UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

GABRIEL SEHN DELAVALD

Uma Análise de Dados das Reações à Crise Política Brasileira no Twitter

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação

Orientador: Prof. Dr. Leandro Krug Wives
Co-orientador: Ms. Vinícius Woloszyn

Porto Alegre
2018

AGRADECIMENTOS

Gostaria de agradecer acima de tudo à minha mãe, pelo dom da curiosidade e apreço pelo conhecimento que me deu desde sempre. Ao meu pai, de onde quer que ele esteja, pela dedicação e exemplo de como enfrentar todas dificuldades que a vida nos impõe.

Agradeço à minha esposa, companheira e melhor amiga, Luísa, pelos anos compartilhados e por me mostrar a cada dia que eu posso ser uma pessoa melhor.

Um agradecimento especial à toda minha família pelo apoio durante todos estes anos, ao grupo da Gurizada e ao time do Void F.C. pelo companheirismo demonstrado durante o curso.

Finalmente, agradeço ao Prof. Leandro Krug Wives e ao Vinícius Woloszyn pela paciência e orientação no desenvolvimento deste trabalho.

RESUMO

A análise de opiniões pessoais nas mídias sociais tornou-se ponto de interesse para diversos trabalhos, com um foco crescente em plataformas como o Twitter, onde a conversação agregada pode ser utilizada, por meio de mineração de dados, como fonte para análise de opiniões. Este trabalho utiliza a mineração de dados no Twitter por meio de uma busca por palavras-chave para construção de um *dataset* baseado na opinião dos usuários a respeito da crise política brasileira e reações à discussão sobre uma intervenção militar no país, sendo os dados coletados datados de janeiro de 2014 a Dezembro de 2017. O *dataset* construído é então utilizado por um algoritmo de análise de sentimentos, baseado no dicionário LIWC em Português Brasileiro, para gerar dados a respeito da opinião pública.

Palavras-chave: Processamento da Lingua Natural. Análise de Redes Sociais. Análise de Polaridade. Mineração de Textos.

A Data Analysis of the Reactions to the Brazilian Political Crisis on Twitter

ABSTRACT

The analysis of personal opinions on social media has become a source of interest to many types of researches, with a growing focus on platforms like Twitter, in which the aggregated chatter can be used, by data mining, as a source for opinion analysis. This paper creates a *dataset* based on users' opinions about the Brazilian political crisis and reactions to the military intervention discussion via data mining from Twitter, with the data being gathered by a keyword search matching the dates of January 2014 to December 2017. The completed *dataset* is then used by a sentiment analysis algorithm, based on the LIWC dictionary for Brazilian Portuguese, to create data about the public opinion on the subject.

Keywords: Natural Language Processing, Social Network Analysis, Polarity Analysis, Text Mining.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
DOM	Document Object Model
HTML	HyperText Markup Language
LIWC	Linguistic Inquiry and Word Count
MIT	Massachusetts Institute of Technology
NILC	Núcleo Interinstitucional de Linguística Computacional
SVM	Support Vector Machine
TWINT	Twitter Intelligence Tool
URL	Uniform Resource Locator

SUMÁRIO

1 INTRODUÇÃO

A análise de opiniões pessoais nas mídias sociais tornou-se ponto de interesse para diversos trabalhos, seja no âmbito político (TUMASJAN et al., 2010) considerando a opinião pública em relação aos candidatos, como no financeiro (BOLLEN; MAO; ZENG, 2011) em que as informações compartilhadas nas redes sociais tornam-se fator-chave na flutuação do mercado, ou na resposta da sociedade (TEKA; NAINI; NIEDERÉE, 2016) a acontecimentos que afetam suas vidas.

O crescimento de uso das mídias sociais como uma plataforma para conversação online tem sido documentado por diversos autores, já que essas redes possibilitam uma maneira para os usuários criarem e compartilharem suas opiniões e emoções para seus amigos, seguidores e público geral através da sua rede de contatos. Entre essas plataformas, o Twitter possui características únicas que o fazem se diferenciar dos demais. Por ter sido desenvolvido com o foco em *micro-blogging*, definido como uma nova forma de comunicação em que usuários descrevem interesses ou expressam atitudes em postagens curtas (JANSEN et al., 2009), a plataforma força os usuários a respeitarem um limite de caracteres para suas postagens, chamadas de *tweets*, inicialmente limitadas em 140 caracteres e atualmente em 280 (TWITTER, 2018).

Considerando a conversação agregada dessas redes como uma forma de sabedoria coletiva, como (ASUR; HUBERMAN, 2010), acredita-se ser possível utilizar essas informações como uma visão geral da opinião pública em relação aos assuntos observados, que no caso deste trabalho foca-se na crise política do Brasil, mais especificamente na discussão sobre a possibilidade de uma intervenção militar e no sentimento do público em relação a esses discursos.

Os pedidos de intervenção militar começaram a ser vistos na mídia em 2014, com a reeleição da presidente Dilma Rousseff em meio a uma crise política evidenciada por grandes esquemas de corrupção sendo descobertos e a uma crise econômica que levou ao recuo do Produto Interno Bruto e ao aumento do desemprego, porém o discurso autoritário tem tornado-se mais comum desde então e muitas pessoas têm compartilhado opiniões sobre o assunto nas redes sociais, como o Twitter e o Facebook (MüZELL, 2017; CASTRO, 2017).

Nesse contexto, o objetivo deste trabalho é gerar um *dataset* com dados extraídos do Twitter relacionados à crise política brasileira, utilizando a palavra-chave "Intervenção Militar"como principal foco da pesquisa e buscando todos *tweets* entre janeiro de 2014 e

dezembro de 2017, e realizar uma análise de polaridade nestes documentos. Baseando-se na hipótese geral de que se uma opinião pode ser categorizada como *positiva* ela está endossando a "Intervenção Militar", e se for categorizada como *negativa* ela a repudia, verificamos tendências da opinião pública e a inclinação geral do discurso encontrado na rede social.

As contribuições deste trabalho são a geração de um *dataset* de *tweets* extraídos do Twitter com opiniões sobre a crise política brasileira (especificamente, sobre a intervenção militar) e uma avaliação destes dados utilizando análise de polaridade com um dicionário em Português Brasileiro.

Este trabalho está organizado em 6 capítulos. No capítulo seguinte serão apresentados os fundamentos, conceitos e tecnologias utilizados na pesquisa e desenvolvimento deste trabalho, ao abordar métodos de extração de dados e análise de sentimentos. No Capítulo 3 é realizada uma análise dos trabalhos relacionados, discutindo as diferentes aplicações e técnicas utilizadas e comparando-as aos métodos utilizados neste trabalho. O Capítulo 4 apresenta a proposta do trabalho e a metodologia adotada para a coleta de dados e análise do *dataset*. No Capítulo 5 apresentamos os resultados da mineração de dados proposta para a construção do *dataset* e fazemos uma análise sobre os dados coletados. Finalmente, a conclusão do trabalho é apresentada no capítulo 6, onde são sugeridas melhorias em cima do projeto proposto e discutidos trabalhos que poderiam ser explorados a partir dos resultados encontrados.

2 FUNDAMENTOS, CONCEITOS E TECNOLOGIAS

Neste capítulo são abordados os fundamentos teóricos e os conceitos gerais utilizados na elaboração do trabalho, assim como as tecnologias utilizadas para o desenvolvimento do projeto.

2.1 Mineração de dados

Com o grande crescimento da Web nas últimas décadas, ela tornou-se uma fonte de informações enorme que se expande sobre diversos assuntos e interesses. Essas informações podem ser encontradas de qualquer forma, seja em textos desestruturados, páginas que possuem uma hierarquia, tabelas estruturadas e arquivos multimídia (imagens, áudios e vídeos) (LIU, 2007).

A mineração de dados, do inglês *Data Mining*, é definida como o processo de descobrimento de padrões úteis em grandes quantidades de dados como bancos de dados, textos, imagens, documentos da Web, etc (SRIVASTAVA et al., 2000). O termo *Web Mining* é utilizado para definir a mineração de dados especificamente na Web, cujo objetivo é descobrir informações úteis a partir de estruturas de links, conteúdo de páginas e uso de dados relacionados com essas páginas e seu acesso.

A mineração de estruturas da Web propõe-se a descobrir informações baseadas nos *links* das páginas, cujo funcionamento é uma peça-chave da tecnologia utilizada em motores de busca como Google, Bing e Yahoo. Em (CHAKRABARTI et al., 1999), os autores explicam que o funcionamento dos links na Web é baseado em "referências"(ou endosso, do inglês *endorsement*), assim, uma página que possui muitos *links* apontados para si demonstra uma influência quanto às páginas que apontam para ela. Esse tipo de mineração é utilizado para descobrir relações entre comunidades que compartilham interesses comuns (LIU, 2007).

A mineração de uso da Web, por outro lado, procura por padrões de acesso do usuário, como *cookies*, históricos de navegação e registros de *logs* de servidores e sistemas de *proxy* (KOSALA; BLOCKEEL, 2000). Esse tipo de informação é utilizado, principalmente, para personalização da navegação ou melhora no sistema para o usuário, porém também pode ser visto na Web como uma forma de aumentar o sucesso de propagandas mostradas para o usuário (SRIVASTAVA et al., 2000).

Finalmente, a mineração de conteúdo da Web, foco principal deste trabalho, busca

Figura 2.1: Utilização da biblioteca TWINT em Python

```
import twint

searchConfig = twint.Config()
searchConfig.Search = "Intervenção Militar"

twint.run.Search(searchConfig)
```

Fonte: Elaborado pelo autor.

extrair informações e conhecimento de conteúdos disponíveis em páginas Web (LIU, 2007). Essa categoria de mineração é utilizada para classificar e agregar automaticamente conteúdos de páginas Web, descobrir padrões para extração de dados úteis e minerar comentários e *reviews* de usuários para a análise de opiniões (KOSALA; BLOCKEEL, 2000; WOLOSZYN; SANTOS; WIVES, 2016).

2.1.1 Twitter Intelligence Tool

Para a extração de dados do Twitter foi utilizada a ferramenta *Twitter Intelligence Tool* (TWINT), que pode ser encontrada gratuitamente e possui código aberto no GitHub (https://github.com/twintproject/twint) com a licença MIT.

O TWINT utiliza-se da ferramenta de pesquisa avançada que é disponibilizada pelo Twitter e minera os dados da página de resultados retornada. Este método é uma melhora em relação ao uso da API de buscas do Twitter, pois essa limita os resultados aos últimos 3200 *tweets* e também a uma taxa de requisições de 180 chamadas a cada 15 minutos. Além disso, para a utilização da API do Twitter é necessário uma conta de desenvolvedor, enquanto o TWINT funciona sem necessitar de nenhum tipo de autenticação (TWITTER, 2018).

A utilização da biblioteca TWINT em um programa em Python pode ser visto na figura 2.1, onde é exemplificada uma busca pelo termo “Intervenção Militar” no Twitter. Os parâmetros adicionados na configuração da busca são utilizados pelo TWINT para gerar URLs customizadas do Twitter, onde, nesse caso, utilizou-se apenas o termo procurado, porém a biblioteca permite a utilização de diversas outras opções que estão listadas na documentação do projeto no GitHub.

Como discutido anteriormente, extrair informações de páginas que possuem uma estrutura conhecida torna-se uma tarefa trivial e uma análise rápida na página de buscas do

Twitter nos mostra um padrão simples para cada *tweet* do resultado. O TWINT aproveita-se desse padrão conhecido de HTML para coletar os resultados da busca avançada do Twitter e transformá-los em uma estrutura de dados no formato que o usuário necessite.

Para gerar estruturas baseadas nas informações Web, o TWINT emprega a biblioteca Beautiful Soup (disponibilizada em https://www.crummy.com/software/BeautifulSoup/, sob a licença MIT), desenvolvida em Python com o objetivo de facilitar a extração de dados de páginas Web, ela prove métodos simples para navegar, buscar e modificar representações de documentos.

Assim como em (GUPTA et al., 2003), em que é utilizada uma técnica de navegação pelo *Document Object Model* (DOM), uma representação do documento como um grupo estruturado de nós e objetos com métodos para acesso, para extrair informações relevantes de uma página Web, a biblioteca Beautiful Soup remonta uma árvore representativa da estrutura HTML para buscar os dados desejados. Essa técnica de extração de dados pode ser facilmente utilizada quando se conhece onde a informação desejada está situada dentro da hierarquia da página, seja um elemento HTML específico ou baseado em um padrão dentro da estrutura do DOM, porém essa abordagem é facilmente defasada quando ocorrem modificações na estrutura HTML das páginas mineradas.

2.2 Análise de sentimentos

Conhecer a opinião de outras pessoas sempre foi um fator importante no momento em que é preciso tomar decisões, seja para decidir que filme assistir, buscar informações sobre um local para jantar ou discutir a opinião de outros sobre um candidato nas eleições (LIU, 2012). Com o crescimento da Web, essas informações tornaram-se cada vez mais próximas das pessoas, seja em fóruns especializados, páginas de profissionais da área ou até mesmo nas redes sociais, onde as pessoas enviam mensagens diariamente sobre o que gostam ou desgostam (PAK; PAROUBEK, 2010).

A construção de sistemas que processam informações subjetivas efetivamente utiliza-se dos métodos computacionais de Análise de Sentimentos, uma área crescente do Processamento de Linguagem Natural que se preocupa não com o assunto de um documento, mas com a opinião que ele expressa. Dentro da área de Análise de Sentimentos, existem ainda diferentes abordagens para a marcação de documentos, seja com uma abordagem de subjetividade, em que o sistema decide se um documento é baseado em fatos (objetivo) ou em opiniões (subjetivo); Análise de força, em que um documento definido, por exemplo,

como tendo opinião positiva é categorizado em fraco, médio ou fortemente positivo; E, finalmente, análise de polaridade (ou orientação), que trata se um texto subjetivo expressa uma opinião positiva ou negativa sobre o assunto (ESULI; SEBASTIANI, 2006).

Em (LIU, 2012), o autor apresenta métodos para classificação de sentimentos de documentos, sendo que a maior parte das técnicas apresentadas na bibliografia utilizam-se de aprendizagem supervisionada, embora também existam abordagens utilizando aprendizagem não supervisionada. A classificação com aprendizagem supervisionada geralmente é classificada como um problema de classificação em duas classes, *positiva* e *negativa*, e normalmente é necessário a utilização de dados classificados, e.g., *reviews* com notas para o treinamento do algoritmo, embora alguns autores começaram a explorar a utilização de dados classificados automaticamente por algum método de extração especializado, um exemplo sendo a presença de *emoticons* representando felicidade ou tristeza (GO; BHAYANI; HUANG, 2009; READ, 2005; PAK; PAROUBEK, 2010). Embora qualquer algoritmo de aprendizagem de máquina possa ser capaz de realizar a classificação de sentimentos para documentos dada um conjunto de dados de treinamento bons suficientes, os melhores resultados da bibliografia têm sido com os algoritmos *Naive Bayes* e *Support Vector Machine* (SVM) (PANG; LEE; VAITHYANATHAN, 2002; LIU, 2012).

Embora sejam menos explorados, os métodos de classificação por aprendizado não supervisionado possuem a vantagem de não serem presos a um domínio específico como é o caso dos algoritmos que necessitam de treinamento (TABOADA et al., 2011; PANG; LEE, 2008; WOLOSZYN et al., 2017; WOLOSZYN et al., 2017; WOLOSZYN; NEJDL, 2018), portanto podem ser reutilizados para diferentes *dataset*s com resultados satisfatórios. Para os métodos não supervisionados podem ser utilizadas classificações baseadas em léxicos de análise sintática (TURNEY, 2002) ou dicionários que indicam a polaridade das palavras (PENNEBAKER; BOOTH; FRANCIS, 2007).

2.2.1 O dicionário LIWC

Com base em pesquisas dos anos 1980, onde foi descoberto que ao pedir para pessoas escreverem sobre eventos emocionais em suas vidas eles notavam uma subsequente melhora na sua saúde física, pesquisadores tentaram encontrar uma ligação entre o conteúdo dos documentos e as melhoras físicas apresentadas, primeiramente via auditores que liam o texto e os avaliavam em múltiplas dimensões, notando porém diversas discrepâncias entre a avaliação de diferentes auditores, além da quantidade de tempo gasto

para estas avaliações e o impacto que os auditores sofriam ao ler textos muito negativos. O passo seguinte, portanto, foi encontrar uma maneira de analisar os textos de maneira automatizada e o resultado foi a criação do *Linguistic Inquiry and Word Count* (LIWC) pelos autores Pennebaker, Francis e Booth em 2001, um programa que analisa documentos procurando e contando palavras relacionadas a categorias com relevância psicológica (TAUSCZIK; PENNEBAKER, 2010).

O dicionário LIWC, lançado inicialmente em 2001, já recebeu atualizações em 2007 e 2015 com mudanças tanto na categorização de palavras quanto nas categorias em si, sendo que categorias com uma taxa muito baixa de palavras foram removidas e outras adicionadas. A atualização de 2007 analisou mais de 100 milhões de palavras, com os detalhes do processo sendo documentado em Pennebaker, Chung, Ireland, Gonzales e Booth (2007), e o dicionário é composto por quase 4,500 palavras e raízes de palavras (e.g., a raiz *brav** que corresponderia tanto a *brave* quanto *bravery*), em que cada uma define uma ou mais categorias ou subdiretórios.

O funcionamento do LIWC baseia-se na análise dos documentos e contagem de palavras contidas nele, sendo que cada palavra é comparada ao dicionário definido pelo usuário e classificada em uma ou mais categorias, assim a porcentagem da ocorrência das palavras e das categorias relacionadas sugere a inclinação do texto a cada uma das categoria psicológicas (PENNEBAKER; BOOTH; FRANCIS, 2007).

Neste trabalho utilizou-se o dicionário LIWC2007 com tradução para Português Brasileiro. O trabalho de tradução foi realizado por 3 equipes, uma do NILC (Núcleo Interinstitucional de Linguística Computacional), uma da empresa Checon Pesquisa e outra da Unisinos (ALUíSIO; CHECCHIA; CHISHMAN, 2013), utilizando vários dicionários bilíngues Português-Inglês e é composto por 127,149 entradas.

A escolha de utilizar este dicionário é principalmente devido ao fato de se encontrar pouca bibliografia utilizando a versão em Português Brasileiro, que foi lançada recentemente (ALUíSIO; CHECCHIA; CHISHMAN, 2013), portanto auxiliando na avaliação da ferramenta e melhorando o ecossistema de análise de sentimentos de textos em português. Também é possível citar a qualidade da avaliação desta ferramenta como verificado na bibliografia (FILHO; PARDO; ALUÍSIO, 2013) que se mostra satisfatória.

3 TRABALHOS RELACIONADOS

A análise de sentimentos no Twitter não é um assunto trivial, levando em consideração o tamanho dos *tweets*, contidos em 140 ou 280 caracteres, as idiossincrasias da plataforma e a possibilidade de uma mensagem conter sentimentos diversos, como pode ser percebido em (BIFET; FRANK, 2010). Esses desafios causaram a utilização de diferentes tipos de abordagens para a análise de sentimento e polaridade dos *tweets*.

A pesquisa realizada em (ASUR; HUBERMAN, 2010) demonstra como conteúdo disponível nas mídias sociais pode ser utilizado para prever efeitos no mundo real. Ao extrair sentimentos de *tweets* do público geral, os autores conseguiram predizer tendências nos resultados de bilheteria com uma performance melhor do que predições quantitativas feitas via mercados artificiais.

Em (BOLLEN; MAO; ZENG, 2011), os autores investigaram se a mudança do humor do público, sendo analisado em uma coleção de larga escala de dados do Twitter, teria correlação ou poderia predizer resultados na bolsa de valores. A análise conseguiu encontrar com sucesso categorias que prediziam mudanças nos indicadores da bolsa de uma maneira melhor que outros analisadores de opinião.

Assim como feito nesta pesquisa, (TUMASJAN et al., 2010) utiliza palavras-chave dos *tweets* para fazer uma filtragem de relevância no conteúdo encontrado, em que foram coletados tweets relacionados a temas da eleição Alemã, incluindo nomes de partidos e de candidatos, e utiliza o dicionário LIWC para categorização dos *tweets* com objetivo de buscar correlações entre a polaridade do discurso e os resultados das eleições.

A utilização de *hashtags* para análise de sentimentos é explorada em (TEKA; NAINI; NIEDERÉE, 2016), onde os autores analisam a discussão corrente sobre a situação dos refugiados na Europa em 2015, extraindo dados da plataforma utilizando *hashtags* relacionadas aos refugiados como filtro. As mesmas *hashtags* depois são utilizadas como um indicador da polaridade em relação ao assunto, sendo categorizadas como positivas, neutras ou negativas, dependendo do seu conteúdo. O uso de *hashtags* também pode ser usado como medida base para relevância do assunto (ASUR; HUBERMAN, 2010; KOULOUMPIS; WILSON; MOORE, 2011) baseado na sua frequência de uso.

Em (FRANÇA; OLIVEIRA, 2014), os autores possuem um objetivo similar ao deste trabalho, ao coletar *tweets* sobre o período de protestos ocorrido de Junho a Agosto de 2013, os autores porém utilizam outro tipo de categorização, baseando-se em métodos de aprendizado supervisionado, utilizando o algoritmo classificador de *Naive Bayes*

treinado com *tweets* classificados por humanos, para análise de polaridade.

Em (FILHO; PARDO; ALUÍSIO, 2013), os autores analisam o dicionário LIWC em Português Brasileiro em uma comparação com outros dois léxicos *Opinion Lexicon* e *SentiLex*, limitando a avaliação às categorias *posemo* e *negemo* do dicionário LIWC, que correspondem respectivamente a emoção positiva e negativa. Os dados apresentados pelos autores mostram que o dicionário LIWC em Português Brasileiro performa melhor em textos positivos para a classificação de opinião e de sentenças, enquanto apresenta dificuldades para indicar negatividade em ambos tipo de documentos.

Os trabalhos recentes (JUNIOR et al., ; RODRIGUES et al., 2017) buscam inferir dados sobre os autores dos documentos analisados, retirados de uma rede social, utilizando o dicionário LIWC em Português Brasileiro. Os autores também propõe uma série de pré-processamentos nos dados do *dataset* para uma melhora na análise de polaridade.

Pode se ver na bibliografia o crescente uso do Twitter como uma plataforma para mineração de dados e análise de opiniões . Embora a bibliografia (FILHO; PARDO; ALUÍSIO, 2013; RODRIGUES et al., 2017; JUNIOR et al.,) aborde trabalhos que utilizam o dicionário LIWC em Português Brasileiro, não encontramos casos da utilização deste dicionário em grandes *dataset*s como o apresentado neste trabalho.

4 PROPOSTA E METODOLOGIA

Neste capítulo apresenta-se o processo proposto e realizado, seus componentes e funcionamento genérico, além da metodologia utilizada para a criação do *dataset*.

4.1 Construção do *dataset*

O foco na construção do *dataset* foi coletar dados históricos do Twitter dos anos de 2014 a 2017, um período que abrange desde a eleição da presidente Dilma Roussef e seu *impeachment* até os acontecimentos mais recentes da crise brasileira na presidência de Michel Temer. Como notado anteriormente, o Twitter disponibiliza uma ferramenta de busca avançada, porém limitações de sua API fizeram com que a abordagem proposta fosse baseada na utilização do TWINT, uma ferramenta existente e disponível para realizar a mineração de dados no Twitter.

Para coletar os dados que foram julgados importantes utilizou-se um processo similar ao de outros trabalhos da bibliografia, tais como (TEKA; NAINI; NIEDERÉE, 2016; OLTEANU et al., 2014; VIEWEG et al., 2010), os quais sugerem a utilização de palavras-chave, termos específicos ou *hashtags* relacionados ao assunto para a coleta dos *tweets*. Embora este método possa apresentar uma certa taxa de erro, ou *noise*, como visto em (VIEWEG et al., 2010), ele faz mais sentido do que utilizar filtros baseados em localização, que é outro método utilizado na bibliografia (TEKA; NAINI; NIEDERÉE, 2016; OLTEANU et al., 2014), principalmente pelo fato que o próprio usuário pode informar sua localização, que pode ser incorreta, no seu perfil do Twitter (TEKA; NAINI; NIEDERÉE, 2016)

A tabela 4.1 lista os parâmetros usados na mineração realizada pelo TWINT para a filtragem de *tweets*, sendo que o termo *Intervenção Militar* foi utilizado como uma expressão única, portanto todas ocorrências deveriam ser das duas palavras juntas. A decisão de utilizar este termo foi baseado na inclinação neutra da expressão em relação às discussões encontradas na Web, sendo utilizada tanto pelo público de esquerda quanto de direita, evitando um viés baseado puramente no termo de busca.

Para criar uma linha base dos resultados, tanto para validar hipóteses sobre a polaridade dos *tweets* quanto para ter uma comparação quanto aos sentimentos encontrados, foram selecionados 10 usuários do Twitter relacionados diretamente a canais de comunicação política, sendo 5 deles considerados esquerdistas e 5 deles considerados de direita

Tabela 4.1: Parâmetros utilizados para coleta de dados do Twitter

Termo de busca:	*Intervenção Militar*
Linguagem:	*Português*
Data Início:	*01/01/2014*
Data Fim:	*31/12/2017*

(ARAUJO, 2017). Nossa hipótese é de que estes resultados criem níveis comparativos para o resultado da análise geral, podendo portanto comparar se a opinião geral do público se compara aos níveis de categorização dos veículos de esquerda ou direita. Os dados minerados desses perfis seguem o mesmo princípio do *dataset* geral e respeitam os parâmetros demonstrados na Tabela 4.1, o número de *tweets* extraído foi de 178 para as mídias de esquerda e 81 dos perfis de direita.

Tabela 4.2: Canais de comunicação política utilizados como linha base

Esquerda:	**Direita:**
Mídia Ninja	Movimento Brasil Livre
Caros Amigos	Folha Política
Carta Capital	Movimento Contra Corrupção
Jornalistas Livres	Imprensa Viva
Revista Fórum	O Antagonista

4.2 Análise de opiniões

Para realizar a compilação de opiniões sobre os dados minerados do Twitter, foi utilizado o dicionário LIWC2007 em Português Brasileiro (ALUíSIO; CHECCHIA; CHISHMAN, 2013). Esse dicionário foi criado a partir da tradução direta do dicionário LIWC2007, e começou a ser explorado recentemente em trabalhos de análise de sentimentos de documentos em Português (JUNIOR et al., ; RODRIGUES et al., 2017; CARVALHO; GUEDES, 2017). Em (FILHO; PARDO; ALUÍSIO, 2013) foi realizada uma avaliação comparativa do dicionário LIWC em Português Brasileiro em relação a outros dois léxicos usados na língua portuguesa para análise de polaridade.

O LIWC é composto do dicionário e do aplicativo para leitura de documentos e categorização de palavras, porém por ser um software proprietário acabamos por utilizar uma algoritmo simples para a leitura dos *tweets* do *dataset* e categorização destes utilizando o dicionário LIWC. O algoritmo utilizado pode ser encontrado no repositório deste projeto (*gdelavald/intervencaoounao* no GitHub), ele faz uma leitura do banco de dados que contém os *tweets* e utiliza a biblioteca do LIWC em Python para classificar as pa-

lavras encontradas em classes do dicionário. Após esta classificação, é feita a contagem de palavras incluídas em cada categoria e sua comparação com a classificação geral das palavras, gerando assim a relevância de cada uma das categorias no documento analisado.

Ao utilizar o dicionário LIWC conseguimos categorizar os *tweets* de acordo com suas características textuais, baseando-se nos aspectos psicológicos de cada palavra. Na análise proposta as categorias mais importantes são as de polaridade, portanto *posemo* (Emoção Positiva) e *negemo* (Emoção Negativa), pois estas são utilizadas para tentar comprovar nossa hipótese de que os usuários do Twitter, ao utilizarem a palavra-chave "Intervenção Militar", declaram suas opiniões de satisfação ou insatisfação no contexto da mensagem. Assim, mensagens que são categorizadas como positivas (e.g. *"O melhor que pode acontecer no Brasil agora é a intervenção militar"*) expressam apoio à Intervenção Militar, enquanto mensagens que são categorizadas como negativas (e.g. *"Não aguento mais ouvir falar em Intervenção Militar, já chega!"*) expressam ideias contrárias à Intervenção.

Além da análise de categorizações em si, podemos também analisar suas mudanças em relação ao tempo, gerando a média através dos anos de 2014 a 2017, verificando assim a evolução da opinião público em relação ao tempo e aos desdobramentos da crise.

5 ANÁLISE DE DADOS E RESULTADOS

Neste capítulo iremos analisar os resultados da extração de dados do Twitter, seguindo a metodologia proposta no capítulo anterior, e fazer uma análise dos resultados, discutindo sobre possíveis interpretações da análise de polaridade dos dados obtidos.

5.1 Análise do *dataset*

A mineração de dados do Twitter retornou o total de 226.267 entradas, a distribuição de *tweets* por ano pode ser vista na Tabela 5.1. Esse número não inclui *retweets*, porém pode incluir duplicações intencionais dos usuários ou mensagens semelhantes com destinatários diferentes.

Tabela 5.1: Distribuição de *tweets* por ano no *dataset*

Ano:	**Número de Tweets:**
2014	43.212
2015	83.816
2016	43.495
2017	55.744
	226.267

A análise realizada pelo LIWC utiliza diferentes categorias, classificadas em processos linguísticos e psicológicos, para agrupar palavras encontradas nos documentos examinados (PENNEBAKER; BOOTH; FRANCIS, 2007). As categorias utilizadas para classificação utilizadas neste trabalho são as seguintes: *posemo* (Emoção Positiva), *negemo* (Emoção Negativa), *affect* (Processos Afetivos), *anger* (Raiva), *anx* (Ansiedade), *sad* (Tristeza), *cogmech* (Processos Cognitivos), *social* (Processos Sociais), *percept* (Processos Perceptivos), *bio* (Processos Biológicos) e *relativ* (Relatividade). Sendo o principal foco nas categorias de polaridade *posemo* e *negemo*.

Nas Tabelas 5.2 e 5.3 pode ser observada a classificação dos documentos dos canais de mídia política em relação às categorias do LIWC utilizada. Para a análise de dados desses canais, foram reunidos 178 *tweets* de esquerda e 81 de direita, uma pequena porção do *dataset* extraído do Twitter. Como análise inicial, nota-se que o discurso das perfis de *direita* são classificados mais acima em relação a Processos Afetivos (*affect*), que envolvem sentimentos (e.g., felicidade, choro, abandono) e abrangem também as categorias de Emoções Positivas e Negativas. Outro detalhe interessante é que ambas categorias de

emoções (*posemo* e *negemo*) possuem níveis mais altos no discurso de *direita* do que da *esquerda*.

Tabela 5.2: Média dos valores de análise para cada categoria dos canais de mídia de esquerda

	posemo	*negemo*	*affect*	*anger*	*anx*
Média:	0.02852	0.02905	0.06236	0.00925	0.00367
Variância:	0.04460	0.04962	0.07313	0.02740	0.01454
	sad	*cogmech*	*social*	*percept*	*bio*
Média:	0.00249	0.32607	0.10066	0.02733	0.01664
Variância:	0.01292	0.14684	0.09059	0.04772	0.03407

Na Tabela 5.4 são listadas as médias de cada categoria quanto a classificação do *dataset*, onde é possível verificar um valor significante para a categoria de Processos Cognitivos (*cogmech*), sendo essa a segunda maior categoria do LIWC (abrangendo categorias como *Ideia*, *Causa*, *Discrepância*, *Tentativa*, *Certeza*, *Inibição* e *Inclusividade*) e é relacionada a explicações e desenvolvimento de ideias (e.g., por causa, saber, deveria) (TAUSCZIK; PENNEBAKER, 2010).

Os valores dos gráficos abaixo foram obtidos calculando uma média da classificação dos *tweets* diária, sendo submetidos então a uma suavização de forma a melhorar a visualização dos dados. Na Figura 5.1 pode se ver a evolução da classificação dos dados através do período de tempo abrangido pelos *tweets*. A variabilidade das categorizações torna-se evidente neste gráfico, demonstrando a heterogeneidade das opiniões coletadas no Twitter, como é ressaltado nas referências (KOULOUMPIS; WILSON; MOORE, 2011; PAK; PAROUBEK, 2010).

Uma análise sobre as categorias de Emoções Positivas e Negativas demonstra uma categorização média parecida com a das mídias de *esquerda* visto anteriormente, porém pelo tamanho pequeno da amostra das mídias políticas não podemos afirmar com certeza essa similaridade.

Tabela 5.3: Média dos valores de análise para cada categoria dos canais de mídia de direita

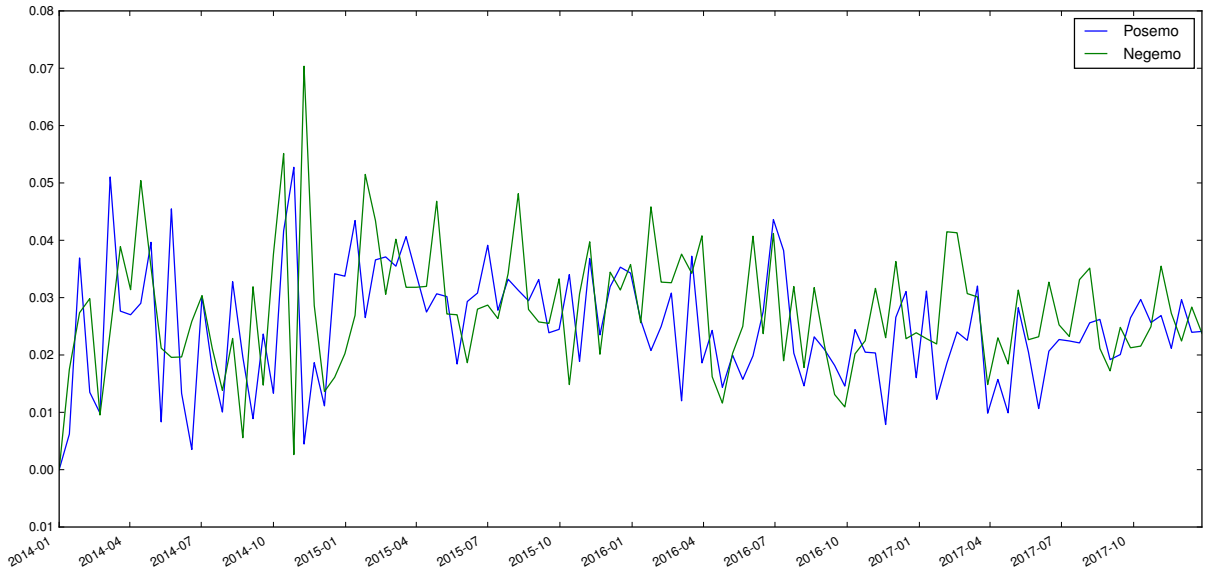
	posemo	*negemo*	*affect*	*anger*	*anx*
Média:	0.04132	0.04117	0.09518	0.00853	0.00671
Variância:	0.04400	0.03778	0.06367	0.02436	0.02106
	sad	*cogmech*	*social*	*percept*	*bio*
Direita:	0.00483	0.33409	0.14257	0.03347	0.04142
Variância:	0.01615	0.13021	0.09669	0.04043	0.05865

Tabela 5.4: Média dos valores de análise para cada categoria para **Intervenção Militar**

	posemo	*negemo*	*affect*	*anger*	*anx*
Média	0.02846	0.03666	0.07004	0.00944	0.00381
Variância	0.04781	0.05418	0.07781	0.02658	0.01705
	sad	*cogmech*	*social*	*percept*	*bio*
Média	0.00761	0.27937	0.11420	0.02974	0.02542
Variância	0.02402	0.18575	0.11322	0.047023	0.04756

Nas imagens abaixo é mostrada a evolução da classificação do *dataset* em cada categoria através do tempo, com a diferenciação dos dados gerais, denominados "Intervenção Militar" na legenda, e dos dados isolados de *esquerda* e *direita* sendo estes os *tweets* específicos de fontes de mídia respectivos.

Figura 5.1: Médias de *posemo* e *negemo* no *dataset* entre 2014 e 2017



Fonte: Elaborado pelo autor

5.2 Análise Temporal

Separando o *dataset* em períodos anuais, é possível identificar padrões que evoluem de ano a ano. A tabela 5.5 dispõe os dados de média e variância para as categorias de Emoção Positiva (*posemo*) e Emoção Negativa (*negemo*) entre os anos de 2014 a 2017,

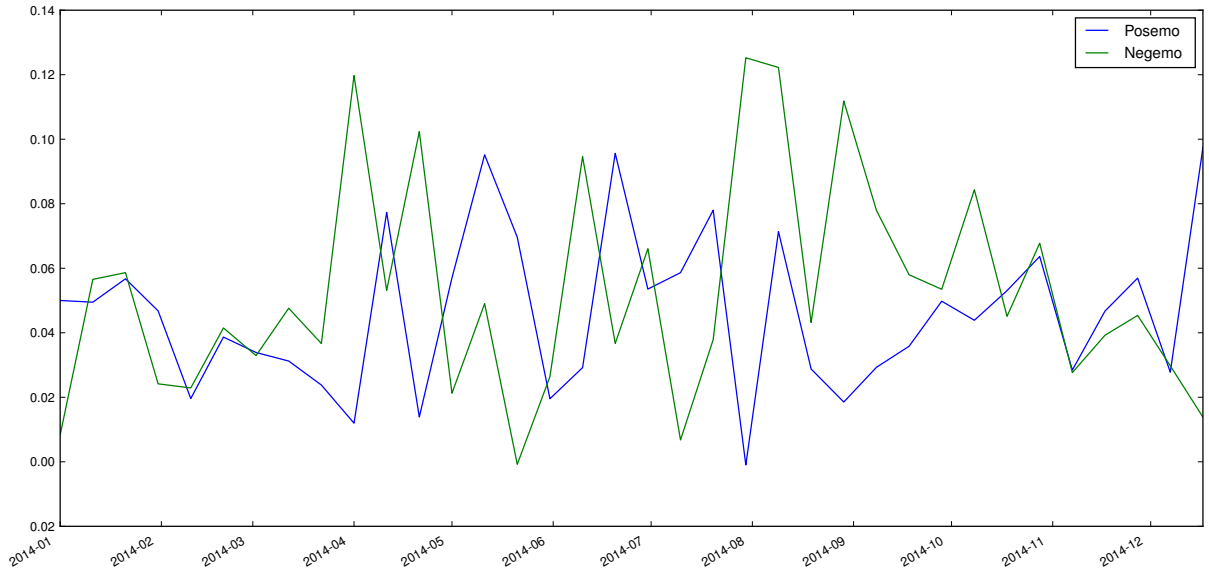
estes mesmos dados podem ser vistos nas figuras abaixo (5.2, 5.3, 5.4, 5.5) que mostram os dados de médias e uma suavização dos valores para uma melhor visualização da evolução através do tempo.

Tabela 5.5: Média dos valores de análise para cada categoria para **Intervenção Militar** através dos anos

		2014	**2015**	**2016**	**2017**
posemo	**Média**	0.04256	0.03898	0.03357	0.03501
	Variância	0.07558	0.07437	0.08078	0.09679
negemo	**Média**	0.05437	0.05168	0.04473	0.03901
	Variância	0.08599	0.08574	0.09739	0.10444

Em ambas as categorias, é possível ver um aumento geral na variância, demonstrando que os dados têm se tornado mais heterogêneos e dispersos, enquanto a média de ambas a categorias têm diminuído levemente no passar dos anos. Uma possível interpretação para a diminuição das médias das duas categorias é que os discursos se tornaram menos exacerbados e mais racionais ao longo dos anos.

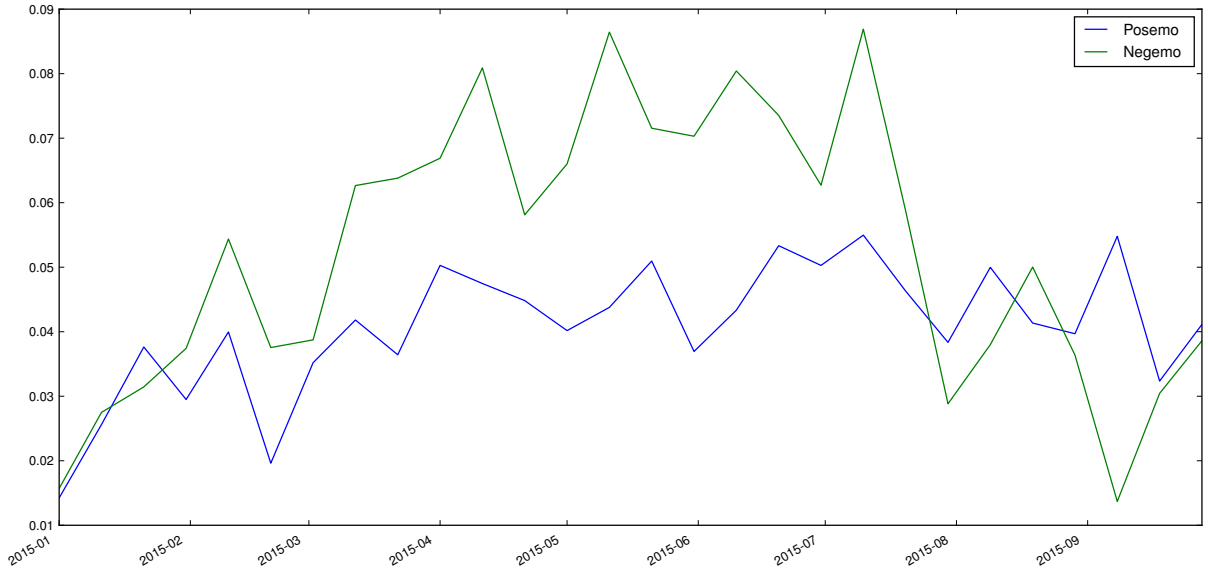
Figura 5.2: Gráfico para média de *posemo* e *negemo* em 2014



Fonte: Elaborado pelo autor

Em 2014 as categorias de *posemo* e *negemo*, mostradas nas figura 5.2, possuem o maior valor entre os períodos verificados, levando em conta o contexto da crise neste ano, que foi logo após os protestos de junho de 2013 (FRANÇA; OLIVEIRA, 2014) em que a popularidade da presidente Dilma Rousseff caiu, seguido de diversas denúncias de corrupção no governo e de uma crise (CASTRO, 2017) que afetou diversos setores da economia, causando discussões acentuadas nas redes sociais.

Figura 5.3: Gráfico para média de *posemo* e *negemo* em 2015

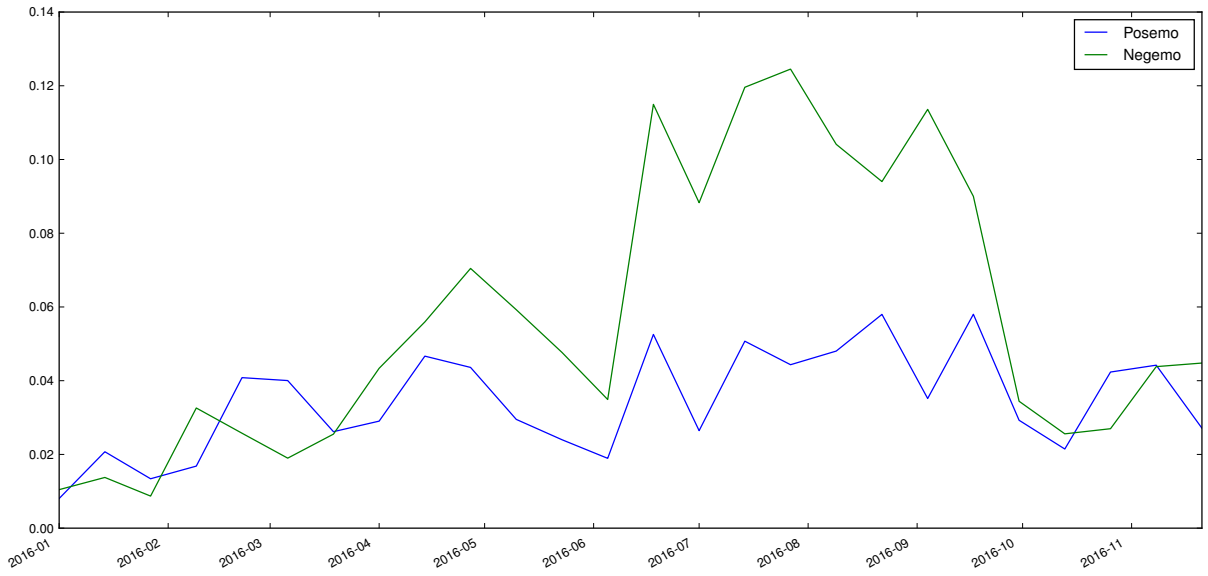


Fonte: Elaborado pelo autor

Em 2015, apesar de possuir o maior volume de *tweets* extraído entre os anos analisados, a análise dos dados mostra uma queda na média de ambas as categorias. A variância se mantêm no mesmo patamar do ano anterior. No gráfico 5.3 porém, é possível verificar um leve crescimento da média no meio do ano, com uma nova queda no final de 2015.

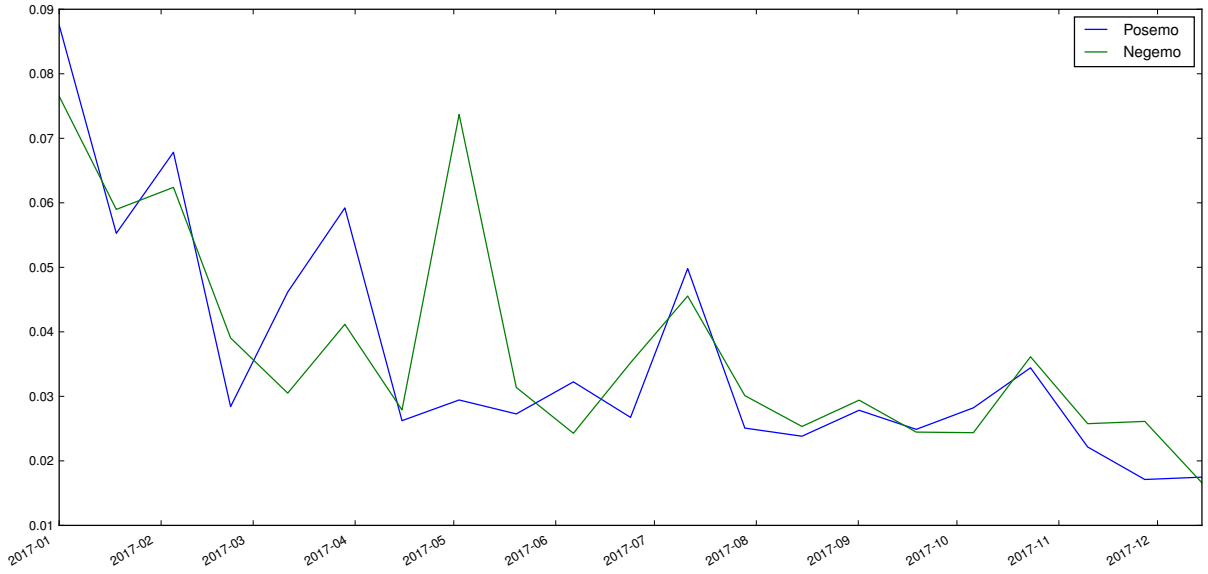
É possível notar a partir do meio do ano de 2016 um crescimento da média da categorização de *negemo*, o que pode ser relacionado ao final do processo de impeachment da presidente Dilma Rouseff, que foi encerrado em agosto daquele ano com a cassação do mandato da presidente.

Figura 5.4: Gráfico para média de *posemo* e *negemo* em 2016



Fonte: Elaborado pelo autor

Figura 5.5: Gráfico para média de *posemo* e *negemo* em 2017



Fonte: Elaborado pelo autor

6 CONCLUSÃO

Neste trabalho apresentou-se uma análise de polaridade e classificação de um *dataset* extraído do Twitter sobre reações à crise política brasileira de 2014 a 2017. Os dados apresentados demonstram a inclinação da opinião pública através do tempo sobre o debate a respeito da intervenção militar. Analisando os dados obtidos podemos verificar que, em média, o discurso dos usuários do Twitter sobre o termo "Intervenção Militar" possui uma conotação negativa, demonstrando um repúdio a ideia de intervenção militar no Brasil pela opinião pública, porém a diferença entre as categorias têm diminuído através dos anos que foram analisados, mostrando a tendência do discurso no Twitter sobre o assunto.

Ainda assim, é necessário admitir que as mensagens mineradas poderiam ser melhor interpretadas se fossem utilizados alguns dos métodos de pré-processamento definidos na bibliografia, como a correção de palavras incorretas devido ao uso de gírias, idiossincrasias do Twitter e redução do *noise* das mensagens (PAK; PAROUBEK, 2010; GO; BHAYANI; HUANG, 2009). Além disso, as mensagens online são sempre sujeitas a ambiguidades, principalmente em casos que uma mensagem pode ser acompanhada de *links* externos que contrariam o discurso, e.g., usuários enviando uma mensagem elogiando uma notícia que critica a intervenção militar. Nesse caso seu discurso apresentaria uma mensagem positiva, porém seu conteúdo na verdade é negativo em relação ao objeto analisado. Esses casos específicos abrem a possibilidade de uma análise de sentimentos mais profunda, analisando não só a mensagem enviada, mas também conteúdos multimídia ou links externos.

Trabalhos futuros podem trazer melhorias a alguns aspectos abordados neste trabalho, incluindo a verificação da qualidade da classificação que pode ser feita utilizando outros métodos de classificação similares ao LIWC como em (FILHO; PARDO; ALUÍSIO, 2013), ou mesmo indo mais a fundo na abordagem proposta de verificar veículos de esquerda e direita como meio de comparação dos resultados, que neste trabalho acabaram por não demonstrar grande influência principalmente pelo número pequeno de *tweets* encontrados por estes veículos e pela sua natureza de serem informativos e não subjetivos. Outro problema que pode ser citado é a grande utilização de *bots* no Twitter como forma de disseminar informações falsas ou opiniões favorecendo um lado da discussão, este problema é abordado em bibliografias que tratam de dados extraídos do Twitter mas não foi discutido neste trabalho, apesar de se mostrar um grande desafio quando estamos

avaliando a opinião pública utilizando este tipo de dados.

Como valor gerado neste projeto, deixamos acessível o *dataset* com todos dados extraídos do Twitter de acordo com a metodologia apresentada neste trabalho. Estes dados são abertos e acompanham o código que foi utilizado para o desenvolvimento da análise de polaridade. Os dados podem ser acessados no repositório *gdelavald/intervencaoounao* no GitHub.

O trabalho também apresenta valor ao utilizar o dicionário LIWC em Português Brasileiro, sendo esse um recurso ainda relativamente novo e que tem aparecido mais frequentemente em trabalhos de classificação textual de documentos em Português.

REFERÊNCIAS

ALUíSIO, S.; CHECCHIA, R.; CHISHMAN, R. Brazilian portuguese liwc 2007 dictionary. 2013. Available from Internet: <http://www.nilc.icmc.usp.br/portlex/index.php/pt/projetos/liwc>.

ARAUJO, P. Z. de. "o antipetismo é pouco politizado", diz o professor pablo ortellado, da usp, ao dcm. por zambarda. **Diário do Centro do Mundo**, Aug 2017. Acesso em: Junho 2018. Available from Internet: <https://bit.ly/2KN8pao>.

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.

BIFET, A.; FRANK, E. Sentiment knowledge discovery in twitter streaming data. In: **Proceedings of the 13th International Conference on Discovery Science**. Berlin, Heidelberg: Springer-Verlag, 2010. (DS'10), p. 1–15. ISBN 3-642-16183-9, 978-3-642-16183-4. Available from Internet: <http://dl.acm.org/citation.cfm?id=1927300.1927301>.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, n. 1, p. 1 – 8, 2011. ISSN 1877-7503. Available from Internet: <http://www.sciencedirect.com/science/article/pii/S187775031100007X>.

CARVALHO, F.; GUEDES, G. P. Night sleep deprivation: Computational analysis of language effects. In: ACM. **Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web**. [S.l.], 2017. p. 221–224.

CASTRO, J. R. Quanto a população brasileira já perdeu com a crise econômica. **Carta Capital**, Sep 2017. Acesso em: Junho 2018. Available from Internet: <http://bitly.com/2uaPqey>.

CHAKRABARTI, S. et al. Mining the link structure of the world wide web. **IEEE computer**, Citeseer, v. 32, n. 8, p. 60–67, 1999.

ESULI, A.; SEBASTIANI, F. Determining term subjectivity and term orientation for opinion mining. In: **11th Conference of the European Chapter of the Association for Computational Linguistics**. [S.l.: s.n.], 2006.

FILHO, P. P. B.; PARDO, T. A. S.; ALUÍSIO, S. M. An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**. [S.l.: s.n.], 2013.

FRANÇA, T. C. de; OLIVEIRA, J. Análise de sentimento de tweets relacionados aos protestos que ocorreram no brasil entre junho e agosto de 2013. In: **Proceedings of the III Brazilian Workshop on Social Network Analysis and Mining (BRASNAN)**. [S.l.: s.n.], 2014. p. 128–139.

GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. **CS224N Project Report, Stanford**, v. 1, n. 12, 2009.

GUPTA, S. et al. Dom-based content extraction of html documents. In: **Proceedings of the 12th International Conference on World Wide Web**. New York, NY, USA: ACM, 2003. (WWW '03), p. 207–214. ISBN 1-58113-680-3. Available from Internet: <http://doi.acm.org/10.1145/775152.775182>.

JANSEN, B. J. et al. Micro-blogging as online word of mouth branding. In: **CHI '09 Extended Abstracts on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2009. (CHI EA '09), p. 3859–3864. ISBN 978-1-60558-247-4. Available from Internet: <http://doi.acm.org/10.1145/1520340.1520584>.

JUNIOR, L. A. da P. et al. Inferindo o sexo de usuarios de redes sociais utilizando o liwc em português do brasil.

KOSALA, R.; BLOCKEEL, H. Web mining research: A survey. **ACM Sigkdd Explorations Newsletter**, ACM, v. 2, n. 1, p. 1–15, 2000.

KOULOUMPIS, E.; WILSON, T.; MOORE, J. D. Twitter sentiment analysis: The good the bad and the omg! **Icwsm**, v. 11, n. 538-541, p. 164, 2011.

LIU, B. **Web data mining: exploring hyperlinks, contents, and usage data**. [S.l.]: Springer Science & Business Media, 2007.

LIU, B. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.

MüZELL, L. Para historiadora, intervenção militar no brasil "não pode mais ser descartada". **Carta Capital**, Sep 2017. Acesso em: Junho 2018. Available from Internet: <https://bit.ly/2zgXa4w>.

OLTEANU, A. et al. Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In: **ICWSM**. [S.l.: s.n.], 2014.

PAK, A.; PAROUBEK, P. Twitter as a corpus for sentiment analysis and opinion mining. In: . [s.n.], 2010. v. 2010, p. 1320–1326. Available from Internet: <http://www.mendeley.com/research/twitter-corpus-sentiment-analysis-opinion-mining-18/>.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends® in Information Retrieval**, v. 2, n. 1–2, p. 1–135, 2008. ISSN 1554-0669. Available from Internet: <http://dx.doi.org/10.1561/1500000011>.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up?: sentiment classification using machine learning techniques. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10**. [S.l.], 2002. p. 79–86.

PENNEBAKER, J. W.; BOOTH, R. J.; FRANCIS, M. E. Linguistic inquiry and word count: Liwc [computer software]. **Austin, TX: liwc. net**, 2007.

READ, J. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the ACL student research workshop**. [S.l.], 2005. p. 43–48.

RODRIGUES, R. G. et al. Inferência de idade utilizando o liwc: identificando potenciais predadores sexuais. In: **XXXVII Congresso da Sociedade Brasileira de Computaç ao**. [S.l.: s.n.], 2017.

SRIVASTAVA, J. et al. Web usage mining: Discovery and applications of usage patterns from web data. **SIGKDD Explor. Newsl.**, ACM, New York, NY, USA, v. 1, n. 2, p. 12–23, jan. 2000. ISSN 1931-0145. Available from Internet: <http://doi.acm.org/10.1145/846183.846188>.

TABOADA, M. et al. Lexicon-based methods for sentiment analysis. **Computational linguistics**, MIT Press, v. 37, n. 2, p. 267–307, 2011.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The psychological meaning of words: Liwc and computerized text analysis methods. **Journal of language and social psychology**, Sage Publications Sage CA: Los Angeles, CA, v. 29, n. 1, p. 24–54, 2010.

TEKA, A.; NAINI, K. D.; NIEDERÉE, C. Welcome or not-welcome: Reactions to refugee situation on social media. **CoRR**, abs/1610.02358, 2016.

TUMASJAN, A. et al. Predicting elections with twitter: What 140 characters reveal about political sentiment. In: **ICWSM**. [S.l.: s.n.], 2010.

TURNEY, P. D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 40th annual meeting on association for computational linguistics**. [S.l.], 2002. p. 417–424.

TWITTER. Twitter api documentation. 2018. Acesso em: Junho 2018. Available from Internet: <https://developer.twitter.com/en/docs>.

VIEWEG, S. et al. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: ACM. **Proceedings of the SIGCHI conference on human factors in computing systems**. [S.l.], 2010. p. 1079–1088.

WOLOSZYN, V. et al. Beatnik: an algorithm to automatic generation of educational description of movies. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S.l.: s.n.], 2017. v. 28, n. 1, p. 1377.

WOLOSZYN, V.; NEJDL, W. Distrustrank: Spotting false news domains. In: ACM. **Proceedings of the 10th ACM Conference on Web Science**. [S.l.], 2018. p. 221–228.

WOLOSZYN, V.; SANTOS, H.; WIVES, L. The influence of readability aspects on the user's perception of helpfulness of online reviews. Revista de Sistemas de Informação da FSMA, 2016.

WOLOSZYN, V. et al. Mrr: an unsupervised algorithm to rank reviews by relevance. In: ACM. **Proceedings of the International Conference on Web Intelligence**. [S.l.], 2017. p. 877–883.