent+Op+LIWC	82,46	84,63	83,53	52,06	43,17	47,20	75,55

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem.

Em nosso primeiro experimento combinamos a detecção de polaridade do aspecto com o método Polaridade das Palavras. Os resultados estão apresentados na Tabela 26. Os resultados foram superiores ao experimento que utiliza o método Polaridade das Palavras somente nas frases. Considerando os léxicos individuais, o LexReLi obteve uma acurácia de 72,19%. Nos resultados para léxicos combinados, o LIWC + SentiLex + OpLexicon obteve 74,88% e a combinação com o LexReLi 77,98%, ambos resultados para a acurácia.

No segundo experimento realizado, utilizamos a análise somente dos adjetivos combinada com a detecção de polaridade do aspecto. Novamente os resultados foram superiores ao método aplicado somente nas frases, como podemos observar na Tabela 27. Entretanto foram inferiores ao experimento anterior. O resultado do léxico LexReLi para a acurácia foi de 62,06%. Para os léxicos combinados o LIWC + OpLexicon + SentiLex alcançou 56,04% e quando combinado ao LexReLi 63,14% para a acurácia.

Em nosso terceiro e último experimento combinamos o método atual com o aquele que prioriza os adjetivos na detecção das polaridades, caso não seja possível o algoritmo passa para a análise da frase pelas palavras, como no primeiro método apresentado. Os resultados obtidos foram pouco superiores ao método que combina aspectos com polaridade das palavras e acabou sendo levemente inferior ao resultado do experimento Preferência aos Adjetivos aplicado somente nas frases, como podemos ver na Tabela 28. Individualmente o léxico LexReLi obteve a melhor acurácia, 72,22%. A combinação LIWC +

Tabela 27 – Resultados Aspectos + Somente Adjetivos.

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	84,30	55,53	66,96	53,74	39,25	45,36	51,96
SentiLex	86,45	48,54	62,17	55,70	45,05	49,81	47,78
LIWC	82,41	36,33	50,43	53,91	21,16	30,39	33,01
LexReLi	83,27	68,84	75,37	57,07	37,88	45,54	62,06
Conciliado	86,10	51,89	64,76	55,46	43,34	48,66	50,02
LIWC+Op+Sent	85,30	59,74	70,27	55,29	42,83	48,27	56,04
LIWC+Sent+Op	86,05	58,16	69,41	55,17	45,56	49,91	55,40
Op+LIWC+Sent	84,54	56,01	67,38	53,95	39,59	45,67	52,41
Op+Sent+LIWC	84,42	56,01	67,34	53,74	39,25	45,36	52,34
Sent+LIWC+Op	85,75	54,14	66,37	54,88	46,08	50,09	52,37
Sent+Op+LIWC	86,12	52,56	65,28	55,39	45,56	50,00	51,03
LexReLi+LIWC+Op+Sent	84,01	69,17	75,87	55,45	41,64	47,56	63,14
LexReLi+LIWC+Sent+Op	84,07	68,74	75,64	55,13	42,15	47,78	62,92
LexReLi+Op+LIWC+Sent	84,05	69,12	75,86	55,33	41,64	47,52	63,10
LexReLi+Op+Sent+LIWC	83,95	69,12	75,82	55,13	41,30	47,22	63,03
LexReLi+Sent+LIWC+Op	84,01	68,65	75,55	54,79	41,98	47,54	62,80
LexReLi+Sent+Op+LIWC	84,01	68,65	75,55	54,79	41,98	47,54	62,80

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem.

Tabela 28 – Resultados Aspectos + Preferência aos Adjetivos

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,93	66,25	74,05	46,64	47,44	47,04	62,13
SentiLex	84,88	58,07	68,96	48,63	54,44	51,37	57,27
LIWC	80,75	79,70	80,22	45,45	27,30	34,12	68,22
LexReLi	82,12	81,76	81,94	56,85	38,23	45,71	72,22
Conciliado	83,90	71,33	77,10	47,11	50,00	48,51	66,65
LIWC+Op+Sent	83,45	83,29	83,37	51,14	46,08	48,47	75,14
LIWC+Sent+Op	84,05	82,72	83,38	51,52	49,32	50,39	75,40
Op+LIWC+Sent	82,91	81,52	82,21	48,68	44,20	46,33	73,35
Op+Sent+LIWC	83,01	74,87	78,73	45,69	45,22	45,45	68,37
Sent+LIWC+Op	83,85	73,05	78,08	48,10	51,88	49,92	68,41
Sent+Op+LIWC	84,11	71,95	77,55	47,68	52,56	50,00	67,70
LexReLi+LIWC+Op+Sent	82,41	87,89	85,06	54,18	43,17	48,05	78,09
LexReLi+LIWC+Sent+Op	82,49	87,70	85,01	53,78	43,69	48,21	78,06
LexReLi+Op+LIWC+Sent	82,38	87,75	84,98	53,72	43,17	47,87	77,98
LexReLi+Op+Sent+LIWC	82,46	84,83	83,62	52,50	43,00	47,28	75,66
LexReLi+Sent+LIWC+Op	82,53	84,59	83,55	52,24	43,69	47,58	75,63
LexReLi+Sent+Op+LIWC	82,55	84,49	83,51	52,13	43,86	47,64	75,59

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

SentiLex + OpLexicon alcançou como resultado 75,40% e o LexReLi combinado com LIWC + OpLexicon + SentiLex 78,09%.

Figura 9 – Resultados dos experimentos realizados com léxicos individuais.

Da mesma forma que nos experimentos anteriores, o modelo *Word2Vec* não obteve resultados satisfatórios. De qualquer forma os resultados das análises orientadas a aspectos combinadas com o *Word2Vec* se encontram na seção de anexos.

5.7 Conclusão

Podemos concluir que a abordagem utilizada para criar o LexReLi foi eficaz, dada a evidente melhoria nos resultados na aplicação dos métodos de identificação de sentimentos usando esse léxico. O gráfico da Figura 9 compara os resultados obtidos pelos léxicos individualmente na análise realizada em nível de aspecto. Outro ponto importante, é que o método aplicado na criação do LexReLi pode ser facilmente utilizado para a criação de léxicos para diferentes contextos. Eventuais melhorias nos métodos de análise de polaridade utilizados em sua criação devem implicar em melhorias no léxico obtido.

O método de combinação de léxicos que utilizou um processo de conciliação para resolver as divergências entre os léxicos mostrou-se menos eficaz, dado que, além de ser mais trabalhoso na criação, obteve resultados inferiores em relação aos métodos que utilizaram uma forma mais simples na combinação dos léxicos. O gráfico da Figura 10 apresenta os resultados para as combinações de léxicos, nele podemos ver os resultados do léxico obtido por conciliação, os piores e os melhores resultados entre as combinações de léxicos, e os melhores resultados obtidos nas combinações realizadas com o LexReLi.

No método mais simples para combinação de léxicos, obtivemos indicações de que

Figura 10 – Resultados dos experimentos realizados com léxicos combinados.

é boa prática dar prioridade ao léxico que obteve o melhor resultado individual, já que nos experimentos observamos que as combinações iniciadas pelo léxico LIWC alcançaram melhores resultados.

Embora ligeiramente superior, os resultados do método de identificação de sentimentos Preferência aos Adjetivos são promissores. Melhorias podem ser implementadas nas duas etapas da abordagem. Na primeira etapa, pode-se trabalhar com métodos mais avançados na detecção das palavras que qualificam os aspectos. Na segunda etapa, é possível melhorar ou aplicar outros métodos para detectar a polaridade no nível da frase.

Conclusões

Apesar de hoje ser uma área muito ativa, a análise de sentimentos baseada em aspectos ainda carece de trabalhos na língua portuguesa. Possuímos apenas um corpus anotado, com textos opinativos e marcações referentes aos aspectos e respectivos sentimentos, o que é muito pouco para uma efetiva avaliação e comparação de métodos de identificação de aspectos e sentimentos. Portanto, o objetivo desse trabalho foi o de estudar e avaliar vários métodos de análise de sentimentos baseada em aspectos em textos em português, mais especificamente, trabalhando com o recém-criado corpus ReLi (FREITAS et al., 2012).

Nos experimentos relacionados a identificação de aspectos, observamos que apesar de simples o método frequencial, **FREQ** Baseline, conseguiu alcançar bons resultados. A limitação desse método reside na sua grande dependência do parâmetro da frequência de corte, o qual não podemos estimar previamente sem um corpus anotado de referência.

Nossa implementação do método de Hu & Liu obteve resultados mais equilibrados, isto é, sem tanta variação com relação a frequência de corte. A implementação do sistema de corte de aspectos não relacionados utilizando o modelo distribucional *Word2Vec*, mostrou-se efetiva elevando os resultados de ambos métodos.

Na identificação de sentimentos, desenvolvemos um método que permite a criação de léxicos sensíveis ao contexto, ou seja, apropriados para a identificação de sentimentos em textos pertencentes a uma mesma categoria, por exemplo, no nosso caso textos avaliativos sobre livros. Dados os resultados dos experimentos, o léxico criado mostrou-se eficaz, e seu método de criação pode ser utilizado facilmente na construção de léxicos focados na análise de contextos diferentes. A combinação de léxicos também se mostrou efetiva, sendo uma maneira simples de contornar a mencionada escassez de recursos na língua portuguesa.

Com relação aos métodos utilizados para a identificação de sentimentos, nosso método Preferência aos Adjetivos, foi ligeiramente superior aos métodos com os quais foi comparado. Entretanto, melhorias podem ser implementadas em suas duas etapas de

avaliação, melhorando seus resultados.

6.1 Dificuldades e Limitações

A principal dificuldade e limitação encontrada no desenvolvimento deste projeto foi a ausência de outros conjuntos de dados com textos opinativos marcados em aspectos e sentimentos.

6.2 Contribuições

Durante o mestrado, o estudante escreveu um artigo sobre mineração de tópicos e aspectos em mensagens da rede social *Twitter* sobre dengue, chikungunya, zika e microcefalia, que foi aceito para o 17º Workshop de Informática Médica (WIM 2017) (MACHADO et al., 2017). Um segundo trabalho sobre métodos não supervisionados para identificação de aspectos foi aceito para o 14º Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2017) (MACHADO; RUIZ; PARDO, 2017). Esses dois trabalhos estão relacionados ao Capítulo 4. Outro trabalho sobre identificação de sentimentos (Capítulo 5) está em desenvolvimento.

Este projeto contribuiu com suas aplicações, avaliações e melhorias implementadas em métodos de análise de sentimentos baseada em aspectos. Os conjuntos de resultados obtidos nos experimentos podem servir como base para direcionar e justificar o uso dos métodos de identificação de aspectos ou sentimentos, além de apontar pontos fortes e fracos em cada abordagem. Especificamente para a identificação de sentimentos, nossa principal contribuição foi o método para criação de léxicos de sentimentos sensíveis ao contexto. Tal método pode ser facilmente aplicado na criação de léxicos relacionados a diferentes temas. Os métodos de criação e combinação de léxicos, foram importantes contribuições dada a escassez de recursos na língua portuguesa, e por serem formas simples e eficazes na obtenção de melhores resultados.

6.3 Trabalhos Futuros

Como trabalhos futuros, pretende-se: (i) avaliar os métodos de análise de sentimentos baseada em aspectos em outros conjuntos de dados, provavelmente em outro idioma, dada a inexistência de outro corpus anotado em português, (ii) implementar melhorias na detecção dos sintagmas nominais nos métodos de identificação de aspectos, (iii) na identificação de sentimentos, buscar outras formas de utilização do modelo distribucional

Word2Vec para a identificação de polaridade das palavras e (iv) avaliar o método de criação de léxicos com outros conjuntos de dados.

Referências

- ASUR, S.; HUBERMAN, B. A. Predicting the future with social media. In: *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. [S.l.: s.n.], 2010. v. 1, p. 492–499.
- BABOUR, A.; KHAN, J. I. Tweet sentiment analytics with context sensitive tone-word lexicon. In: IEEE COMPUTER SOCIETY. *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01*. [S.l.], 2014. p. 392–399.
- DING, X.; LIU, B.; YU, P. S. A holistic lexicon-based approach to opinion mining. In: ACM. *Proceedings of the 2008 International Conference on Web Search and Data Mining*. [S.l.], 2008. p. 231–240.
- FILHO, P. P. B.; PARDO, T. A. NILC USP: A hybrid system for sentiment analysis in twitter messages. In: CITESEER. *Second Joint Conference on Lexical and Computational Semantics (* SEM)*. [S.l.], 2013. v. 2, p. 568–572.
- FILHO, P. P. B.; PARDO, T. A. S. Buscaopinioes: Searching for opinions over the internet. In: UNIVERSIDADE DE SÃO PAULO–USP. *International Conference on Computational Processing of the Portuguese Language, 11; Software Demonstration Workshop*. [S.l.], 2014.
- FILHO, P. P. B.; PARDO, T. A. S.; ALUÍSIO, S. M. An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. *9th Brazilian Symposium in Information and Human Language Technology – STIL*, p. 215–219, 2013. ISSN 21565570.
- FONSECA, E. R.; ROSA, J. L. G. Mac-Morpho revisited: Towards robust part-of-speech tagging. In: *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*. [S.l.: s.n.], 2013. p. 98–107.
- FREITAS, C. et al. Vampiro Que Brilha... Rá! Desafios Na Anotação De Opinião Em Um Corpus De Resenhas De Livros. In: *XI Encontro de Linguística de Corpus (ELC 2012)*. São Paulo: [s.n.], 2012.
- GANU, G.; ELHADAD, N.; MARIAN, A. Beyond the Stars: Improving Rating Predictions using Review Text Content. In: *WebDB*. [S.l.: s.n.], 2009. v. 9, p. 1–6.
- HARTMANN, N. et al. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In: *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*. [S.l.]: Sociedade Brasileira de Computação, 2017. p. 122–131.

- HARTMANN, N. S. et al. A large corpus of product reviews in portuguese: Tackling out-of-vocabulary words. In: EUROPEAN LANGUAGE RESOURCES ASSOCIATION-ELRA. *International Conference on Language Resources and Evaluation, 9th.* [S.l.], 2014.
- HONG, Y.; SKIENA, S. The wisdom of bookies? Sentiment analysis vs. the NFL point spread. In: *Proceedings of the International Conference on Weblogs and Social Media (icWSm-2010)*. [S.l.: s.n.], 2010.
- HRIDOY, S. A. A. et al. Localized Twitter opinion mining using sentiment analysis. *Decision Analytics*, v. 2, n. 1, p. 8, 2015. ISSN 2193-8636.
- HU, M.; LIU, B. Mining and summarizing customer reviews. In: ACM. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.], 2004. p. 168–177.
- HU, M.; LIU, B. Mining Opinion Features in Customer Reviews. *19th National Conference on Artificial intelligence*, p. 755–760, 2004.
- KOLCHYNA, O. et al. Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination. p. 32, Jul 2015.
- KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, v. 40, n. 10, p. 4065–4074, 2013. ISSN 09574174.
- KOULOUMPIS, E.; WILSON, T.; MOORE, J. Twitter Sentiment Analysis : The Good the Bad and the OMG ! *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, p. 538–541, 2011. ISSN 03781097.
- LIU, B. Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, v. 2, p. 627–666, 2010.
- LIU, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.
- LIU, B.; HU, M.; CHENG, J. Opinion observer: Analyzing and comparing opinions on the web. In: ACM. *Proceedings of the 14th International Conference on World Wide Web*. [S.l.], 2005. p. 342–351.
- LIU, K.-L.; LI, W.-J.; GUO, M. Emoticon Smoothed Language Models for Twitter Sentiment Analysis. *AAAI*, p. 1678–1684, 2012.
- LIU, Y. et al. Arsa: A sentiment-aware model for predicting sales performance using blogs. In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2007. (SIGIR '07), p. 607–614.
- MACHADO, M. T.; RUIZ, E. E. S.; PARDO, T. A. S. Analysis of unsupervised aspect term identification methods for portuguese reviews. In: *XIV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*. Uberlândia, MG: 6th Brazilian Conference on Intelligent Systems (BRACIS), 2017. p. 239–249.

- MACHADO, M. T. et al. Mineração de tópicos e aspectos em microblogs sobre dengue, chikungunya, zika e microcefalia. In: *XVII Workshop de Informática Médica (WIM)*. São Paulo: XXXVII Congresso da Sociedade Brasileira de Computação, 2017. p. 2058–2067.
- MIKOLOV, T. et al. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, p. 1–12, 2013. ISSN 15324435.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, Now Publishers Inc., v. 2, n. 1-2, p. 1–135, 2008.
- PAVLOPOULOS, I. *Aspect based sentiment analysis*. Tese (Doutorado) — Department of Informatics, Athens University of Economics and Business, 2014.
- PAVLOPOULOS, J.; ANDROUTSOPOULOS, I. Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method. In: *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)@ EACL*. [S.l.: s.n.], 2014. p. 44–52. ISBN 9781937284923.
- PENNEBAKER, J. W.; FRANCIS, M. E.; BOOTH, R. J. Linguistic Inquiry and Word Count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, v. 71, n. 2001, p. 2001, 2001.
- PONTIKI, M. et al. Semeval-2015 task 12: Aspect based sentiment analysis. In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, 2015. p. 486–495.
- POPESCU, A.-M.; ETZIONI, O. Extracting product features and opinions from reviews. In: *Natural language processing and text mining*. [S.l.]: Springer, 2007. p. 9–28.
- PSOMAKELIS, E. et al. Comparing methods for twitter sentiment analysis. *KDIR 2014 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, p. 225–232, 2014.
- ROSSI, R. G. et al. Inductive model generation for text classification using a bipartite heterogeneous network. *Journal of Computer Science and Technology*, Springer, v. 29, n. 3, p. 361–375, 2014.
- SAIF, H.; HE, Y.; ALANI, H. Alleviating data sparsity for twitter sentiment analysis. In: *2nd Workshop on Making Sense of Microposts (#MSM2012): Big things come in small packages at the 21st International Conference on the World Wide Web (WWW'12)*. Lyon, France: CEUR Workshop Proceedings (CEUR-WS. org), 2012. p. 2–9.
- SEBASTIANI, F. Machine learning in automated text categorization. *ACM Computing Surveys*, ACM, v. 34, n. 1, p. 1–47, 2002.
- SILVA, M. J. et al. Automatic Expansion of a Social Judgment Lexicon for Sentiment Analysis. *Technical Report TR 1008 University of Lisbon Faculty of Sciences LASIGE*, v. 6694, n. December, 2010.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, Elsevier Ltd, v. 45, n. 4, p. 427–437, 2009. ISSN 03064573.

- SOUZA, M.; VIEIRA, R. Sentiment Analysis on Twitter Data for Portuguese Language. *Proceedings of the 10th international conference on Computational Processing of the Portuguese Language*, p. 241–247, 2012. ISSN 03029743.
- SOUZA, M. et al. Construction of a Portuguese Opinion Lexicon from multiple resources. *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*, p. 59–66, 2011.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: principles and methods. *Data & knowledge engineering*, Elsevier, v. 25, n. 1, p. 161–197, 1998.
- TABOADA, M. et al. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, v. 37, n. 2, p. 267–307, 2011. ISSN 0891-2017.
- THET, T. T.; NA, J.-C.; KHOO, C. S. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science*, Sage Publications, p. 823–848, 2010.
- TUMASJAN, A. et al. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, v. 10, p. 178–185, 2010.
- TURING, A. M. Computing machinery and intelligence. *Mind*, JSTOR, v. 59, n. 236, p. 433–460, 1950.
- VOSOUGHI, S.; ZHOU, H.; ROY, d. Enhanced twitter sentiment classification using contextual information. In: *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Lisboa, Portugal: Association for Computational Linguistics, 2015. p. 16–24.
- WEI, C.-P. et al. Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews. *Information Systems and E-Business Management*, Springer, v. 8, n. 2, p. 149–167, 2010.

Anexos

Figura 11 – Resultados Hu & Liu corpus ReLi com sintagmas.

Tabela 29 – Resultados Hu & Liu corpus ReLi estendido com sintagmas

Frequência de Corte	Precisão	Revocação	Medida F
1%	35.54	56.01	43.49
2%	40.26	56.97	47.18
3%	41.79	48.67	44.97
4%	43.35	46.51	44.88
5%	44.64	46.76	45.68
6%	44.64	46.76	45.68
7%	44.64	46.76	45.68
8%	42.89	42.93	42.91
9%	42.89	42.93	42.91
10%	42.89	42.93	42.91

Valores de Precisão, Revocação e Medida F dados em Porcentagem

Figura 12 – Resultados Hu & Liu corpus ReLi estendido com sintagmas.

Figura 13 – Resultados Hu & Liu corpus ReLi estendido sem sintagmas.

Tabela 30 – Resultados Hu & Liu corpus ReLi estendido sem sintagmas

Frequência de Corte	Precisão	Revocação	Medida F
1%	42.18	59.05	49.21
2%	49.14	58.72	53.51
3%	53.49	50.68	52.05
4%	57.15	49.40	52.99
5%	59.07	49.76	54.02
6%	59.07	49.76	54.02
7%	59.07	49.76	54.02
8%	57.28	44.90	50.34
9%	57.28	44.90	50.34
10%	57.28	44.90	50.34

Valores de Precisão, Revocação e Medida F dados em Porcentagem

Tabela 31 – Resultados Preferência aos Adjetivos *Word2Vec* Versão 1

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,13	65,58	73,32	43,95	42,15	43,03	60,45
SentiLex	83,68	57,20	67,96	46,00	48,12	47,04	55,21
LIWC	82,69	71,33	76,59	37,54	43,17	40,16	65,16
LexReLi	82,04	82,19	82,11	58,89	34,47	43,49	71,74
Conciliado	83,11	70,90	76,52	46,04	48,63	47,30	66,02
LIWC+Op+Sent	82,64	82,72	82,68	48,44	42,32	45,17	73,87
LIWC+Sent+Op	83,40	82,00	82,69	48,47	45,90	47,15	74,09
Op+LIWC+Sent	82,08	81,14	81,61	45,57	41,30	43,33	72,41
Op+Sent+LIWC	82,23	74,44	78,14	44,82	43,52	44,16	67,66
Sent+LIWC+Op	83,09	72,24	77,29	46,96	50,00	48,43	67,36
Sent+Op+LIWC	83,06	71,57	76,88	46,13	49,83	47,91	66,80
LexReLi+LIWC+Op+Sent	81,56	90,00	85,57	57,57	36,35	44,56	78,24
LexReLi+LIWC+Sent+Op	81,62	89,95	85,58	57,03	36,69	44,65	78,28
LexReLi+Op+LIWC+Sent	81,59	89,76	85,48	56,88	36,69	44,61	78,13
LexReLi+Op+Sent+LIWC	81,97	86,84	84,33	55,81	37,71	45,01	76,07
LexReLi+Sent+LIWC+Op	82,01	86,84	84,35	55,61	38,05	45,19	76,15
LexReLi+Sent+Op+LIWC	82,03	86,74	84,32	55,33	38,05	45,10	76,07

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Tabela 32 – Resultados Preferência aos Adjetivos *Word2Vec* Versão 2

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,04	65,87	73,47	44,50	42,15	43,30	60,67
SentiLex	83,62	58,16	68,61	48,05	48,29	48,17	56,00
LIWC	81,99	72,14	76,75	36,98	39,76	38,32	65,05
LexReLi	81,74	82,29	82,01	58,56	33,28	42,44	71,55
Conciliado	83,04	71,47	76,82	46,76	47,95	47,35	66,32
LIWC+Op+Sent	82,34	81,91	82,12	46,55	41,47	43,86	73,05
LIWC+Sent+Op	83,21	81,86	82,53	47,99	44,88	46,38	73,76
Op+LIWC+Sent	82,03	81,28	81,65	45,32	40,44	42,74	72,34
Op+Sent+LIWC	82,41	74,92	78,49	45,37	43,52	44,43	68,04
Sent+LIWC+Op	82,89	73,05	77,66	48,09	49,32	48,69	67,85
Sent+Op+LIWC	83,10	72,47	77,42	47,80	50,17	48,96	67,59
LexReLi+LIWC+Op+Sent	81,17	89,95	85,33	56,55	34,64	42,96	77,83
LexReLi+LIWC+Sent+Op	81,42	90,00	85,49	56,91	35,84	43,98	78,13
LexReLi+Op+LIWC+Sent	81,23	89,71	85,26	55,98	35,15	43,19	77,76
LexReLi+Op+Sent+LIWC	81,47	86,74	84,03	55,18	36,35	43,83	75,70
LexReLi+Sent+LIWC+Op	81,70	86,79	84,17	55,84	37,54	44,90	76,00
LexReLi+Sent+Op+LIWC	81,72	86,69	84,13	55,56	37,54	44,81	75,93

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Tabela 33 – Resultados Aspectos + Somente Adjetivos *Word2Vec* Versão 1

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,73	62,09	71,3	47,01	44,2	45,56	58,17
SentiLex	85,88	59,12	70,03	46,77	51,88	49,19	57,53
LIWC	83,65	54,86	66,26	39,91	45,56	42,55	52,82
LexReLi	83,18	69,6	75,79	53,73	41,81	47,02	63,51
Conciliado	85,59	61,7	71,71	48,51	50,0	49,24	59,14
LIWC+Op+Sent	84,76	64,43	73,21	49,82	46,76	48,24	60,56
LIWC+Sent+Op	85,37	62,57	72,21	49,16	50,0	49,58	59,81
Op+LIWC+Sent	83,89	62,33	71,52	47,63	44,54	46,03	58,43
Op+Sent+LIWC	83,89	62,33	71,52	47,44	44,2	45,76	58,36
Sent+LIWC+Op	85,55	60,08	70,58	47,83	50,68	49,21	58,02
Sent+Op+LIWC	85,49	60,08	70,57	48,2	50,34	49,25	57,94
LexReLi+LIWC+Op+Sent	83,88	69,51	76,02	54,72	43,52	48,48	63,81
LexReLi+LIWC+Sent+Op	83,95	69,12	75,82	54,09	44,03	48,54	63,63
LexReLi+Op+LIWC+Sent	83,92	69,46	76,01	54,7	43,69	48,58	63,81
LexReLi+Op+Sent+LIWC	83,92	69,46	76,01	54,51	43,34	48,29	63,74
LexReLi+Sent+LIWC+Op	83,97	68,98	75,74	53,77	43,86	48,31	63,48
LexReLi+Sent+Op+LIWC	83,97	68,98	75,74	53,77	43,86	48,31	63,48

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Tabela 34 – Resultados Aspectos + Somente Adjetivos *Word2Vec* Versão 2

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	84,35	56,25	67,49	53,12	39,25	45,14	52,52
SentiLex	86,47	48,64	62,25	55,58	45,05	49,76	47,85
LIWC	82,41	36,33	50,43	53,91	21,16	30,39	33,01
LexReLi	83,27	68,84	75,37	57,07	37,88	45,54	62,06
Conciliado	86,14	52,08	64,92	54,98	43,34	48,47	50,17
LIWC+Op+Sent	85,35	59,98	70,45	54,88	43,17	48,33	56,30
LIWC+Sent+Op	86,13	58,54	69,71	54,79	45,90	49,95	55,78
Op+LIWC+Sent	84,57	56,68	67,87	53,33	39,59	45,45	52,93
Op+Sent+LIWC	84,45	56,68	67,83	53,12	39,25	45,14	52,86
Sent+LIWC+Op	85,78	54,57	66,71	54,33	46,08	49,86	52,71
Sent+Op+LIWC	86,12	53,18	65,76	54,83	45,56	49,77	51,51
LexReLi+LIWC+Op+Sent	84,01	69,17	75,87	55,45	41,64	47,56	63,14
LexReLi+LIWC+Sent+Op	84,08	68,79	75,67	55,23	42,32	47,92	62,99
LexReLi+Op+LIWC+Sent	84,05	69,12	75,86	55,33	41,64	47,52	63,10
LexReLi+Op+Sent+LIWC	83,95	69,12	75,82	55,13	41,30	47,22	63,03
LexReLi+Sent+LIWC+Op	84,02	68,69	75,59	54,89	42,15	47,68	62,88
LexReLi+Sent+Op+LIWC	84,02	68,69	75,59	54,89	42,15	47,68	62,88

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Tabela 35 – Resultados Aspectos + Preferência aos Adjetivos *Word2Vec* Versão 1

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,69	66,59	74,17	44,79	46,25	45,51	62,13
SentiLex	84,73	58,98	69,55	48,40	54,10	51,09	57,91
LIWC	82,72	72,62	77,34	38,84	43,34	40,97	66,21
LexReLi	82,57	80,71	81,63	55,66	41,13	47,30	72,04
Conciliado	84,09	71,09	77,04	46,21	53,07	49,40	67,14
LIWC+Op+Sent	83,70	81,62	82,65	47,82	48,63	48,22	74,39
LIWC+Sent+Op	84,21	80,66	82,40	47,34	51,71	49,43	74,32
Op+LIWC+Sent	83,04	80,13	81,56	45,26	46,42	45,83	72,75
Op+Sent+LIWC	82,99	74,25	78,37	44,27	47,44	45,80	68,37
Sent+LIWC+Op	83,97	71,95	77,49	46,38	54,61	50,16	68,15
Sent+Op+LIWC	84,00	71,37	77,17	45,91	54,61	49,88	67,70
LexReLi+LIWC+Op+Sent	82,65	87,31	84,92	53,58	44,71	48,74	77,98
LexReLi+LIWC+Sent+Op	82,71	87,03	84,81	52,89	45,22	48,76	77,87
LexReLi+Op+LIWC+Sent	82,66	87,17	84,86	53,24	44,88	48,70	77,91
LexReLi+Op+Sent+LIWC	82,79	84,49	83,63	52,51	44,71	48,29	75,78
LexReLi+Sent+LIWC+Op	82,82	84,20	83,50	51,95	45,39	48,45	75,70
LexReLi+Sent+Op+LIWC	82,84	84,11	83,47	51,75	45,39	48,36	75,63

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem

Tabela 36 – Resultados Aspectos + Preferência aos Adjetivos *Word2Vec* Versão 2

Léxico	Positivo			Negativo			Acurácia
	Prec	Rev	F	Prec	Rev	F	
OpLexicon	83,83	67,26	74,63	46,12	47,61	46,85	62,95
SentiLex	84,80	59,02	69,60	48,42	54,95	51,48	58,13
LIWC	80,49	81,19	80,84	43,80	28,33	34,40	69,61
LexReLi	82,08	81,76	81,92	56,85	38,23	45,71	72,22
Conciliado	83,68	72,14	77,48	46,88	51,37	49,02	67,59
LIWC+Op+Sent	83,35	83,63	83,49	50,28	46,76	48,45	75,55
LIWC+Sent+Op	83,96	83,20	83,58	50,69	50,00	50,34	75,93
Op+LIWC+Sent	82,84	82,24	82,54	47,57	45,05	46,28	74,09
Op+Sent+LIWC	82,74	75,73	79,08	45,64	46,42	46,02	69,31
Sent+LIWC+Op	83,65	73,72	78,37	47,83	52,73	50,16	69,12
Sent+Op+LIWC	83,85	73,05	78,08	47,58	53,75	50,48	68,82
LexReLi+LIWC+Op+Sent	82,38	87,94	85,07	54,06	43,17	48,01	78,13
LexReLi+LIWC+Sent+Op	82,46	87,75	85,02	53,67	43,69	48,17	78,09
LexReLi+Op+LIWC+Sent	82,35	87,79	84,99	53,60	43,17	47,83	78,02
LexReLi+Op+Sent+LIWC	82,43	84,87	83,63	52,28	43,00	47,19	75,70
LexReLi+Sent+LIWC+Op	82,47	84,68	83,56	52,03	43,69	47,50	75,70
LexReLi+Sent+Op+LIWC	82,53	84,59	83,55	51,92	43,86	47,55	75,66

Precisão (Prec), Revocação (Rev), Medida F (F) e Acurácia dados em Porcentagem