



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS QUIXADÁ
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

RÔMULO CÉSAR COSTA DE SOUSA

**IDENTIFICANDO SENTIMENTOS DE TEXTOS EM PORTUGUÊS COM O
SENTIWORDNET TRADUZIDO**

QUIXADÁ – CEARÁ

2016

RÔMULO CÉSAR COSTA DE SOUSA

IDENTIFICANDO SENTIMENTOS DE TEXTOS EM PORTUGUÊS COM O
SENTIWORDNET TRADUZIDO

Monografia apresentada no curso de Ciência da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Ciência da Computação. Área de concentração: Computação.

Orientador: Prof. Dr. Paulo de Tarso Guerra Oliveira

QUIXADÁ – CEARÁ

2016

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S698i Sousa, Rômulo César Costa de.
Identificando sentimentos de texto em português com o SentiWordNet traduzido / Rômulo César Costa de Sousa. – 2016.
49 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Ciência da Computação, Quixadá, 2016.
Orientação: Prof. Dr. Paulo de Tarso Guerra Oliveira.

1. Linguística - Processamento de dados. 2. Mineração de dados. 3. Algoritmos. 4. Tradução. I. Título.
CDD 004

RÔMULO CÉSAR COSTA DE SOUSA

IDENTIFICANDO SENTIMENTOS DE TEXTOS EM PORTUGUÊS COM O
SENTIWORDNET TRADUZIDO

Monografia apresentada no curso de Ciência da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Ciência da Computação. Área de concentração: Computação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Paulo de Tarso Guerra Oliveira (Orientador)
Campus Quixadá
Universidade Federal do Ceará – UFC

Profa. Me. Lívia Almada Cruz
Campus Quixadá
Universidade Federal do Ceará - UFC

Prof. Me. Victor Aguiar Evangelista de Farias
Campus Quixadá
Universidade Federal do Ceará - UFC

A Deus.

Aos meus pais, Ocinele e Antônio César.

AGRADECIMENTOS

A Deus por ter me dado saúde e força para superar as dificuldades.

A toda minha família que, com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa.

A todos os professores do curso de ciência da computação, que foram tão importantes na minha vida acadêmica.

Ao professor Paulo de Tarso pela paciência na orientação e o grande incentivo que tornaram possível a realização desse trabalho.

Aos meus amigos e colegas de graduação André Davys, Jonas Costa, Sergio Filho, Ana Paula Dantas, Jhonata Matias e Wallinson Deives Batista.

“Tente mover o mundo - o primeiro passo será mover a si mesmo.”

(Platão)

RESUMO

Análise de Sentimento é o campo de estudo que analisa opiniões descritas de forma textual. Pesquisas na área de análise de sentimentos frequentemente são aplicadas a textos em inglês e línguas como o português ainda são pouco exploradas. Nesse trabalho, é apresentada a implementação de classificadores semânticos que realizam uma análise de sentimentos em textos escritos em português. Utilizamos nessa classificação o recurso léxico SentiWordNet, traduzido de forma automática para o português, e comparamos o seu desempenho em relação ao SentiLex, recurso léxico já em português. Os textos analisados são comentários do site Tripadvisor, um portal online de viagens que fornece informações e opiniões de conteúdos relacionados ao turismo. Os resultados mostram que a classificação usando o SentiWordNet traduzido tem, em muitos casos, desempenho melhor.

Palavras-chave: Análise de Sentimentos. Classificação. SentiWordNet. Tradução Automática.

ABSTRACT

Sentiment Analysis is the field of study that examines opinion written in text. Researches in the field of sentiment analysis is often applied to english texts, and other languages like Brazilian portuguese are little explored. In this study it's shown an implementation to semantic classifiers that analyze sentiments on texts written in portuguese. On these classifications, We use the lexical resource SentiWordNet, translated automatically to portuguese, and compare it's performance with SentiLex, a lexical resource in portuguese. The analyses texts are comments are extracted from the website TripAdvisor, an online travel portal that provides informations and opinions related to tourism. The results show that the classification using the translated SentiWordNet has, in most of the cases, a better performance.

Keywords: Sentiment Analysis. Classification. SentiWordNet. Automatic Translation.

LISTA DE FIGURAS

Figura 1 – Exemplo de uma SVM com duas classes	17
Figura 2 – Exemplo de <i>synset</i> no SentiWordNet	19
Figura 3 – Exemplo de requisição HTTP para Google API em python	23
Figura 4 – Exemplo de Json resposta da requisição	24
Figura 5 – Exemplo de <i>synset</i> SentiWordNet traduzido	24
Figura 6 – Classificação de polaridade	33
Figura 7 – Exemplo negação	35

LISTA DE TABELAS

Tabela 1 – Aplicação do Tree-Tagger	25
Tabela 2 – Etiquetas do Tree-Tagger	25
Tabela 3 – Uso do Tree-Tagger no exemplo	35
Tabela 4 – Exemplo da interpretação do SentiWordNet traduzido	36
Tabela 5 – Avaliação das configurações #1, #2, #3 e #4	39
Tabela 6 – Avaliação das configurações #5, #6, #7 e #8	39
Tabela 7 – Avaliação geral das configurações #1, #2, #3 e #4	40
Tabela 8 – Avaliação geral das configurações #5, #6, #7 e #8	40
Tabela 9 – Comparação das avaliações	40
Tabela 10 – Avaliação das configurações #1 e #2	47
Tabela 11 – Avaliação das configurações #3 e #4	48
Tabela 12 – Avaliação das configurações #5 e #6	48
Tabela 13 – Avaliação das configurações #7 e #8	48
Tabela 14 – Avaliação geral das configurações #1 e #2	48
Tabela 15 – Avaliação geral das configurações #3 e #4	49
Tabela 16 – Avaliação geral das configurações #5 e #6	49
Tabela 17 – Avaliação geral das configurações #7 e #8	49

LISTA DE ALGORITMOS

Algoritmo 1	–	<i>Soma das pontuações dos termos</i>	26
Algoritmo 2	–	<i>Média das sentenças e média do comentário</i>	28
Algoritmo 3	–	<i>Método base</i>	30
Algoritmo 4	–	<i>Posição dos adjetivos</i>	32

SUMÁRIO

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Processamento de Linguagem Natural	14
2.2	Análise de Sentimentos	15
2.3	Classificadores de Sentimentos	16
2.3.1	<i>Classificadores de Aprendizagem de Máquina</i>	16
2.3.2	<i>Classificadores Semânticos</i>	18
2.4	Recursos Léxicos	18
2.4.1	<i>SentiWordNet</i>	18
2.5	Regras Linguísticas de Negação	19
3	TRABALHOS RELACIONADOS	20
3.1	Classificação por meio de Aprendizagem de Máquina	20
3.2	Classificação Semântica	21
4	IDENTIFICANDO EMOÇÕES COM O SENTIWORDNET	23
4.1	Tradução automática do SentiWordNet	23
4.2	Classificação gramatical e simplificação do texto	24
4.3	Classificadores de polaridade	26
4.3.1	<i>Soma das pontuações dos termos</i>	26
4.3.2	<i>Média das sentenças e média do comentário</i>	27
4.3.3	<i>Método base</i>	29
4.3.4	<i>Posição dos adjetivos</i>	31
4.4	Implementação	33
4.5	Exemplo de classificação	35
5	RESULTADOS	37
5.1	Métodos de avaliação	37
5.2	Resultados	38
6	DISCUSSÃO	42
7	CONSIDERAÇÕES FINAIS	44
	REFERÊNCIAS	45
	APÊNDICE A – AVALIAÇÕES	47

1 INTRODUÇÃO

Com a popularização das redes sociais, blogs e portais de opiniões, é cada vez mais comum que usuários possam enviar e compartilhar mensagens que expressam sentimentos. Junto com o crescimento dessas ferramentas, também tem crescido a atenção dada à extração de informações em mensagens que expressam opiniões. Milhões de mensagens escritas em vários idiomas são enviadas todos os dias, mensagens essas que contêm informações úteis e que poderiam ser usadas para muito mais do que apenas comunicação. Extrair informações dessas mensagens através de um processo automático pode ser muito útil (DUARTE, 2013).

O campo de estudo que analisa as opiniões e sentimentos das pessoas em textos chama-se **análise de sentimento** (MEDHAT; HASSAN; KORASHY, 2014). O interesse da indústria e da academia neste campo de estudo é, em parte, devido ao seu potencial de aplicações, tais como: marketing, relações públicas e campanhas políticas (FREITAS; VIEIRA, 2015). Empresas e organizações podem estar interessadas nas seguintes perguntas:

- O que as pessoas pensam sobre seus produtos, serviços etc?
- O quão positivo (ou negativo) as pessoas pensam sobre seus produtos?
- Como as pessoas preferem seus produtos?

Partidos políticos podem estar interessados em saber se as pessoas apoiam ou não seu plano de governo. Organizações sociais podem estar interessadas na opinião das pessoas sobre assuntos atuais. Essas informações poderiam, por exemplo, serem extraídas de textos originados em redes sociais, *blogs* e portais de opiniões, pois os usuários desses serviços, todos os dias, publicam textos expressando opiniões sobre diversos assuntos (PAK; PAROUBEK, 2010).

Freitas e Vieira (2015) afirmam que pesquisas na área de análise de sentimentos frequentemente são aplicadas a textos em inglês e que outras línguas são menos exploradas. No Brasil, pesquisas nesta área ainda estão em fase inicial, mas a comunidade científica está fazendo um esforço para que recursos e técnicas a serem utilizadas nestas tarefas sejam criados. A criação de novos recursos léxicos para a língua portuguesa é um exemplo desse esforço. Segundo Freitas e Vieira (2015), temos atualmente apenas quatro léxicos escritos para a língua portuguesa: OpLexicon¹, SentiLex², Brazilian Portuguese Linguistic Inquiry and Word Count (LIWC)³ e Onto.PT⁴. No entanto, o Brasil é o quinto país com o maior número de usuários na Internet

¹ <http://ontolp.inf.pucrs.br/Recursos/downloads-OpLexicon.php>

² http://dmir.inesc-id.pt/project/SentiLex-PT_01

³ <http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>

⁴ <http://ontopt.dei.uc.pt/>

(MUNDIAL, 2016), o que motiva diretamente o estudo de análise de sentimentos em textos escritos em português.

Diversos trabalhos foram desenvolvidos na área de análise de sentimentos. Dosciatti e Ferreira (2013) aplicam algoritmos de aprendizagem de máquina para classificar textos jornalísticos escritos em português do Brasil. A classificação é feita segundo as emoções expressas no texto. Neethu e Rajasree (2013) constroem alguns classificadores, usando varias técnicas de aprendizagem de máquina para identificar a polaridade de mensagens escritas em inglês, as mensagens são retiradas do Twitter. No trabalho de Freitas e Vieira (2015), foram analisados comentários escritos em português retirados do TripAdvisor⁵, um portal *online* de viagens que fornece informações e opiniões de conteúdos relacionados ao turismo. Para realizar essa análise as autoras construíram um classificador semântico usando os recursos léxicos disponíveis para o português.

Este trabalho propõe a implementação de classificadores semânticos para realizar análise de sentimentos de comentários escritos em português coletados do site TripAdvisor. A análise de sentimentos classifica os comentários em duas classes: positivo ou negativo, de acordo com o sentimento expresso no texto. O classificador é construído usando o recurso léxico SentiWordNet, uma ferramenta que atribui uma nota para o grau de positividade, negatividade e objetividade das palavras.

O SentiWordNet está disponível apenas em inglês, por isso, neste trabalho, esse recurso foi traduzido para o português de forma automática. Em comparação com os recursos léxicos disponíveis em português, o SentiWordNet contém um maior número de palavras, além de ser um recurso muito usado em trabalhos de classificação de sentimentos.

O objetivo desse trabalho é verificar a viabilidade e o desempenho do uso do recurso léxico SentiWorNet traduzido de forma automática para o português.

O restante desse trabalho é dividido da seguinte maneira: o Capítulo 2 apresenta os conceitos de processamento de linguagem natural, análise de sentimentos, classificadores de sentimentos e recursos léxicos; o Capítulo 3 mostra alguns trabalhos relacionados a identificação de sentimentos em texto; o Capítulo 4 apresenta como foi realizado a tradução do SentiWordNet, classificação gramatical do texto, os algoritmos de classificação e o processo de implementação; o Capítulo 5 apresenta os resultados obtidos; o Capítulo 6 traz algumas discussões sobre o resultados; e por fim, o Capítulo 7 apresenta as considerações finais sobre este trabalho.

⁵ <http://www.tripadvisor.com.br/>

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, é apresentada uma visão geral sobre os conceitos abordados neste trabalho. Nas subseções 2.1, 2.2, 2.3 e 2.4 são apresentados os conceitos relacionados a Processamento de Linguagem Natural, Análise de Sentimentos, Classificadores de Sentimentos e Recursos Léxicos, respectivamente.

2.1 Processamento de Linguagem Natural

Processamento de linguagem natural (PLN) é uma subárea de pesquisa da inteligência artificial que explora como os computadores podem ser usados para compreender e manipular linguagem natural, seja ela escrita ou falada. Pesquisadores da área de PLN têm como objetivo reunir o conhecimento sobre como os seres humanos entendem a linguagem natural para que ferramentas e técnicas computacionais para compreensão e manipulação da linguagem possam ser desenvolvidas (CHOWDHURY, 2003).

Coppin (2010) afirma que, cada vez mais, é preciso que computadores sejam capazes de entender linguagens naturais. Hoje em dia, por exemplo, muitos sistemas de telefonia são aptos a compreender um conjunto restrito de comandos e perguntas para ajudar clientes em centrais telefônicas, sem a intervenção humana. Além disso, a quantidade de dados textuais que existe na Internet alcançou proporções não gerenciáveis. Com isso, torna-se quase impossível fazer buscas nesse grande volume de dados usando apenas comandos SQL (*Structured Query Language*), por exemplo. A ideia é que as pessoas possam fazer perguntas em suas próprias linguagens.

Segundo Allen (2003), o principal desafio em PLN é a ambiguidade presente na linguagem humana. O trabalho de Brascher (2002) elenca algumas das principais causas de ambiguidade na linguagem natural:

- **Ambiguidade lexical:** a palavra “cobre”, pode ser referir ao metal cobre ou ao verbo cobrir.
- **Ambiguidade sintática:** “Maria leu a notícia sobre a greve no campus”, a frase pode significar que Maria leu sobre uma greve quando estava no campus ou significar que a greve ocorre no campus.
- **Ambiguidade semântica:** a frase “um rio corre através de cada país europeu” pode significar que um único rio corre através de todos os países ou diferentes

rios correm através de diferentes países.

- **Ambiguidade predicativa:** na frase “a crítica deste autor” não se sabe se o autor é objeto da crítica ou agente da crítica.
- **Ambiguidade pragmática:** na frase “Paulo vai à escola” não se sabe se Paulo é estudante ou se ele está indo à escola neste momento.

Allen (2003) argumenta que todas essas formas de ambiguidade podem tornar o PLN complexo. A ambiguidade torna a maioria das técnicas desenvolvidas para análise de linguagens formais ineficazes, quando aplicadas em linguagens com ambiguidade.

2.2 Análise de Sentimentos

As informações textuais de todo o mundo podem ser classificadas em dois grandes grupos principais: fatos e opiniões (LIU, 2010). Os **fatos** são expressões objetivas sobre entidades, eventos, produtos e etc. Já as **opiniões** são geralmente expressões que descrevem sentimentos e avaliações. **Análise de Sentimento** é o campo de estudo que analisa essas opiniões (FREITAS; VIEIRA, 2015). As técnicas de análise de sentimento são utilizadas para extrair e determinar automaticamente sentimentos e emoções expressos em linguagem natural (DUARTE, 2013). Neste trabalho, se extrairá dos textos apenas expressões de sentimentos como positivo e negativo.

Opiniões são importantes, pois, em alguns contextos, a tomada de decisão se torna mais fácil após uma análise de outras opiniões e pontos de vista (FREITAS; VIEIRA, 2015). Liu (2010) relata que, quando uma organização queria buscar opiniões do público sobre os seus produtos e serviços, essa organização realizava pesquisas de opinião e entrevistas. No entanto, com a expansão do volume de opiniões na *Web* nos últimos anos, esse processo vem mudando.

Em muitos casos, as opiniões estão escondidas entre as páginas dos *blogs* ou entre milhões de postagens diárias no Twitter, por exemplo. Por isso, a tarefa de procurar informações relevantes e extrair opiniões de textos da *Web* pode ser difícil para leitores humanos. Nesse contexto, a análise de sentimentos automatizada se faz necessária.

Análise de sentimentos pode também ser aplicada na bolsa de valores, notícias jornalísticas, debates políticos, etc (MEDHAT; HASSAN; KORASHY, 2014). Em debates políticos, por exemplo, é possível descobrir a opinião das pessoas sobre candidatos e partidos políticos, em uma eleição. Os resultados das eleições podem ser previstos a partir de postagens que falam sobre política.

A análise de sentimentos em texto não é uma atividade trivial. Uma simples busca por palavras-chave como “bom” e “ruim” não é suficiente para determinar o sentimento expresso no texto, por exemplo na frase “João não é um pessoa boa”, a presença da palavra “boa” não torna a sentença positiva (DOSCIATTI; FERREIRA, 2013). Por isso, técnicas avançadas para análise de sentimento precisam ser aplicadas.

Segundo Maynard e Funk (2011), as principais técnicas para classificação de sentimento são divididas em três abordagens: abordagem de aprendizagem de máquina, abordagem semântica e abordagem híbrida. A **abordagem de aprendizagem de máquina** se baseia na aplicação dos famosos algoritmos da área (naïve bayes, máquina de vetores de suporte, redes neurais etc). A **abordagem semântica** é baseada no uso de recursos léxicos, uma coleção de termos pré-compilados capaz de valores às palavras segundo seu grau de positividade e negatividade. Essa abordagem pode ser aplicada de duas maneiras, uma com uso de dicionários e recursos para atribuição de notas às palavras, e outra com o uso de corpus que utilizam métodos estáticos e semânticos para encontrar a polaridade. A **abordagem híbrida** combina as duas outras abordagens, essa técnica é utilizada na maioria dos métodos de análise de sentimentos.

2.3 Classificadores de Sentimentos

Existem várias maneiras de classificar sentimentos em textos. Os principais meios usam algoritmos de aprendizado de máquina e classificadores, com base na semântica do texto (DUARTE, 2013). A classificação é responsável por separar os textos em diferentes classes, como por exemplo positivos e negativos, no caso específico deste trabalho.

2.3.1 *Classificadores de Aprendizagem de Máquina*

Esse tipo de classificador faz uso de algoritmos de aprendizagem de máquina, que são treinados com um conjunto de dados previamente rotulados e, assim, são capazes de classificar uma nova instância de acordo o conhecimento adquirido (DUARTE, 2013).

Naive Bayes é o classificador de aprendizagem de máquina simples e comumente o mais utilizado (MAYNARD; FUNK, 2011). Esse classificador probabilístico é baseado no **teorema de Bayes**:

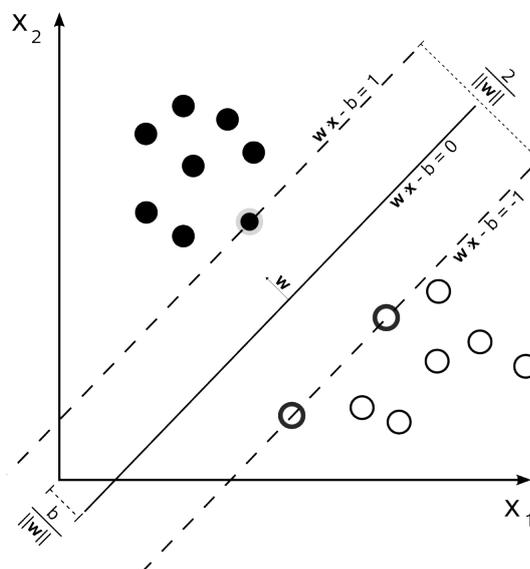
$$P(c|f_1...f_n) = \frac{P(c)P(f_1...f_n|c)}{P(f_1...f_n)}$$

Onde c é a classe e os f_n 's são as características levadas em consideração para realizar a classificação.

Para classificar um elemento é preciso que a probabilidade $P(c|f_1...f_n)$ seja calculada para cada classe c . Então o elemento pertencerá à classe com a maior probabilidade calculada. No caso especial da classificação de sentimentos, as classes poderiam ser representadas por positivo e negativo e as palavras presentes no texto poderiam representar as características levadas em consideração.

As **Máquina de Vetores de Suporte** (*support vector machine* ou **SVM**) é um outro classificador de aprendizagem de máquina muito utilizado. As SVM são embasadas pela teoria do aprendizado estatístico, desenvolvida por Vapnik (VAPNIK, 2013). Essa técnica converte as características relevantes para a classificação em um ponto disposto em um hiperplano. Esse hiperplano é dividido em várias classes, com base em um conjunto de treino inicial. A posição do ponto no hiperplano, perto ou longe de uma classe, indica a qual classe pertence uma determinada instância submetida ao classificador. Há também uma margem de separação entre classes, e os pontos contidos neste margem são tratados como neutros. A Figura 1 mostra um exemplo de SVM com duas classes.

Figura 1 – Exemplo de uma SVM com duas classes



Fonte: Fonte: Duarte (2013)

2.3.2 *Classificadores Semânticos*

Classificadores semânticos atribuem sentimentos aos textos usando dicionários, corpus e recursos léxicos que contêm palavras com polaridades previamente atribuídas. A classificação semântica leva em consideração a relação de uma determinada palavra com o restante do texto. Os vários sentidos de uma mesma palavra são levados em consideração para fornecer uma classificação mais específica (DUARTE, 2013). Classificação baseada em dicionários e classificação baseada em corpus são as duas técnicas mais usadas na classificação semântica.

Técnicas baseadas em dicionários usam sinônimos, antônimos e classes gramaticais das palavras combinados com recursos léxicos como OpLexicon, SentiLex, Onto.PT e SentiWordNet para determinar a polaridade de um texto. O uso de um bom recurso léxico é crucial para o bom desempenho dessa técnica. Já as técnicas baseadas em corpus usam grandes corpora linguísticos com o intuito de determinar para cada palavra um valor estático que definirá sua polaridade. Portanto, técnicas baseadas em corpus podem não ser tão eficientes quanto as técnicas baseadas em dicionário (DANG; ZHANG; CHEN, 2010). Nesse trabalho usamos classificadores semânticos para determinar a polaridade em textos.

2.4 Recursos Léxicos

Recursos léxicos são as principais ferramentas linguísticas empregadas na tarefa de análise de sentimentos. Eles são usados como dicionários de polaridade e são consultados no processo de atribuição de polaridade (positivo, negativo ou neutro). Em português, existem apenas quatro recursos léxicos: Oplexicon, LIWC, SentiLex, e Onto.PT (FREITAS; VIEIRA, 2015). Este trabalho utilizará outro recurso léxico, o SentiWordNet, descrito na próxima subseção.

2.4.1 *SentiWordNet*

O SentiWordNet é um recurso léxico escrito em inglês para a mineração de opiniões (PANG; LEE, 2008). Esse recurso léxico possui 117.374 entradas vindas de anotações automáticas de todos os *synsets* do WordNet 3.0. Um *synset* no WordNet representa um conjunto de palavras que são sinônimos entre si e um comentário que as descreve (CAVALCANTI et al., 2012). Cada *synset* presente no SentiWordNet está associado a três

pontuações numéricas Obj(s), Pos(s) e Neg(s), que descrevem o quão objetivo, positivo e negativo os termos contidos no *synset* são. Cada uma das três pontuações varia entre 0.0 e 1.0 e a sua soma total vale 1.0 para cada *synset*, pois $\text{Obj(s)} + \text{Pos(s)} + \text{Neg(s)} = 1$. A Figura 2 mostra um exemplo de um *synset* no SentiWordNet

Figura 2 – Exemplo de *synset* no SentiWordNet

```

synset
1 Pos: 0
2 Neg: 0.625
3 {ugly, horrifying, horrible, frightful, atrocious}
4 provoking horror

```

Fonte: Produzido pelo autor

O que motivou a atribuição de valores numéricos para um *synset*, ao invés de valores direto a um termo, foi a possibilidade de um mesmo termo ter diferentes sentidos e cada um desses sentidos merece uma pontuação diferente (ESULI; SEBASTIANI, 2006). No SentiWordNet, por exemplo, a palavra “*broken*” está relacionada a dois *synset*, $S1 = \{\textit{wiped out, impoverished, broken}\}$ que tem como descrição “*destroyed financially or the broken fortunes of the family*” e pontuações $\text{Obj(s)}=0.5$, $\text{Pos(s)}=0$ e $\text{Neg(s)}=0.5$, e $S2 = \{\textit{broken}\}$ que tem a descrição “*physically and forcibly separated into pieces or cracked or split*” e pontuações $\text{Obj(s)}=0.875$, $\text{Pos(s)}=0$ e $\text{Neg(s)}=0.125$. É possível notar que o SentiWordNet tem uma pontuação diferente para diferentes significados de um mesmo termo.

2.5 Regras Linguísticas de Negação

Freitas e Vieira (2015) afirmam que a negação é uma construção linguística que tem um papel importante na definição da polaridade de um texto. Segundo Schwenter (2005), no português do Brasil existem três maneiras de expressar negação, são elas:

- **Negação do tipo I:** onde a partícula de negação aparece antes do verbo. Exemplo:
A Cláudia não veio à festa.
- **Negação do tipo II:** onde a partícula de negação aparece antes e depois do verbo. Exemplo: A Cláudia não veio à festa não.
- **Negação do tipo III:** onde a partícula de negação aparece depois do verbo. Exemplo: A Cláudia veio à festa não.

Nesse trabalho identificaremos apenas as negações do tipo I.

3 TRABALHOS RELACIONADOS

3.1 Classificação por meio de Aprendizagem de Máquina

No trabalho de Dosciatti e Ferreira (2013), os autores apresentam uma abordagem que utiliza SVM para identificar emoções em textos escritos em português do Brasil. O trabalho usa um corpus de texto composto de notícias extraídas de um jornal *online* (a manchete da notícia e uma pequena descrição dela). Cada notícia do corpus é previamente rotulada com as seis emoções básicas: medo, surpresa, raiva, desgosto, tristeza e alegria. Os textos previamente rotulados são submetidos a um classificador MVS em configuração multi-classe, obtendo uma taxa de acerto de 61%.

Propomos desenvolver nesse trabalho um classificador de sentimentos para textos escritos em português do Brasil, assim como no trabalho de Dosciatti e Ferreira (2013), mas esse classificador será de apenas de duas classes (positivo e negativo).

Neethu e Rajasree (2013) classificam postagens retiradas do Twitter em duas classes: positivo e negativo, usando varias técnicas de aprendizagem de máquina. Os classificadores em geral encontram mais dificuldades para classificar textos vindos do Twitter, pois essas são geralmente mensagens curtas, cheias de gírias e erros ortográficos. Para contornar esse problema, os autores fazem um pré-processamento dos *tweets*, retirando os erros ortográficos e substituindo as gírias e abreviações por seus respectivos significados relacionados. As técnicas de aprendizagem de máquina: Naïve Bayes, Máquinas de Vetores de Suporte e Máxima Entropia são usadas na classificação.

Assim como no trabalho de Neethu e Rajasree (2013), neste trabalho pretendemos classificar textos em duas classes (positivo e negativo), porém a classificação será destinada a textos escritos em português do Brasil e não a *tweets* escritos em inglês como feito por eles.

Uma outra aplicação de análise de sentimentos em textos do Twitter é descrita no trabalho de Pak e Paroubek (2010). Eles afirmam que textos de *microblogging* hoje tornaram-se uma ferramenta de comunicação muito popular entre os usuários de Internet. Milhões de usuários compartilham opiniões sobre diferentes aspectos da vida cotidiana. Portanto, *microbloggings* originados do Twitter e outras redes sociais são uma rica fonte de dados para análise de sentimento. Pak e Paroubek (2010) constroem de maneira automática um corpus, que é usado para gerar um classificador de sentimento, capaz de determinar sentimentos positivos, negativos e neutros para textos de *microbloggings* escritos em inglês originados no Twitter. Dois classificadores são

construídos, um com base em Naïve Bayes e outro com base em MVS. O classificador Naïve Bayes mostra melhores resultados. O grande diferencial do trabalho de Pak e Paroubek (2010) é que o corpus usado para validar os classificadores é construído de forma automática e não com anotações manuais como feito em outros trabalhos aqui descritos.

Assim como no trabalho de Pak e Paroubek (2010), também construiremos um classificador de sentimentos, porém usando classificadores semânticos.

3.2 Classificação Semântica

Freitas e Vieira (2015) apresentam um classificador semântico de polaridade (positivo ou negativo) para a classificação de comentários escritos em português retirados do TripAdvisor, um portal *online* de opiniões sobre viagens e acomodações. Freitas e Vieira (2015) comparam diferentes recursos léxicos em português e ainda o impacto da aplicação de diferentes regras linguísticas no processo de atribuição dos sentimentos. O conjunto de dados usado por Freitas e Vieira (2015) contém 194 comentários escritos em português do Brasil e publicados entre março de 2010 e maio de 2014. A atribuição dos sentimentos para aos comentários foi feita de forma manual, conduzida por dois anotadores, um linguista e um cientista da computação, ambos falantes nativos do português.

No nosso trabalho o classificador semântico será implementado usando o SentiWordNet traduzido. Os testes serão feitos e comparados com o resultados obtidos no trabalho de Freitas e Vieira (2015), o mesmo conjunto de dados também será utilizado.

Cavalcanti et al. (2012) apresentam um classificador semântico para análise de citações de artigos científicos. Eles argumentam que o impacto de uma publicação é muitas vezes medido pelo número de citações que recebeu, sendo considerado atualmente um fator determinante para a relevância do trabalho publicado. No entanto, um índice alto de citações não significa necessariamente que uma publicação teve uma avaliação ou retorno positivo dos autores que a citaram, pois uma citação pode representar uma crítica negativa. Cavalcanti et al. (2012), a fim de superar essa limitação, utilizam análise de sentimento para classificar citações como positivas ou negativas e, assim, propor uma melhor ordenação das publicações, considerando o impacto positivo real da citação no artigo e não apenas o número de citação. Esse classificador foi construído através da extração de adjetivos, advérbios, verbos e substantivos das citações com o auxílio de uma ferramenta de *Pos-Tagger*¹. Em seguida, o recurso léxico SentiWordNet

¹ Pos-Tagger são sistemas que leem textos e atribuem classes morfológicas para as palavras.

é utilizado para identificar o grau de positividade e negatividade para cada termo extraído da citação.

Assim como no trabalho de Cavalcanti et al. (2012), também pretendemos usar o recurso léxico SentiWordNet para criar um classificador semântico, porém será feita a tradução desse recurso para o português, pois tal encontra-se disponível apenas na língua inglesa.

4 IDENTIFICANDO EMOÇÕES COM O SENTIWORDNET

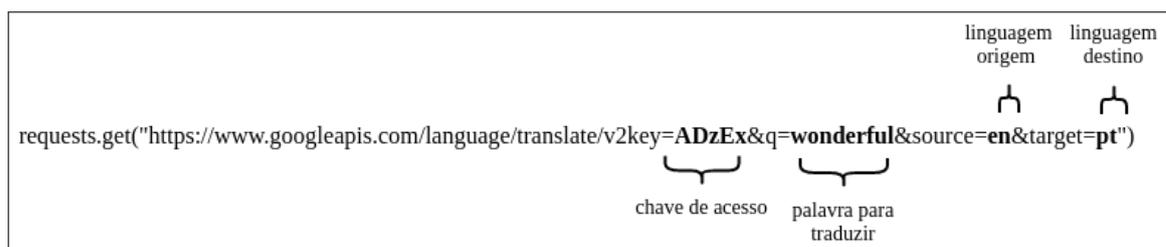
Nesse Capítulo apresenta-se como o SentiWordNet foi traduzido, os classificadores semânticos implementados no trabalho e as ferramentas utilizadas na implementação.

4.1 Tradução automática do SentiWordNet

Para utilizar o SentiWordNet em classificações de textos em português, foi preciso traduzir esse recurso de forma automática. No processo de tradução foi utilizado a API (*Application Programming Interface*) do Google Tradutor¹, uma plataforma que dá suporte à tradução automática de textos entre vários pares de linguagens diferentes. A API permite que sites e programas interajam com o Google Tradutor. Essa plataforma está disponível como um serviço pago, porém o Google disponibiliza um período de avaliação que fornece 300 dólares em crédito para serem usados em até 60 dias, a versão de avaliação foi suficiente para realizar a tradução.

Construímos um *script* para realizar a tradução usando a linguagem Python. O *script* realiza para cada palavra presente no SentiWordNet uma requisição para o serviço de tradução do Google, então esse serviço responde a requisição com um Json² que contém a palavra traduzida. A Figura 3 mostra os detalhes da requisição.

Figura 3 – Exemplo de requisição HTTP para Google API em python



Fonte: Produzido pelo autor

A requisição é feita via HTTP para o serviço de tradução do Google, na requisição é enviada uma chave de acesso para identificação do cliente, a palavra para a tradução, a linguagem de origem da palavra e a linguagem de destino da tradução. O serviço de tradução responde a requisição com uma palavra traduzida formatada usando Json. A Figura 4 mostra um exemplo do Json resposta da requisição.

¹ <http://cloud.google.com/translate/docs/>

² <http://www.json.org/>

Figura 4 – Exemplo de Json resposta da requisição

```

1  {
2    "data": {
3      "translations": [
4        {
5          "translatedText": "maravilhoso"
6        }
7      ]
8    }
9  }

```

Fonte: Produzido pelo autor

O SentiWordNet conta com mais de 117.000 mil *synsets*, cada *synsets* tem em média três palavras. Para concluir a tradução de todo o recurso léxico, o script de tradução foi executado em uma máquina virtual hospedada nos servidores da Amazon³. Para o melhor gerenciamento do recurso léxico, o SentiWordNet traduzido foi armazenado em um banco de dados Postgres⁴. A Figura 5 mostra um exemplo de *synsets* no SentiWordNet traduzido.

Figura 5 – Exemplo de *synset* SentiWordNet traduzido

```

1  Pos: 0
2  Neg: 0.625
3  {feio, horripilante, horrível, medonho, atroz}
4

```

Fonte: Produzido pelo autor

4.2 Classificação gramatical e simplificação do texto

Um passo importante para a classificação de polaridade em textos é a categorização de todas as palavras segundo sua classe gramatical. Essa etapa ajuda na aplicação das regras linguísticas usadas na classificação. Ferramentas de anotação morfossintática do tipo *pos-tagger* fazem essa classificação com base no papel que a palavra desempenha no contexto que é usada. Esse tipo de ferramenta faz o uso de categorias gramaticais básicas como: substantivo, verbo, adjetivo, advérbio e etc, para realizar a categorização.

Foi usado nesse trabalho o Tree-Tagger⁵, um tipo de etiquetador desenvolvido

³ <http://aws.amazon.com/pt/>

⁴ <http://www.postgresql.org/>

⁵ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

por Helmut Schmid no Instituto de Linguística Computacional da Universidade de Stuttgart (SCHMID, 1994). O Tree-Tagger é capaz de etiquetar textos em português atribuindo a cada palavra sua classe gramatical correspondente. O Tree-Tagger não marca só palavras mas também símbolos existentes no texto e ainda é capaz de tornar a sentença o mais simples possível, levando os verbos a sua forma infinitiva, retirando as flexões dos adjetivos e etc. O Tree-Tagger usa árvore de decisão binária para realizar a classificação gramatical. A árvore é construída de forma recursiva a partir de um conjunto de treino (SCHMID, 1994).

A Tabela 1 mostra a sentença “Esse hotel é bonito e calmo” etiquetada com o Tree-Tagger, onde DD indica um determinante demonstrativo, NCMS indica um substantivo comum masculino simples, VMI indica um verbo masculino indicativo, AQ0 indica um adjetivo de qualificação, C indica uma conjunção e Fp indica a presença de um ponto final.

Tabela 1 – Aplicação do Tree-Tagger

Termo	Esse	hotel	é	bonito	e	calmo	.
Lema	esse	hotel	ser	bonito	e	calmo	.
Etiqueta	DD	NCMS	VMI	AQ0	C	AQ0	Fp

Fonte: Produzido pelo autor

A Tabela 2 mostra as principais etiquetas de marcação do Tree-Tagger e suas respectivas classes gramaticais.

Tabela 2 – Etiquetas do Tree-Tagger

Etiqueta do Tree-Tagger	Uso gramatical
AQ0	Adjetivo de qualificação
AO0	Advérbio ordinal
RG	Advérbio geral
RN	Advérbio negativo
DD	Determinante demonstrativo
DP	Determinante possessivo
NCM	Substantivo comum masculino
NCF	Substantivo comum feminino
VM	Verbo principal
VA	Verbo auxiliar
P	Pronomes
C	Conjunções

Fonte: Produzido pelo autor

4.3 Classificadores de polaridade

Nesse seção são apresentados os métodos de classificação semântica implementados no trabalho. A polaridade dos textos é atribuída usando os valores presentes no SentiWordNet traduzido com os métodos de classificação aqui apresentados.

4.3.1 Soma das pontuações dos termos

O método de **soma das pontuações dos termos** (*term score summation*) foi proposto em (HAMOUDA; ROHAIM, 2011). Esse método usa os valores de todas as palavras presente no texto para realizar a classificação. Nessa técnica é feito o somatório das notas de positividade e negatividade de todos os termos encontrados no comentário. Então o sentimento do comentário é determinado com base no maior valor dos somatórios.

Algoritmo 1: Soma das pontuações dos termos

Entrada: *comentario*

Saída: classificação (*POSITIVO* ou *NEGATIVO*)

```

1 início
2   somaPos = 0, somaNeg = 0
3   para cada palavra ∈ comentario faça
4     somaPos += buscarValorPosNoRecurso(palavra)
5     somaNeg += buscarValorNegNoRecurso(palavra)
6   fim
7   se somaPos >= somaNeg então
8     retorne POSITIVO
9   senão
10    retorne NEGATIVO
11  fim
12 fim
```

O Algoritmo 1 descreve o funcionamento da técnica de soma das pontuações dos termos implementado nesse trabalho. Na linha 2, as variáveis que armazenam os valores dos somatórios são iniciadas com zero. Nas linhas 4 e 5, são adicionados aos somatórios os valores de positividade e negatividade de cada palavra presente no texto que tem correspondência no SentiWordNet traduzido. A função *buscarValorPosNoRecurso* retorna o valor de positividade

presente no SentiWordNet traduzido relacionado a palavra consultada, e a função *buscarValorNegNoRecurso* retorna o valor de negatividade presente no SentiWordNet traduzido relacionado a palavra consultada. Por fim, na linha 7 os valores dos dois somatórios são comparados, se valor do somatório dos valores positivos for maior ou igual do que o valor do somatório dos valores negativos, a classificação retorna “positivo”, e “negativo” caso contrário.

4.3.2 Média das sentenças e média do comentário

O método de **média das sentenças e média do comentário** (*average on sentence and average on review*) também foi proposto em (HAMOUDA; ROHAIM, 2011). Esse método é semelhante a soma das pontuações dos termos, porém ele usa a média dos valores das sentenças para calcular a média dos valores de todo o texto. Para cada sentença presente no comentário, uma média das notas de positivo e negativo é calculada com base nas notas de todas as palavras encontradas na sentença. Então, calcula-se uma média geral de todo o comentário, usando as médias de todas as sentenças. Se o valor de positividade da média geral for maior que valor de negatividade da média geral, o texto é classificado como “positivo”, e “negativo” caso contrário.

Algoritmo 2: Média das sentenças e média do comentário

Entrada: *comentario*
Saída: classificação (*POSITIVO* ou *NEGATIVO*)

```

1 início
2   somaPosComentario = 0, somaNegComentario = 0
3   para cada sentenca ∈ comentario faça
4     somaPosSentenca = 0, somaNegSentenca = 0
5     para cada palavra ∈ sentenca faça
6       somaPosSentenca += buscarValorPosNoRecurso(palavra)
7       somaNegSentenca += buscarValorNegNoRecurso(palavra)
8     fim
9     mediaPosSentenca = somaPosSentenca / numeroDePalavrasEm(sentenca)
10    mediaNegSentenca = somaNegSentenca / numeroDePalavrasEm(sentenca)
11    somaPosComentario += mediaPosSentenca
12    somaNegComentario += mediaNegSentenca
13  fim
14  mediaPosComentario = somaPosComentario / numSentencasEm(comentario)
15  mediaNegComentario = somaNegComentario / numSentencasEm(comentario)
16  se mediaPosComentario ≥ mediaNegComentario então
17    retorne POSITIVO
18  senão
19    retorne NEGATIVO
20  fim
21 fim

```

O Algoritmo 2 descreve o funcionamento da técnica de média das sentenças e média do comentário implementado nesse trabalho. Na linha 2 do algoritmo as variáveis que armazenam os valores dos somatórios do comentário são iniciadas com zero. Então para cada sentença presente no comentário é preciso calcular suas médias de valores positivos e negativos. Na linha 4 as variáveis que armazenam os somatórios das sentenças são iniciados com zero e nas linhas 6 e 7 os valores dessas variáveis são incrementados de acordo com os valores do SentiWordNet traduzido para cada palavra presente na sentença. Nas linhas 9 e 10 do algoritmo os valores dos somatórios das sentenças são usados para calcular as médias dos valores, dividindo os

somatórios pela quantidade de palavras na sentença. Nas linhas 11 e 12 os valores das médias são incrementados aos somatórios do comentário. E, finalmente, nas linhas 14 e 15 é calculado o valor de média para todo o comentário usando os somatórios de médias das sentenças e o número de sentenças presente no comentário. Então na linha 16, as médias de positivo e negativo são comparadas e a classificação determinada.

4.3.3 Método base

O **método base** (*baseline*) foi proposto em (FREITAS; VIEIRA, 2015). Para identificar a polaridade nesse método, o primeiro passo é identificar os substantivos presentes no texto, para cada um dos substantivos notas de positividade e negatividade são atribuídas. Essa atribuição é feita com a soma das notas dos adjetivos que estão três posições antes e três posições depois do substantivo. Para atribuir uma polaridade ao texto inteiro, é preciso somar todas as notas dos substantivos: se o valor de positividade for maior que valor de negatividade o texto é classificado como “positivo”, e “negativo” caso contrário.

Algoritmo 3: Método base

Entrada: *comentario***Saída:** classificação (*POSITIVO* ou *NEGATIVO*)

```

1 início
2   somaPosComentario = 0, somaNegComentario = 0
3   para cada sentenca ∈ comentario faça
4     para cada palava ∈ sentenca faça
5       somaPosSubstantivo = 0, somaNegSubstantivo = 0
6       se classificaGramatical(palavra) = SUBSTANTIVO então
7         para cada palava W até três posições do substantivo faça
8           se classificaGramatical(W) = ADJETIVO então
9             somaPosSubstantivo += buscarValorPosNoRecurso(W)
10            somaNegSubstantivo += buscarValorNegNoRecurso(W)
11           fim
12         fim
13       fim
14       somaPosComentario += somaPosSubstantivo
15       somaNegComentario += somaNegSubstantivo
16     fim
17   fim
18   se somaPosComentario ≥ somaNegComentario então
19     retorne POSITIVO
20   senão
21     retorne NEGATIVO
22   fim
23 fim

```

O Algoritmo 3 descreve o funcionamento do método base implementado nesse trabalho. Na linha 2 do algoritmo, as variáveis que armazenam os valores dos somatórios do comentário são iniciadas com zero. Então em cada sentença no comentário, procura-se os substantivos como feito na linha 6. Para atribuir valor ao substantivo procura-se adjetivos três posições antes e três posições depois, então os valores dos adjetivos são resgatados no SentiWordNet traduzido e incrementados aos somatórios dos substantivos como feito na linhas

9 e 10. Na linhas 14 e 15, são incrementados aos somatórios do comentário os valores dos somatórios dos substantivos. E na linha 18 os somatórios dos comentários são comparados e a classificação determinada.

4.3.4 Posição dos adjetivos

O método de classificação **posição dos adjetivos** (*adjectives position*) também foi proposto em (FREITAS; VIEIRA, 2015). Nesse método os substantivos são identificados e, para cada um deles, notas de positividade e negatividade são atribuídas. A atribuição é feita com a procura do adjetivo na posição imediatamente antes do substantivo. Se não é encontrado, então procura-se depois do substantivo. Esse processo é repetido até um outro substantivo ser encontrado ou até o um fim da sentença. A nota final atribuída ao substantivo é a mesma nota do primeiro adjetivo encontrado. Para atribuir uma polaridade ao texto inteiro, é preciso somar todas notas dos substantivos: se o valor de positividade for maior que o valor de negatividade o texto é classificado como “positivo”, e “negativo” caso contrário.

Algoritmo 4: Posição dos adjetivos

Entrada: *comentario*
Saída: classificação (*POSITIVO* ou *NEGATIVO*)

```

1 início
2   somaPosComentario = 0, somaNegComentario = 0
3   para cada sentenca ∈ comentario faça
4     para cada palava ∈ sentenca faça
5       somaPosSubstantivo = 0, somaNegSubstantivo = 0
6       se classificaGramatical(palavra) = SUBSTANTIVO então
7         se existe adjetivo anterior a palavra então
8           adjetivo = buscaAdjetivoAnterior(palavra, sentenca)
9         senão
10          adjetivo = buscaPrimeiroAdjetivoPosterior(palavra, sentenca)
11        fim
12        somaPosSubstantivo += buscaValorPosNoRecurso(adjetivo)
13        somaNegSubstantivo += buscaValorNegNoRecurso(adjetivo)
14      fim
15      somaPosComentario += somaPosSubstantivo
16      somaNegComentario += somaNegSubstantivo
17    fim
18  fim
19  se somaPosComentario ≥ somaNegComentario então
20    retorne POSITIVO
21  senão
22    retorne NEGATIVO
23  fim
24 fim

```

O Algoritmo 4 mostra o funcionamento do método de posição dos adjetivos implementado nesse trabalho. Na linha 2 do algoritmo, as variáveis que armazenam os valores dos somatórios do comentário são iniciadas com zero. Então em cada sentença no comentário, procura-se os substantivos como feito na linha 6. Para atribuir valor ao substantivo procura-se o primeiro adjetivo imediatamente antes como feito na linha 8 com a função *buscaAdjetivoAnterior*.

Se o adjetivo não existe, procura-se o primeiro adjetivo posterior, a função *buscaPrimeiroAdjetivoPosterior* usada na linha 10 retorna, se houver, o primeiro adjetivo posterior ao substantivo, essa busca é feita até o fim da sentença ou até encontrar outro substantivo. Nas linhas 16 e 17, os valores dos somatórios são incrementados com os valores atribuídos os substantivos. Na linha 19, os valores dos somatórios do comentários são comparados e a classificação determinada.

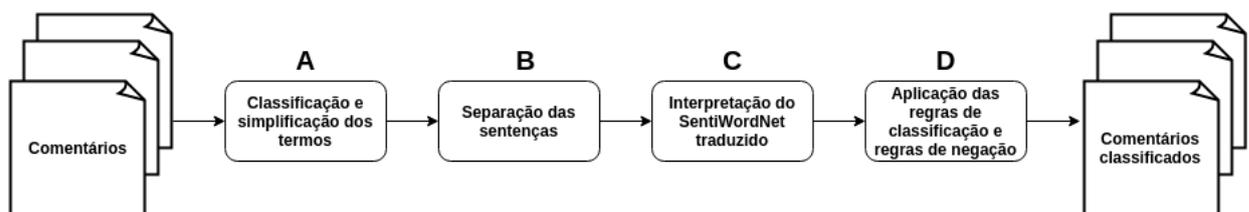
4.4 Implementação

A identificação de polaridade nesse trabalho é feita combinando o pré-processamento dos comentários com a classificação semântica usando o SentiWordNet traduzido.

A implementação dos classificadores foi feita usando a linguagem Python⁶. A escolha da linguagem de programação se dá pela sintaxe transparente, boas funcionalidades para o tratamento de seqüências de textos (*strings*) e suporte nativo para o *kit* de ferramentas *Natural Language Toolkit* (NLTK). O NLTK⁷ é conjunto de ferramentas *open source* desenvolvido em Python que estabelece uma infraestrutura para criação de programas de processamento de linguagem natural (BIRD; KLEIN; LOPER, 2009).

A Figura 6 descreve os passos para construção dos classificadores e sua seqüência de execução. Os passos na Figura 6 mostram a interação entre as ferramentas apresentadas e os algoritmos descritos na seção anterior.

Figura 6 – Classificação de polaridade



Fonte: Produzido pelo autor

Os dois primeiros passos descrevem o processamento linguístico dos comentários. No passo A a classificação gramatical e simplificação dos termos é feita usando a ferramenta Tree-Tagger. O texto é aplicado na ferramenta de marcação morfossintática e então as palavras presentes no texto são classificadas segundo sua classe gramatical, os verbos são levados ao

⁶ <http://www.python.org/>

⁷ <http://www.nltk.org/>

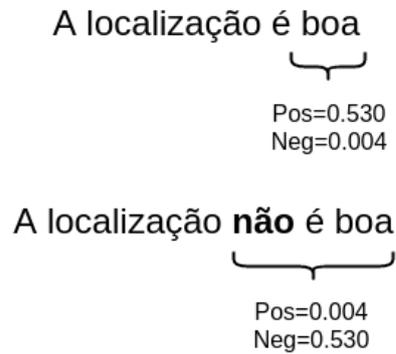
infinitivo e substantivos a sua forma contraída.

No passo **B** é realizada a separação das sentenças. Essa tarefa é um problema processamento de linguagem natural conhecido como *sentence boundary disambiguation (SBD)*. O problema consiste em saber onde as sentenças de um texto começam e terminam. Porém, a separação das sentenças não é uma tarefa trivial, os sinais de pontuação dos textos muitas vezes são ambíguos. Um ponto final, por exemplo, pode indicar o fim de uma sentença, uma abreviação, um número decimal ou um endereço de e-mail (REYNAR; RATNAPARKHI, 1997). Para realizar a separação de sentenças nesse trabalho, usamos a função *sentTokenize* disponível no NTLK.

Os últimos passos descrevem o funcionamento da classificação. No passo **C** é realizada a interpretação do SentiWordNet traduzido. Nessa etapa notas de positividade e negatividade são atribuídas às palavras. O SentiWordNet traduzido foi armazenado em um banco de dados Postgres. Quando as notas de uma palavra são solicitadas, realiza-se uma busca no banco de dados por todos os *synset* que contém a palavra, então a consulta devolve a média dos valores positivo e negativo.

No passo **D** a classificação é finalmente realizada. Nesse passo aplica-se as regras de classificação descritas no capítulo anterior. Cada classificador semântico tem seu próprio procedimento para usar os valores do SentiWordNet traduzido e classificar os textos como positivo ou negativo. Nessa etapa, as regras de negação são levadas em consideração no método base e no método da soma dos adjetivos. Valores de positividade e negatividade dos termos são invertidos se palavras como “não”, “nunca”, “ninguém”, “nada” e etc aparecerem no texto. Procura-se a partícula de negação uma posição antes do verbo que precede adjetivo, de acordo com a regra de negação I. A Figura 7 mostra um exemplo da inversão dos valores de polaridade quando uma negação é encontrada antes do verbo.

Figura 7 – Exemplo negação



Fonte: Produzido pelo autor

4.5 Exemplo de classificação

Como exemplo mostraremos a classificação do seguinte comentário usando a regra de soma das pontuações dos termos:

“Hotel excelente, funcionários super simpáticos e atenciosos.”

O passo de classificação e simplificação dos termos é feito usando o Tree-Tagger, os resultados desse processo são mostrados Tabela 3.

Tabela 3 – Uso do Tree-Tagger no exemplo

Termo	Hotel	excelente,	,	funcionários	super	simpáticos	e	atenciosos	.
Elemento	hotel	excelente	,	funcionário	super	simpático	e	atencioso	.
Classificação	NCMS	AQ0	Fc	NCMP	AQ0	AQ0	CC	AQ0	Fp

Fonte: Produzido pelo autor

Após a marcação do Tree-Tagger os comentários são separados por sentenças, no caso do exemplo há apenas uma frase. Para cada elemento presente na sentença é feita uma consulta no SentiWordNet traduzido para atribuir valores de positividade e negatividade aos elementos. Os resultados dessa atribuição são mostrados Tabela 4.

Tabela 4 – Exemplo da interpretação do SentiWordNet traduzido

Elemento	hotel	excelente	,	funcionário	super	simpático	e	atencioso	.
Valor Positivo	0.000	1.000	-	0.000	0.380	0.437	-	0.416	-
Valor Negativo	0.000	0.000	-	0.000	0.100	0.312	-	0.083	-

Fonte: Produzido pelo autor

Então, para atribuir a polaridade do comentário usando a regra de soma das pontuações dos termos, é preciso somar os valor positivos e negativos e verificar qual a maior soma. O somatório dos valores positivos vale 2,233 e o somatório dos valores negativos vale 0,4955. Como o valor do somatório positivo é maior, o comentário é classificado com positivo.

5 RESULTADOS

Nesse capítulo são apresentados os métodos de avaliações dos classificadores e os resultados obtidos com avaliação.

5.1 Métodos de avaliação

Para que seja possível realizar uma comparação, usamos as mesmas métricas de avaliação apresentadas em Freitas e Vieira (2015): precisão, cobertura e *f-measure*.

Comentários positivos classificados corretamente são considerados verdadeiros positivos (TP). Comentários positivos classificados como negativos são considerados falsos negativos (FN). Comentários negativos classificados corretamente são considerados verdadeiros negativos (TN). Comentários negativos classificados como positivo são considerados falsos positivos (FP). Com essas medidas calculamos precisão, cobertura e *f-measure*.

O cálculo de precisão descreve a razão entre os comentários classificados corretamente e o total de comentários.

$$PrecisaoPositiva = \frac{TP}{TP + FP}$$

$$PrecisaoNegativa = \frac{TN}{TN + FN}$$

O cálculo de cobertura consiste na razão entre os comentários corretamente classificados e total de comentários de deveriam ser corretamente classificados.

$$CoberturaPositiva = \frac{TP}{TP + FN}$$

$$CoberturaNegativa = \frac{TN}{TN + FP}$$

F-measure é a média harmônica de precisão e cobertura.

$$F - measurePositivo = \frac{2 * PrecisaoPositiva * CoberturaPositiva}{PrecisaoPositiva * CoberturaPositiva}$$

$$F - measureNegativo = \frac{2 * PrecisaoNegativa * CoberturaNegativa}{PrecisaoNegativa * CoberturaNegativa}$$

Essa medida mostra o quão preciso e robusto um classificador é. *F-measure* é utilizada como métrica de avaliação quando tanto precisão e cobertura são importantes para a medida. Por

exemplo, um classificador com um alto valor de precisão mas com uma baixa cobertura, pode ser considerado muito preciso, porém esse classificador está errando um número significativo de instancias difíceis de classificar. Então esse classificador pode não ser muito útil. Por isso a medida de *F-measure* é usada na avaliação.

5.2 Resultados

O *dataset* usado nos experimentos contém 194 comentários rotulados, (106 da classe positiva e 88 da classe negativa) e foi construído em (FREITAS; VIEIRA, 2015). Os comentários foram retirados do Tripadvisor e rotulados por dois anotadores, ambos falantes nativos do Português: um linguista e outro cientista da computação. Com os rótulos dos comentários é possível saber a polaridade do comentário e quais características (quartos, localização, atendimento, limpeza e custo-benefício) o comentário está relacionado.

Na apresentação dos resultados, é comparado a eficiência da utilização do SentiWordNet traduzido com o SentiLex. O recurso léxico SentiLex foi escolhido para realizar as comparações pois, esse recurso mostrou o melhor desempenho em uma comparação feita entre recursos léxicos em português no trabalho de Freitas e Vieira (2015). As configurações seguintes mostram como os classificadores foram organizados para o experimento:

- Configuração #1: usa a técnica do método base e o SentiWordNet traduzido.
- Configuração #2: usa a técnica do método base e o SentiLex
- Configuração #3: usa o método de posição dos adjetivos e o SentiWordNet traduzido.
- Configuração #4: usa o método de posição dos adjetivos e o SentiLex
- Configuração #5: usa a técnica de soma das pontuações dos termos e o SentiWordNet traduzido.
- Configuração #6: usa a técnica de soma das pontuações dos termos e o SentiLex.
- Configuração #7: usa a técnica de média das sentenças e média do comentário e o SentiWordNet traduzido.
- Configuração #8: usa a técnica de média das sentenças e média do comentário e o SentiLex.

A Tabela 5 mostra o valor de *f-measure* obtido com avaliação das configurações #1, #2, #3 e #4. A avaliação divide os comentários que falam de quartos, localização, atendimento, limpeza e custo-benefício, podendo haver interseção de comentários nesses conjuntos. Foi calculado o valor de *f-measure* das classes positiva e negativa para cada conjunto de comentários que mencionam uma característica. Com o uso do SentiWordNet traduzido foi possível obter

resultados próximos os resultados obtidos com aplicação do SentiLex, um recurso léxico em português.

Tabela 5 – Avaliação das configurações #1, #2, #3 e #4

características	# comen.	#1		#2		#3		#4	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Quarto	128	0.74	0.59	0.75	0.59	0.73	0.55	0.74	0.54
Localização	73	0.83	0.55	0.81	0.55	0.82	0.51	0.83	0.52
Atendimento	47	0.88	0.47	0.92	0.66	0.89	0.55	0.90	0.63
Limpeza	9	0.90	0.85	0.90	0.85	0.90	0.85	0.90	0.85
Custo-benefício	4	0.50	0.50	0.80	0.66	0.50	0.50	0.80	0.66
Média		0.770	0.592	0.836	0.662	0.768	0.592	0.834	0.640
		0.681		0.749		0.680		0.737	

Fonte: Produzido pelo autor

A Tabela 6 mostra o valor de *f-measure* obtido com avaliação das configurações #5, #6, #7 e #8. É possível notar algumas melhorias com o uso do SentiWordNet traduzido. Na configuração #5, que usa o recurso léxico traduzido, foi obtido 0.66 de *f-measure* da classe negativa na avaliação dos comentários que faltam de quarto, e na configuração #6 foi obtido 0.60 de *f-measure* para os mesmo comentários.

Tabela 6 – Avaliação das configurações #5, #6, #7 e #8

características	# comen.	#5		#6		#7		#8	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Quarto	128	0.78	0.66	0.77	0.60	0.75	0.58	0.74	0.57
Localização	73	0.82	0.59	0.80	0.46	0.80	0.48	0.80	0.52
Atendimento	47	0.88	0.52	0.92	0.66	0.86	0.44	0.92	0.66
Limpeza	9	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
Custo-benefício	4	0.50	0.50	0.40	0.00	0.40	0.00	0.40	0.00
Média		0.728	0.586	0.710	0.476	0.694	0.432	0.714	0.482
		0.657		0.593		0.563		0.598	

Fonte: Produzido pelo autor

A Tabela 7 mostra os valores de *f-measure* obtidos com avaliação das configurações #1, #2, #3 e #4. Nessa avaliação os comentários não foram divididos por características, portanto a tabela apresenta uma avaliação geral de todo o corpus. As configurações #1 e #3, que usam o SentiWordNet traduzido mostram resultados melhores ou iguais as configurações que usam o SentiLex.

Tabela 7 – Avaliação geral das configurações #1, #2, #3 e #4

avaliação	# comen.	#1		#2		#3		#4	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
todos comentários	194	0.78	0.66	0.76	0.61	0.78	0.63	0.78	0.62

Fonte: Produzido pelo autor

Como na tabela anterior, a Tabela 8 mostra os valores de *f-measure* obtidos com avaliação geral das configurações #5, #6, #7 e #8. As configurações #5 e #7 que usam o SentiWordNet traduzido apresentaram melhores avaliações em todos os casos.

Tabela 8 – Avaliação geral das configurações #5, #6, #7 e #8

avaliação	# comen.	#5		#6		#7		#8	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
todos comentários	194	0.81	0.70	0.78	0.62	0.80	0.66	0.77	0.63

Fonte: Produzido pelo autor

A Tabela 9 mostra os resultados das configurações #1 e #3 em comparação aos resultados obtidos e apresentados em (FREITAS; VIEIRA, 2015) usando o mesmo corpus de comentários. As avaliações #1 (FREITAS; VIEIRA, 2015) e #3 (FREITAS; VIEIRA, 2015) foram obtidas em (FREITAS; VIEIRA, 2015) com os mesmos parâmetros das configurações #1 e #3, respectivamente.

Tabela 9 – Comparação das avaliações

características	# comen.	#1		#1 (Freitas; Vieira, 2015)		#3		#3 (Freitas; Vieira, 2015)	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Quarto	128	0.74	0.59	0.44	0.46	0.73	0.55	0.46	0.40
Localização	73	0.83	0.55	0.31	0.50	0.82	0.51	0.75	0.37
Atendimento	47	0.88	0.47	0.36	0.76	0.89	0.55	0.84	0.36
Limpeza	9	0.90	0.85	0.67	0.80	0.90	0.85	0.67	0.67
Custo-benefício	4	0.50	0.50	0.00	0.00	0.50	0.50	0.00	0.00
Média		0.770	0.592	0.356	0.504	0.768	0.592	0.544	0.360
		0.681		0.482		0.680		0.456	

Fonte: Produzido pelo autor

As configurações #1 e #3 avaliadas nesse trabalho mostram melhores resultados em quase todas as avaliações, devido o uso do SentiWordNet traduzido e algumas diferenças na implementação do método base e no método de posição dos adjetivos. A identificação de substantivos é um exemplo dessas diferenças, as configurações de Freitas e Vieira (2015) fazem o uso de ferramentas como o TreeTagger e *Domain Ontology* nessa etapa, enquanto as

configurações #1 e #3 usam apenas o TreeTagger. As identificações das regras de negação também são contadas como diferenças, nesse trabalho usamos as regras de negação do Tipo I, enquanto (FREITAS, 2015) usam os três tipos de negação da língua portuguesa.

É possível notar uma grande melhora na avaliação dos comentários que falaram de localização na classe positiva, enquanto a configuração #1 obteve 0.83 de *f-measure* a configuração #1 (FREITAS; VIEIRA, 2015) obteve 0.31. Mas também é possível encontrar algumas pioras, por exemplo, a configuração #1 na avaliação dos comentário que falaram de atendimento na classe negativa obteve 0.47, enquanto a configuração #1 (FREITAS; VIEIRA, 2015) obteve 0.76 na mesma avaliação.

6 DISCUSSÃO

Nos experimentos realizados nesse trabalho, os classificadores que usam o SentiWordNet traduzido apresentam resultados próximos e alguns em casos melhores que os classificadores que usam um recurso léxico em português. O SentiWordNet traduzido apresenta resultados satisfatórios mesmo com os erros de tradução cometidos pela tradução automática.

Na Tabela 5 do Capítulo 5, podemos observar pequenas melhorias nas avaliações dos classificadores que usam o SentiWordNet traduzido. Na configuração #1 que usa o recurso traduzido, foi obtido 0.83 de *f-measure* da classe positiva na avaliação dos comentários que falam de localização, e na configuração #2 foi obtido 0.82 na mesma avaliação. Em algumas avaliações, as configurações que usam o recurso léxico já em português foram melhores. A maior diferença na Tabela 5 aconteceu na avaliação dos comentários que falaram de atendimento, a configuração #1 obteve 0.47 de *f-measure* da classe negativa e a configuração #2 obteve 0.66 de *f-measure* na mesma classe.

Uma provável razão para essas diferenças são os erros de tradução que ocorrem na tradução automática do SentiWordNet. Por exemplo o *synset* S1={ *courageous, brave* } como valores Pos=0.375 e Neg=0.125 foi traduzido para S2={ corajoso, bravo }, porém em português do Brasil é mais comum usar a palavra bravo para expressar irritação. Com essa tradução a palavra bravo passa a assumir um valor de positividade erroneamente.

Na Tabela 6, os classificadores que usam o SentiWordNet traduzido mostraram uma maior vantagem sobre os que usam o SentiLex. Na avaliação dos comentários que falam de quartos, as configurações #5 e #7 têm resultados melhores que as configurações #6 e #8. As maiores diferenças acontecem na avaliação da classe negativa. Nos comentários que falaram de localização a configuração #5 obteve 0.59 de *f-measure* e a configuração #6 obteve 0.46. Na avaliação dos comentários que falam de atendimento os classificadores que usam o SentiLex saíram-se melhores.

Quando as avaliações foram feitas de forma geral, ou seja usando todo o corpus sem a separação por características (quarto, localização e etc) as configurações que usam o SentiWordNet traduzido tiveram uma melhor ou igual avaliação em todos os casos. Na Tabela 7 a configuração #1 obteve 0.78 de *f-measure* na classe positiva e 0.66 na classe negativa enquanto a configuração #2 obteve 0.76 na classe positiva e 0.61 na classe negativa. É possível notar uma diferença ainda mais significativa na Tabela 8, onde por exemplo, a configuração #5 obteve 0.70 de *f-measure* na classe negativa e a configuração #6 obteve apenas 0.62 na mesma classe.

Na Tabela 9, as técnicas método base e posição dos adjetivos implementados nesse trabalho são comparados as implementações em (FREITAS; VIEIRA, 2015). Os principais motivos das diferenças se dão pelo uso do SentiWordNet traduzido, diferenças na implementação, diferenças na aplicação das regras de negação (enquanto esse trabalho usa apenas regras de negação do tipo I, as implementações das comparadas usam os três tipos de negação da língua portuguesa).

7 CONSIDERAÇÕES FINAIS

Nesse trabalho apresentamos classificadores semânticos de polaridade (positivo e negativo) para análise de textos escritos em Português do Brasil. Os classificadores usaram o SentiWordNet para atribuir notas de positivo e negativos as palavras. O SentiWordNet foi traduzido de forma automática para o Português, mesmo com os erros de tradução, os classificadores que usaram o SentiWordNet traduzido obtiveram um bom desempenho. O SentiWordNet traduzido nesse trabalho está disponível em um repositório público no GitHub¹. É possível investir em modos de traduções automáticas mais elaboradas para aplicar em recursos léxicos disponíveis em outras línguas.

Na comparação da aplicação do SentiWordNet traduzido e do SentiLex, (recurso já em português) na maioria dos casos o SentiWordNet traduzido obteve resultados melhores ou iguais os do SentiLex. Nos casos em que foi pior a diferença das avaliações não foram discrepantes.

O SentiWordNet traduzido obteve uma maior vantagem sobre o SentiLex quando os classificadores precisavam consultar um maior número de palavras. Por exemplo, os classificadores das configurações #5 e #6 descritos na Seção 5.2 precisam realizar uma consulta no recurso léxico para cada palavra presente no texto. Já os classificadores das configurações #1 e #2 realizam consultas no recurso léxico apenas para os adjetivos. Isso se dá pelo fato do SentiWordNet traduzido ter mais palavras que o SentiLex.

Ainda foi realizado uma comparação com os resultados obtidos por Freitas e Vieira (2015) com os classificadores método base e posição dos adjetivos. Os classificadores implementados nesse trabalho, obtiveram em alguns casos melhores resultados. As diferenças se devem ao uso do SentiWordNet traduzido e diferenças na identificação dos substantivos e das regras de negação na implementação dos classificadores. Os resultados mostram assim o uso promissor do SentiWordNet traduzido em comparação ao recurso léxico SentiLex.

Para os trabalhos futuros será revista a tradução do SentiWordNet de forma que seja possível melhorá-la. Também será considerado realizar a tradução automática de outros recursos léxicos populares em análise de sentimentos. E ainda a expansão do corpus de comentários usado nesse trabalho.

¹ <<https://github.com/romulocosta100/SentiWordNetTraduzido.git>>

REFERÊNCIAS

- ALLEN, J. F. Natural language processing. John Wiley and Sons Ltd., 2003.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with Python**. [S.l.]: "O'Reilly Media, Inc.", 2009.
- BRASCHER, M. A ambigüidade na recuperação da informação. IASI, 2002.
- CAVALCANTI, D. C.; PRUDÊNCIO, R. B.; PRADHAN, S. S.; SHAH, J. Y.; PIETROBON, R. S. Análise de sentimento em citações científicas para definição de fatores de impacto positivo. In: **Proceedings of the IV International Workshop on Web and Text Intelligence (WTI)**. [S.l.: s.n.], 2012. p. 1–10.
- CHOWDHURY, G. G. Natural language processing. **Annual review of information science and technology**, Wiley Online Library, v. 37, n. 1, p. 51–89, 2003.
- COPPIN, B. Inteligência artificial/ben coppin; tradução e revisão técnica jorge duarte pires valério. **Rio de Janeiro: LTC**, 2010.
- DANG, Y.; ZHANG, Y.; CHEN, H. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. **Intelligent Systems, IEEE**, IEEE, v. 25, n. 4, p. 46–53, 2010.
- DOSCIATTI, M. M.; FERREIRA, E. C. L. P. C. Identificando emoções em textos em português do brasil usando máquina de vetores de suporte em solução multiclasse. **ENIAC-Encontro Nacional de Inteligência Artificial e Computacional. Fortaleza, Brasil**, 2013.
- DUARTE, E. S. **Sentiment analysis on twitter for the portuguese language**. Tese (Doutorado) — Faculdade de Ciências e Tecnologia, 2013.
- ESULI, A.; SEBASTIANI, F. Sentiwordnet: A publicly available lexical resource for opinion mining. In: CITESEER. **Proceedings of LREC**. [S.l.], 2006. v. 6, p. 417–422.
- FREITAS, L. A. de. **Feature-level sentiment analysis applied to brazilian portuguese reviews**. Tese (Doutorado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2015.
- FREITAS, L. d; VIEIRA, R. Exploring resources for sentiment analysis in portuguese language. In: IEEE. **2015 Brazilian Conference on Intelligent Systems (BRACIS)**. [S.l.], 2015. p. 152–156.
- HAMOUDA, A.; ROHAIM, M. Reviews classification using sentiwordnet lexicon. In: **World Congress on Computer Science and Information Technology**. [S.l.: s.n.], 2011.
- LIU, B. Sentiment analysis and subjectivity. **Handbook of natural language processing**, v. 2, p. 627–666, 2010.
- MAYNARD, D.; FUNK, A. Automatic detection of political opinions in tweets. In: SPRINGER. **The semantic web: ESWC 2011 workshops**. [S.l.], 2011. p. 88–99.
- MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. **Ain Shams Engineering Journal**, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.

MUNDIAL, B. Relatório sobre o desenvolvimento mundial 2016. **DIVIDENDOS DIGITAIS: visão geral**. Washington, DC, 2016.

NEETHU, M.; RAJASREE, R. Sentiment analysis in twitter using machine learning techniques. In: IEEE. **Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on**. [S.l.], 2013. p. 1–5.

PAK, A.; PAROUBEK, P. Twitter as a corpus for sentiment analysis and opinion mining. In: **LREc**. [S.l.: s.n.], 2010. v. 10, p. 1320–1326.

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.

REYNAR, J. C.; RATNAPARKHI, A. A maximum entropy approach to identifying sentence boundaries. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the fifth conference on Applied natural language processing**. [S.l.], 1997. p. 16–19.

SCHMID, H. Probabilistic part-of-speech tagging using decision trees. In: CITeseer. **Proceedings of the international conference on new methods in language processing**. [S.l.], 1994. v. 12, p. 44–49.

SCHWENTER, S. A. The pragmatics of negation in brazilian portuguese. **Lingua**, Elsevier, v. 115, n. 10, p. 1427–1456, 2005.

VAPNIK, V. **The nature of statistical learning theory**. [S.l.]: Springer Science & Business Media, 2013.

APÊNDICE A – AVALIAÇÕES

O Apêndice A mostra as avaliações de forma mais detalha. Configurações:

- Configuração #1: usa a técnica do método base e o SentiWordNet traduzido.
- Configuração #2: usa a técnica do método base e o SentiLex
- Configuração #3: usa o método de posição dos adjetivos e o SentiWordNet traduzido.
- Configuração #4: usa o método de posição dos adjetivos e o SentiLex
- Configuração #5: usa a técnica de soma das pontuações dos termos e o SentiWordNet traduzido.
- Configuração #6: usa a técnica de soma das pontuações dos termos e o SentiLex.
- Configuração #7: usa a técnica de média da sentenças e média do comentário e o SentiWordNet traduzido.
- Configuração #8: usa a técnica de média da sentenças e média do comentário e o SentiLex.

Quadro 1 – Legenda para as tabelas

Símbolo	Legenda
c	característica
Q	quarto
Lo	localização
A	atendimento
Li	limpeza
C-B	custo-benefício
T	todos comentários
P	precisão
C	cobertura
F	<i>f-measure</i>

Fonte: Elaborado pelo autor

Tabela 10 – Avaliação das configurações #1 e #2

c	#1						#2					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
Q	0.66	0.85	0.74	0.74	0.50	0.59	0.66	0.88	0.75	0.77	0.48	0.59
Lo	0.77	0.89	0.83	0.68	0.45	0.55	0.77	0.85	0.81	0.63	0.50	0.55
A	0.82	0.94	0.88	0.66	0.36	0.47	0.87	0.97	0.92	0.85	0.54	0.66
Li	0.83	1.00	0.90	1.00	0.75	0.85	0.83	1.00	0.90	1.00	0.75	0.85
C-B	0.50	0.50	0.50	0.50	0.50	0.50	0.66	1.00	0.80	1.00	0.50	0.66

Fonte: Produzido pelo autor

Tabela 11 – Avaliação das configurações #3 e #4

c	#3						#4					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
Q	0.64	0.85	0.73	0.72	0.44	0.55	0.64	0.88	0.74	0.75	0.43	0.54
L	0.75	0.89	0.82	0.66	0.41	0.51	0.76	0.91	0.83	0.71	0.41	0.52
A	0.85	0.94	0.89	0.71	0.45	0.55	0.87	0.94	0.90	0.75	0.54	0.63
L	0.83	1.00	0.90	1.00	0.75	0.85	0.83	1.00	0.90	1.00	0.75	0.85
C-B	0.50	0.50	0.50	0.50	0.50	0.50	0.66	1.00	0.80	1.00	0.50	0.66

Fonte: Produzido pelo autor

Tabela 12 – Avaliação das configurações #5 e #6

c	#5						#6					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
Q	0.70	0.88	0.78	0.80	0.56	0.66	0.67	0.92	0.77	0.84	0.46	0.60
Lo	0.79	0.85	0.82	0.65	0.54	0.59	0.74	0.87	0.80	0.60	0.37	0.46
A	0.84	0.91	0.88	0.62	0.45	0.52	0.87	0.97	0.92	0.85	0.54	0.66
Li	0.75	0.60	0.66	0.60	0.75	0.66	0.75	0.60	0.66	0.60	0.75	0.66
C-B	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50

Fonte: Produzido pelo autor

Tabela 13 – Avaliação das configurações #7 e #8

c	#7						#8					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
Q	0.65	0.88	0.75	0.77	0.46	0.58	0.65	0.86	0.74	0.75	0.46	0.57
Lo	0.75	0.85	0.80	0.58	0.41	0.48	0.76	0.85	0.80	0.61	0.45	0.52
A	0.82	0.91	0.86	0.57	0.36	0.44	0.87	0.97	0.92	0.85	0.54	0.66
Li	0.75	0.60	0.66	0.60	0.75	0.66	0.75	0.60	0.66	0.60	0.75	0.66
C-B	0.33	0.50	0.40	0.00	0.00	0.00	0.33	0.50	0.40	0.00	0.00	0.00

Fonte: Produzido pelo autor

Tabela 14 – Avaliação geral das configurações #1 e #2

c	#1						#2					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
T	0.70	0.87	0.78	0.79	0.56	0.66	0.68	0.86	0.76	0.76	0.51	0.61

Fonte: Produzido pelo autor

Tabela 15 – Avaliação geral das configurações #3 e #4

c	#3						#4					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
T	0.69	0.88	0.78	0.79	0.53	0.63	0.69	0.90	0.78	0.81	0.51	0.62

Fonte: Produzido pelo autor

Tabela 16 – Avaliação geral das configurações #5 e #6

c	#5						#6					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
T	0.73	0.92	0.81	0.86	0.59	0.70	0.68	0.91	0.78	0.83	0.50	0.62

Fonte: Produzido pelo autor

Tabela 17 – Avaliação geral das configurações #7 e #8

c	#7						#8					
	Pos			Neg			Pos			Neg		
	P	C	F	P	C	F	P	C	F	P	C	F
T	0.70	0.93	0.80	0.87	0.53	0.66	0.69	0.88	0.77	0.79	0.52	0.63

Fonte: Produzido pelo autor