UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

DIEGO COSTA TUMITAN

Atributos Discriminantes Baseados em Sentimento para a Predição de Pesquisas Eleitorais: Um Estudo de Caso no Cenário Brasileiro

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação

Orientador: Prof. Dr. Karin Becker

Porto Alegre
2014

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Costa Tumitan, Diego

Atributos Discriminantes Baseados em Sentimento para a Predição de Pesquisas Eleitorais: Um Estudo de Caso no Cenário Brasileiro / Diego Costa Tumitan. – Porto Alegre: PPGC da UFRGS, 2014.

111 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2014. Orientador: Karin Becker.

1. Mineração de Opiniões. 2. Previsão baseada em sentimento. 3. Classificação de Sentimento. 4. Conteúdo gerado por usuário. I. Becker, Karin. II. Título.

AGRADECIMENTOS

Agradeço primeiramente a professora Karin, pela paciência, orientação e me motivar nos momentos mais difíceis. À minha família, por todo o suporte que me deram durante o mestrado, especialmente a minha irmã pelas inúmeras horas de FaceTime que me ajudaram a suportar a saudade e a distância.

Aos amigos que tive a oportunidade de conhecer durante o mestrado, em especial: Juan Luis, Majed, Guilherme, Júlia, Vitório, Anderson, Bruno, Alan, Jonas, Matheus e Carlos. Além de todos os outros colegas do laboratório 213, que me acompanharam nestes anos. Agradeço à UFRGS e todos os professores do PPGC que me auxiliaram com esta pesquisa. Agradeço ao CNPq pela bolsa de estudos e auxílio financeiro.

“A sorte favorece a mente bem preparada.”

— LOUIS PASTEUR

RESUMO

O sucesso da mineração de opiniões para processar automaticamente grandes quantidades de conteúdo opinativo disponíveis na Internet tem sido demonstrado como uma solução de baixa latência e mais barata para a análise de opinião pública. No presente trabalho foi investigado se é possível prever variações de intenção de voto com base em séries temporais de sentimento extraídas de comentários de notícias, utilizando três eleições brasileiras como estudo de caso. As contribuições deste estudo de caso são: a) a comparação de duas abordagens para a mineração de opiniões em conteúdo gerado por usuários em português do Brasil; b) a proposta de dois tipos de atributos discriminantes para representar o sentimento em relação a candidatos políticos a serem usados para a previsão, c) uma abordagem para prever variações de intenção de voto que é adequada para cenários de dados esparsos. Foram desenvolvidos experimentos para avaliar a influência dos atributos discriminantes propostos em relação a acurácia da previsão, e suas respectivas preparações. Os resultados mostraram uma acurácia de 70% na previsão de variações de intenção de voto positivas e negativas. Estas contribuições são importantes passos em direção a um *framework* que é capaz de combinar opiniões de diversas fontes para encontrar a representatividade de uma população alvo, de modo que se possa obter previsões mais confiáveis.

Palavras-chave: Mineração de Opiniões. Previsão baseada em sentimento. Classificação de Sentimento. Conteúdo gerado por usuário.

Sentiment-based Features for Predicting Election Polls: A Case Study on the Brazilian Scenario

ABSTRACT

The success of opinion mining for automatically processing vast amounts of opinionated content available on the Internet has been demonstrated as a less expensive and lower latency solution for gathering public opinion. In this work, we investigate whether it is possible to predict variations in vote intention based on sentiment time series extracted from news comments, using three Brazilian elections as case study. The contributions of this case study are: a) the comparison of two approaches for opinion mining in user-generated content in Brazilian Portuguese; b) the proposition of two types of features to represent sentiment behavior towards political candidates that can be used for prediction, c) an approach to predict polls vote intention variations that is adequate for scenarios of sparse data. We developed experiments to assess the influence on the forecasting accuracy of the proposed features, and their respective preparation. Our results display an accuracy of 70% in predicting positive and negative variations. These are important contributions towards a more general framework that is able to blend opinions from several different sources to find representativeness of the target population, and make more reliable predictions.

Keywords: Opinion Mining, Sentiment-based prediction, Sentiment classification, User-generated content.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

C	Cumulativo
CP	Curto Prazo
DIJA	Down Jones Industrial Average
DP	Desvio Padrão
LIWC	Linguistic Inquiry and Word Counts
MO	Mineração de Opiniões
NLTK	Natural Language Toolkit
POS	*Part-Of-Speech*
PLN	Processamento de Linguagem Natural
RSS	*Rich Site Summary*
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine

SUMÁRIO

1 INTRODUÇÃO

Opiniões são fundamentais para quase todas as atividades humanas e são as principais influenciadoras de nossos comportamentos. Governos, empresas e organizações dependem de opinião pública para definir estratégias para melhorar os serviços que prestam, ou aumentar o sucesso e visibilidade de suas marcas, entidades e causas que representam. Telefonemas, pesquisas, questionários e plebiscitos estão entre as técnicas mais comuns para investigar a opinião de uma população alvo. No entanto, estas abordagens são caras, demoradas e, portanto, a sua frequência depende de orçamento e definição de amostra. Elas também possuem alta latência, o que pode afetar a precisão dos resultados. De fato, quanto mais longo o tempo decorrido entre a coleta das opiniões e disponibilidade dos resultados de sua análise, maior a probabilidade de as pessoas mudarem a sua opinião ou terem sido influenciadas.

Um possível complemento ou alternativa para estas técnicas é explorar automaticamente a opinião que as pessoas expõem cada vez mais em redes sociais (e.g. Facebook, Twitter, blogs), sites, notícias on-line, etc. A mineração de opiniões visa identificar automaticamente o conteúdo opinativo, e determinar o sentimento, percepção ou a atitude das pessoas em relação a uma entidade ou tópico (LIU, 2012). Usando mineração de opiniões, é possível analisar automaticamente este vasto e rico conteúdo gerado por usuários e desenvolver soluções de baixa latência, mais baratas e com um grau razoável de acurácia.

No contexto político, o levantamento da opinião das pessoas sobre seus representantes, organizações públicas, políticos e seus partidos, pode auxiliar em decisões políticas, ações de campanha, políticas governamentais, estratégias de marketing, etc. Abordagens tradicionais envolvem pesquisas caras (e, portanto, infrequentes) para a detecção a popularidade de políticos, aprovação de governo, intenção de voto, etc. O potencial da mineração de opiniões para opiniões mais atualizadas tem sido demonstrado em relação a mídias sociais, particularmente o Twitter (BOLLEN; PEPE; MAO, 2009; PAK, 2010; ASUR; HUBERMAN, 2010; TUMASJAN et al., 2010; GUERRA et al., 2011).

Assumindo que sentimento humano possa ser caracterizado por técnicas automáticas em um nível de precisão aceitável, a próxima questão é se este sentimento pode ser usado para prever o comportamento futuro. Previsões baseadas em sentimento foram utilizadas para fins tais como prever o movimento do mercado de ações (BOLLEN; MAO; ZENG, 2011; GILBERT; KARAHALIOS, 2010), resultados de eleições ou de pesquisas (O'CONNOR et al., 2010; TUMASJAN et al., 2010), desempenho de vendas (LIU et al., 2007) e bilheteria de filmes (ASUR; HUBERMAN, 2010). Estes trabalhos têm três características em comum: a) a fonte primária

de sentimento é o Twitter, b) foram feitos considerando a língua inglesa, e c) usam abordagens estatísticas (e.g. regressão linear) ou de aprendizado de máquina (e.g. redes neurais) para previsão, com base em séries temporais.

1.1 Problema

A maioria dos trabalhos em predição baseada em sentimento usa séries temporais longas e diárias, tanto para o sentimento, quanto para as variáveis a serem previstas, pois a precisão dos métodos de previsão utilizados é altamente dependente da quantidade de dados disponíveis para análise, seja o comprimento da série temporal, ou o número de séries temporais que podem ser consideradas para as variáveis da mesma natureza. No entanto, valores históricos para alguns tipos de variáveis podem ser esparsos.

Por exemplo, nos Estados Unidos da América, a intenção de voto para a eleição presidencial é coletada diariamente por várias organizações (e.g. emissoras de televisão, companhias de marketing, etc.). Já no Brasil, o cenário é completamente diferente, pois as pesquisas de intenção de voto só podem ser publicadas por companhias de consultoria de pesquisa autorizadas e mediante várias restrições metodológicas que garantem sua neutralidade. As principais organizações de coleta de opinião, tais como Ibope e Datafolha, publicam aproximadamente doze pesquisas por ano, a maioria concentrada no mês que precede as eleições. Consequentemente, o tempo decorrido entre quaisquer duas pesquisas publicadas varia enormemente, de dias a meses. Além da intenção de voto, existem vários outros contextos que apresentam dados temporais esparsos. No contexto governamental, dados do censo, indicadores de aprovação do governo e relatórios anuais de governos, são exemplos de dados esparsos. Por exemplo, pode-se desejar investigar se indicadores públicos de saúde, educação ou segurança podem ser preditos utilizando a opinião pública expressa sobre estes serviços. Contudo, as opiniões podem ser coletadas e processadas diariamente, mas estes indicadores são esparsos no tempo.

Além disso, a representatividade de usuários do Twitter para predição em certos domínios, em particular da política, tem sido severamente questionada (SCHOEN et al., 2013). Certamente, usar uma única fonte de opinião pode introduzir um viés no comportamento observado e no modelo de previsão, pois as opiniões expressas são representativas de uma população específica. Então, diferentes fontes de opinião devem ser exploradas, de modo que se possa analisar o comportamento subjacente e o público que representam. Por exemplo, no domínio político, foi observado em (TUMITAN; BECKER, 2013) que os autores de comentários de notícias têm um comportamento diferentes quando comparado aos dos usuários do Twitter (TUMASJAN

et al., 2010; PAK, 2010; BOLLEN; MAO; ZENG, 2011). Ao invés de apoiar ou criticar os candidatos, eles expõem suas visões e convicções sobre política em geral. Entender o papel, o comportamento, e como ponderar opiniões de diversas fontes, são passos fundamentais em direção a um problema real e desafiador, que é combinar diferente fontes de opinião para constituir uma amostra representativa da população em geral.

1.2 Objetivo

O objetivo desse trabalho é o desenvolvimento de um estudo de caso que permita investigar se é possível prever variações em intenção de votos baseando-se no sentimento expresso em conteúdo gerado por usuário em jornais, e dados de pesquisas de opiniões pública que são esparsos. Mais especificamente, foi considerado o cenário político brasileiro, no qual resultados de pesquisas eleitorais só podem ser publicados mediante restrições. Os dados resultados de pesquisas de intenção de voto caracterizam dados esparsos, onde a maioria delas são no mês que precede o dia das eleições.

O estudo de caso desenvolvido pode ser dividido em dois aspectos: mineração de opiniões e predição de indicadores externos. O processo de mineração de opiniões desenvolvido neste estudo de caso leva em consideração e é adaptado à fonte de dados utilizada, que são comentários gerados por usuários em notícias política on-line referentes a 3 eleições brasileiras. O jornal utilizado neste estudo de caso foi o Folha de São Paulo, que possui audiência nacional. Também foram consideradas as propriedades intrínsecas destes comentários, tais como português do Brasil, linguagem informal, formas de expressão de opinião, etc. No processo de mineração de opiniões foram experimentadas e comparadas duas abordagens de classificação de sentimento: baseada em dicionário e em aprendizado de máquina. Na primeira abordagem, foi utilizado um dicionário de sentimentos para classificar os termos presentes nos comentários das notícias. Já na segunda abordagem, foram utilizadas técnicas de aprendizado de máquina supervisionado para classificar o conteúdo analisado.

O objetivo da etapa de predição é prever indicadores públicos de aceitação. Neste estudo de caso, foram utilizadas pesquisas públicas de intenção de voto, as quais são poucos frequentes, irregulares e concentradas no mês que precede a eleição. Estas pesquisas são providas por uma instituição de credibilidade reconhecida, o Datafolha. Foram propostos atributos discriminantes baseados no sentimento extraído dos comentários de notícias para prever a variação das intenções de voto. Os atributos discriminantes propostos são: a) variações do total de sentimento extraído dos comentários; b) total de menções aos candidatos, e; c) explosões de ex-

pressão de sentimento. Estes atributos discriminantes foram preparados para representar tanto o efeito cumulativo do sentimento (desde o início da campanha eleitoral), quanto seu efeito de curto prazo (desde a última pesquisa de intenção de voto publicada). O primeiro traduz a estratégia geral da campanha, enquanto que o segundo representa o curso de ações tomadas em resposta o resultado das pesquisas. Além disso, foi analisado o poder preditivo destes atributos discriminantes.

1.3 Contribuição

Este trabalho investiga através de um estudo de caso se é possível prever variações de intenção de voto utilizando o sentimento expresso em comentário de notícias. As contribuições desta pesquisa são:

- O processo de mineração de opiniões adaptado a comentários gerados por usuários escritos em português do Brasil como reação a notícias políticas, os quais possuem erros tipográficos, linguagem não-estruturada, linguagem chula, gírias, internetês, etc.;
- A comparação do desempenho de duas abordagens de classificação de sentimento no processo mineração de opiniões (baseada em dicionário, e em aprendizado de máquina);
- A proposta de métricas que sumarizam totais de sentimento e menções aos candidatos observados. Também foram propostas métricas baseadas em explosões de expressão de sentimento;
- A maneira que a previsão foi tratada no estudo de caso desenvolvido como um problema de classificação, de modo que se possa lidar com um cenário com dados temporais esparsos;
- Experimentos analisando a influência dos atributos discriminantes baseados em sentimento propostos sobre a variação da intenção de voto, de forma cumulativa e a curto prazo;

Estes são importantes passos em direção a um *framework* que é capaz de combinar opiniões de diversas fontes para encontrar a representatividade de uma população alvo, de modo que se possa obter previsões mais confiáveis. Os resultados preliminares deste trabalho foram apresentados por meio de diversas publicações com resultados parciais (TUMITAN; BECKER, 2013; BECKER; TUMITAN, 2013; TUMITAN; BECKER, 2014).

1.4 Organização do Trabalho

O trabalho está estruturado da seguinte forma. O Capítulo 2 apresenta a fundamentação teórica. O Capítulo 3 discute os trabalhos relacionados. O Capítulo 4 apresenta a descrição do estudo de caso realizado no presente trabalho. Os capítulos 5 e 6 detalham e apresentam os resultados do estudo de caso desenvolvido quanto à mineração de opiniões e a predição de intenção de voto, respectivamente. O Capítulo 7 apresenta conclusões e discute trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

O propósito deste capítulo é apresentar os conceitos subjacentes à mineração de opiniões, e caracterizar cada uma das etapas do processo, descrevendo os problemas envolvidos, e as técnicas que podem ser utilizadas. Boa parte deste capítulo consiste em um texto previamente publicado em (BECKER; TUMITAN, 2013).

2.1 Mineração de Opiniões

A *mineração de opiniões*, também chamada de *análise de sentimento* ou *análise de subjetividade* (LIU, 2012; TSYTSARAU; PALPANAS, 2012; PANG; LEE, 2008), é uma disciplina recente que congrega pesquisas de mineração de dados, linguística computacional, recuperação de informações, inteligência artificial, entre outras. A mineração de opiniões é definida em (LIU, 2010) como qualquer estudo feito computacionalmente envolvendo opiniões, sentimentos, avaliações, atitudes, afeições, visões, emoções e subjetividade, expressos de forma textual. O problema da mineração de opiniões pode ser estruturado em termos das seguintes tarefas genéricas (TSYTSARAU; PALPANAS, 2012): a) identificar as opiniões expressas sobre determinado assunto ou alvo em um conjunto de documentos; b) classificar a orientação ou polaridade desta opinião, isto é, se tende a positiva ou negativa; e c) apresentar os resultados de forma agregada e sumarizada. A polaridade da opinião define o sentimento, percepção ou atitude do público em relação ao alvo da opinião.

Como mostra a Figura 2.1, boa parte dos trabalhos nesta área concentrou-se no desenvolvimento de técnicas para detecção e sumarização automáticas de opinião sobre revisões de produtos e serviços (HU; LIU, 2004; PANG; LEE; VAITHYANATHAN, 2002; TURNEY, 2002; DAVE; LAWRENCE; PENNOCK, 2003; GHANI et al., 2006; ARCHAK; GHOSE; IPEIROTIS, 2007). Posteriormente, o foco ampliou-se para entidades específicas (e.g. políticos, celebridades, marcas) em redes sociais (PAK, 2010; GUERRA et al., 2011) ou notícias (GODBOLE; SRINIVASAIAH; SKIENA, 2007). A mineração de opiniões sobre textos menos estruturados, como notícias e blogs, também tem sido alvo de bastante atenção (BALAHUR et al., 2009; BALAHUR et al., 2010; KU; LIANG; CHEN, 2006). Outra vertente importante é a análise de opinião do Twitter, em particular visando estabelecer modelos preditivos (ASUR; HUBERMAN, 2010; BOLLEN; MAO; ZENG, 2011; TUMASJAN et al., 2010).

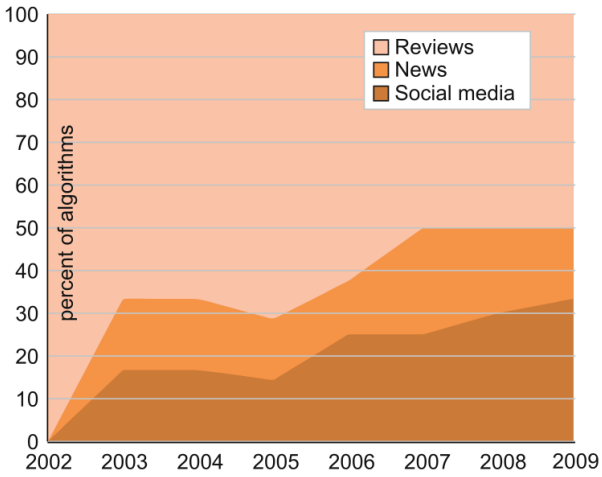

Figura 2.1 – Distribuição dos trabalhos sobre domínios alvo (Fonte: (TSYTSARAU; PALPANAS, 2012)).

2.2 Definições

A mineração de opiniões opera sobre porções de texto de quaisquer tamanho e formato, tais como páginas web, posts, comentários, *tweets*, revisões de produto, etc. Toda opinião é composta de pelo menos dois elementos chave: um *alvo* e um *sentimento* sobre este alvo (LIU, 2012). Um alvo pode ser uma entidade, aspecto de uma entidade, ou tópico, representando um produto, pessoa, organização, marca, evento, etc. Já um sentimento representa uma atitude, opinião ou emoção que o autor da opinião tem a respeito do alvo. A *polaridade* de um sentimento corresponde a um ponto em alguma escala que representa a avaliação positiva, neutra ou negativa do significado deste sentimento (TSYTSARAU; PALPANAS, 2012). O conceito de *aspecto*, também denominado característica ou propriedade, permite que uma entidade seja vista através de diferentes perspectivas ou atributos, ou como uma hierarquia de partes e subpartes (LIU, 2010).

Mais formalmente, uma opinião corresponde a uma quíntupla $(\mathbf{e}_i, \mathbf{a}_{ij}, \mathbf{s}_{ijkl}, \mathbf{h}_k, \mathbf{t}_l)$ (LIU, 2010), onde:

- $\mathbf{e}_i$: é o nome de uma entidade;
- $\mathbf{a}_{ij}$: é um aspecto da entidade $\mathbf{e}_i$ (opcional);
- $\mathbf{s}_{ijkl}$: é a polaridade do sentimento sobre aspecto $\mathbf{a}_{ij}$ que tem como alvo a entidade $\mathbf{e}_i$;
- $\mathbf{h}_k$: é o detentor do sentimento (i.e., quem expressou o sentimento), também chamado de fonte de opinião;
- $\mathbf{t}_l$: é o instante no qual a opinião foi expressa por $\mathbf{h}_k$.

Os termos *sentimento* e *opinião* frequentemente são usados como sinônimo neste con-

texto. A polaridade de um sentimento pode ser classificada em classes discretas (e.g. positiva, negativa ou neutra), ou como um intervalo que representa a intensidade deste sentimento, tipicamente [-1, 1].

Neste trabalho, usaremos os termos *sentimento* e *opinião* como sinônimos. O trabalho considera apenas as entidades como alvo do sentimento, ignorando aspectos específicos.

2.3 Níveis de Análise Textual

A detecção do sentimento em um texto pode ocorrer em diferentes granularidades, sendo que a decisão do nível está sujeita ao contexto e aplicação. A análise pode ser em nível de (LIU, 2012):

- *Documento*: nesse nível, a tarefa é classificar se um documento, tratado como um todo, expressa um sentimento positivo ou negativo. Esta granularidade é adequada quando o documento trata de uma única entidade, por exemplo, um documento que forneça uma opinião sobre um dado produto;
- *Sentença*: determina o sentimento de uma sentença específica de um documento. Este nível é bastante utilizado quando um mesmo documento contém opiniões sobre várias entidades. Ele também permite identificar e distinguir sentenças objetivas (fatos) e subjetivas (opiniões).
- *Entidade e Aspecto*: este nível foca na opinião expressa, independentemente dos construtos utilizados para expressá-la (e.g. documentos, sentenças, orações). Neste caso, o alvo da opinião pode ser uma entidade ou algum de seus aspectos. No exemplo "Adoro minha câmera X porque a qualidade de sua lente é excepcional. Pena que o preço seja tão alto", observa-se que existem três opiniões em 2 sentenças: sobre a câmera X e sobre dois de seus aspectos (preço e lente). Este é o nível mais complexo de análise, o qual tem sido bastante estudado no contexto de revisões de produtos e serviços (e.g. (HU; LIU, 2004; THET; NA; KHOO, 2010; GHANI et al., 2006)).

Neste trabalho, a detecção de sentimento ocorrerá em nível de sentença, pois frequentemente um comentário de jornal não possui uma opinião prevalecente e possui múltiplas referências as entidades observadas.

2.4 Tipos de Opiniões e Formas de Expressão

Opiniões referem-se a conteúdo subjetivo, escrito em linguagem natural. A forma como as opiniões estão expressas influencia diretamente a habilidade de processá-las corretamente. A mineração de opiniões tem origens em comum com a linguística computacional, com a qual compartilha problemas e desafios (LIU, 2012).

Opiniões podem ser *regulares* ou *comparativas*; *diretas* ou *indiretas*, *implícitas* ou *explícitas*. Em opiniões regulares, o autor da opinião expressa um sentimento, atitude, emoção ou percepção sobre um alvo ("Este filme é muito bom"). Já as opiniões comparativas expressam o sentimento com base na relação de similaridades ou diferenças entre duas ou mais entidades, ou preferência quanto a algum aspecto compartilhado ("O teclado deste telefone é muito melhor do que o do meu telefone antigo"). As opiniões podem ser diretas ("Este suco é muito bom"), ou indiretas ("minha gripe piorou depois que tomei este remédio" – implicando opinião negativa sobre o remédio através do seu efeito sobre a gripe). Finalmente, opiniões explícitas expressam diretamente o sentimento, enquanto que as implícitas o sugerem indiretamente ("Formou-se um vale no colchão que comprei na semana passada").

A maioria dos trabalhos concentra-se em opiniões regulares, diretas e explícitas, por serem mais fáceis de serem tratadas. Este é o tipo de opinião que este trabalho assume como premissa.

É comum o uso de palavras de sentimento (e.g. ótimo, péssimo) para detectar opiniões, mas seu uso não é condição necessária, nem suficiente, para detectar uma opinião e classificar sua polaridade. Primeiro, as palavras de opinião podem ser positivas ou negativas de acordo com o contexto. Por exemplo, na sentença "este smartphone é muito caro", a palavra de sentimento "caro" é negativa, enquanto que em "este amigo me é muito caro", ela é positiva. Segundo, nem toda opinião é expressa com palavras de sentimento (e.g. "comprei este casaco na semana passada, e já está cheio de bolinhas"), ou vice-versa (e.g. "se encontrar um bom livro, vou lê-lo"). A negação é outra questão que deve ser tratada, já que inverte o sentido da opinião ("Este filme não é nada bom").

Um problema enfrentado é a coreferência, onde diferentes tipos de menções designam a uma mesma entidade. Por exemplo, as expressões "Dilma", "Presidenta", "Presidenta Dilma Rousseff" referem-se à mesma pessoa, devendo ser reconhecidas e unificadas. Nesse mesmo contexto, a coreferência também trata da resolução de pronomes, com o objetivo de relacionar um pronome a uma determinada entidade. Por exemplo, no texto "Paris é uma cidade maravilhosa. Ela é um excelente lugar para se visitar. Seus restaurantes são muito reconhecidos", os

pronomes “ela” e “seus” referem-se a Paris. O tratamento da coreferência e dos pronomes é extremamente importante para a análise de sentimentos nos níveis de sentença e de aspectos, já que estes níveis analisam o sentimento de forma isolada (i.e., cada sentença ou opinião), com efeito direto sobre a revocação.

A ironia/sarcasmo é um dos problemas mais difíceis de se tratar (“Ontem vendo o horário político, vi propostas bem novas e inovadoras: melhorar a saúde, educação e emprego. Por que ninguém havia prometido isto antes?”). O uso de sarcasmo é muito comum em alguns domínios, como discussões políticas e esportivas, opiniões sobre arte (filmes, bandas), etc (SARMENTO et al., 2009; TURNEY, 2002; BALAHUR; KOZAREVA; MONTOYO, 2009). Os trabalhos encontrados na literatura para identificação de sarcasmo/ironia fazem uso de artifícios, como: frequência do sinal de exclamação e interrogação, palavras capitalizadas, interjeições (e.g. “ah, oh, yeah”), *emoticons* (e.g. “;-)”) e superlativos (CARVALHO et al., 2009; LIU, 2012).

Neste trabalho, as opiniões expressas nos comentários de notícias serão abordadas como regulares, diretas e explícitas. Os problemas de resolução de pronomes e ironia/sarcasmo não foram abordados nesse trabalho.

2.5 Recursos Básicos de Processamento de Linguagem Natural

A mineração de opiniões tem como objetivo analisar o sentimento expressos em porções de texto de qualquer tamanho (LIU, 2010), portanto diversos recursos de processamento de linguagem natural (PLN) podem ser utilizados para se atingir este objetivo. No restante desta seção serão detalhados e exemplificados recursos de processamento de linguagem natural que podem auxiliar no processo de mineração de opiniões.

2.5.1 Tokenização

Um tokenizador (*tokenizer*) divide um texto bruto em *tokens*. O *token* é o componente mais básico em uma determinada análise. Existem diferentes abordagens de separação de *tokens*, por exemplo, por espaço em branco e Treebank, onde a validade de um *token* é dada de acordo com o objetivo da análise de um documento. Por exemplo, em um *tweet*, exemplos de elementos válidos são *hashtags*, usuário do Twitter que compôs a mensagem, *url*, palavras, etc.

Um tokenizador recebe com entrada uma porção de texto. Considere o seguinte *tweet*

como exemplo:

"AAAAAMEEI a nova música do Justin Bieber!!
;-) #beliebers http://youtu.be/Ys7-6_t7OEQ"

Na abordagem por espaço em branco, o resultado da tokenização do *tweet* é: "AAAAAMEEI", "a", "nova", "música", "do", "Justin, "Bieber!!", ":-D", "#beliebers", "http://youtu.be/Ys7-6_t7OEQ".

Já um tokenizador TreeBank separa um texto em *tokens* utilizando as convenções determinadas por Penn Treebank em diversos corpora públicos anotados sintaticamente e semanticamente[1]. Um tokenizador TreeBank assume que o texto a ser tokenizado está segmentado em sentenças. Uma outra característica deste tokenizador é que ele assume que qualquer ponto final (.) presente no texto analisado está ligado ao *token* que o precede (e.g. abreviações). Então, o resultado da tokenização do mesmo *tweet* acima é: "AAAAMEEI", "a, "nova", "música", "do, "Justin, "Bieber", "!", "!", ":", "-D", "#", "beliebers", "http", ":", "//youtu.be/Ys7-6_t7OEQ".

Especificamente para a mineração de opinião podem ser desenvolvidas heurísticas que não apenas dividem o texto em *tokens*, mas os pré-processam para melhorar sua análise. Exemplos de heurísticas são:

- Isolar emoticons;
- Respeitar marcações específicas de domínios e do Twitter (e.g. @usuário #hashtag);
- Sublinhar marcações importantes (e.g. Justin_Bieber);
- Capturar palavrões mascarados (e.g. p@l*v.r@0);
- Preservar letras maiúsculas quando necessário (e.g. nomes próprios);
- Regularizar o alongamento de palavras (e.g. AAAAAMEEI $\rightarrow$ AMEI);
- Capturar expressões idiomáticas significativas (e.g. "bater as botas").

O resultado da tokenização voltada a análise de sentimento é: "AMEI", "a", "nova", "música", "do", "Justin_Bieber", "!", "!", ":-D", "#beliebers", "http://youtu.be/Ys7-6_t7OEQ".

2.5.2 Marcação Morfológica

A marcação morfológica, também conhecida como *part-of-speech tagging* (POS), de uma determinada palavra dá informações importantes sobre ela e seus vizinhos. No processamento de linguagem natural, a marcação morfológica auxilia na lematização e análise sintática de um texto. As classes morfológicas mais comuns são: substantivos, verbos, adjetivos, prepo-

[1] http://www.cis.upenn.edu/treebank/

sições, advérbios e conjunções.

O processo de marcação morfológica em mineração de opiniões possui grande importância, pois serve, por exemplo, para ajudar a identificar palavras de sentimento, já que a maioria das palavras de sentimento são adjetivos (e.g. lindo, bonito, maravilhoso). Também podem pode existir palavras de sentimento em outras classes morfológicas, como em verbos (e.g. enganar, recomendar) ou substantivos (e.g. aberração).

O processo de marcação morfológica tem como entrada um texto (e.g. "Ir para casa") e um conjunto de rótulos, o qual consiste no conjunto de marcações que serão atribuídas ao texto como saída deste processo. Um exemplo de conjunto de rótulos é mostrado na Figura 2.2, o qual foi definido por Penn Treebank. A saída do processo de marcação morfológica é o texto com sua respectiva *tag* (e.g. "Ir (**VB**) para (**IN**) casa (**NN**)").

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	*and, but, or*	SYM	Symbol	+,%, &
CD	Cardinal number	*one, two, three*	TO	"to"	*to*
DT	Determiner	*a, the*	UH	Interjection	*ah, oops*
EX	Existential 'there'	*there*	VB	Verb, base form	*eat*
FW	Foreign word	*mea culpa*	VBD	Verb, past tense	*ate*
IN	Preposition/sub-conj	*of, in, by*	VBG	Verb, gerund	*eating*
JJ	Adjective	*yellow*	VBN	Verb, past participle	*eaten*
JJR	Adj., comparative	*bigger*	VBP	Verb, non-3sg pres	*eat*
JJS	Adj., superlative	*wildest*	VBZ	Verb, 3sg pres	*eats*
LS	List item marker	*1, 2, One*	WDT	Wh-determiner	*which, that*
MD	Modal	*can, should*	WP	Wh-pronoun	*what, who*
NN	Noun, sing. or mass	*llama*	WP$	Possessive wh-	*whose*
NNS	Noun, plural	*llamas*	WRB	Wh-adverb	*how, where*
NNP	Proper noun, singular	*IBM*	$	Dollar sign	$
NNPS	Proper noun, plural	*Carolinas*	#	Pound sign	#
PDT	Predeterminer	*all, both*	"	Left quote	(' or ")
POS	Possessive ending	*'s*	"	Right quote	(' or ")
PP	Personal pronoun	*I, you, he*	(	Left parenthesis	([, (, {, <)
PP$	Possessive pronoun	*your, one's*	)	Right parenthesis	( ],), }, >)
RB	Adverb	*quickly, never*	,	Comma	,
RBR	Adverb, comparative	*faster*	.	Sentence-final punc	(. ! ?)
RBS	Adverb, superlative	*fastest*	:	Mid-sentence punc	(: ; ... – -)
RP	Particle	*up, off*			

Figura 2.2 – Exemplo de um conjunto de rótulos (Extraído de (JURAFSKY; MARTIN, 2009)).

Existem diversas ferramentas disponíveis que realizam a marcação morfológica de um texto. Um exemplo de ferramenta é o Natural Language Toolkit (NLTK)[2], o qual realiza este processo para diversas línguas, como o inglês, português e espanhol. Também são exemplos de ferramentas o Stanford Log-linear Part-Of-Speech Tagger (TOUTANOVA et al., 2003) e o TreeTagger (SCHMID, 1994), os quais estão disponíveis para diversas línguas, onde o último também está disponível para português.

[1]http://nltk.org

2.5.3 *Stemming* e Lematização

Para se entender o processo de *stemming* de um texto deve-se primeiramente entender o que é um *stem*. O *stem* ou radical é a parte que sobra da palavra após a remoção de seu afixo. Por exemplo, o *stem* das palavras bonita, bonitas, boniteza, bonito e bonitos, é "bonit". É importante ressaltar que um *stem* não precisa ser necessariamente uma palavra válida de uma determinada língua, contudo, precisa captar o significado original da palavra.

A utilidade de um *stemmer* (i.e. aplicação que realiza o *stemming*) é reduzir as formas variantes de diversas palavras a um único radical, de forma que se possa unificar palavras com a mesma semântica.

Uma desvantagem do processo de *stemming* é que ele pode introduzir ruídos nos dados previamente inexistentes. Na mineração de opiniões, um *stemmer* pode levar um classificador de sentimento a errar a polaridade de uma palavra de sentimento. Por exemplo, a palavra "extravagância" possui polaridade positiva, enquanto a palavra "extravagante" possui polaridade negativa. O *stem* de ambas as palavras é "extravag", no qual não se pode desambiguar a polaridade, confundindo o classificador.

Existem poucos *stemmers* disponíveis para a língua portuguesa. Um exemplo é o RSLP, o qual remove sufixos baseando-se em regras e exceções (ORENGO; HUYCK, 2001). Já para a língua inglesa existem mais opções, tais como o Porter[3] e o Lancaster (PAICE, 1990). Um *stemmer* recebe como entrada uma porção de texto (e.g. "Voltando do trabalho encontrei meu amigo"), e obtém como saída o texto processado (e.g. "volt do trabalh encontr meu amig").

Assim como o *stemming*, a lematização tem como objetivo reduzir formas flexionadas de uma determinada palavra. No entanto, a lematização de uma palavra envolve o uso de análise de vocabulário e morfológica, e tem como objetivo retornar a palavra a sua base, como em uma entrada de dicionário (i.e., lema). Portanto, considerando o mesmo exemplo, o *lema* das palavras bonita, bonitas, boniteza, bonito e bonitos, é "bonito". Novamente, o NLTK possui diversos lematizadores para diferentes línguas, incluindo português.

2.6 Palavras

Palavras é a ferramenta de processamento de linguagem natural mais popular para a língua portuguesa. O Palavras é um analisador morfológico baseado em dicionário e gramática (BICK, 2000), e foi desenvolvido para diversas aplicações, como anotação de corpora,

[2]http://tartarus.org/martin/PorterStemmer/

ensino de gramática e tradução automática. Além disso, o Palavras também trata relações de dependência de uma forma inovadora, permitindo a transformação da notação utilizada em uma estrutura de árvore.

O Palavras recebe como entrada uma porção de texto a ser analisada (e.g. "A bateria deste celular dura muito"). Após isto o texto é analisado e passa por uma série de processos, como tokenização, *Part-of-speech tagging*, lematização, analisador de dependências, etc. A saída do Palavras é composta do texto, seu lema, informação morfológica (*part-of-speech*, inflexões gramaticais (e.g. número, gênero, pessoa), marcações sintáticas (e.g. predicado, sujeito), entre outros.

A saída também contém informações de ligações de dependência do texto analisado. A partir destas ligações é possível derivar a estrutura de árvore anteriormente mencionada. Um exemplo de árvore construída a partida da saída do Palavras é mostrado na Figura 2.3. A partir desta árvore é possível analisar as dependências de palavras de uma oração, e por exemplo, encontrar o alvo de sentimento de uma determinada palavra.

Um dos pontos negativos do Palavras é que o texto de entrada deve ser corretamente estruturado gramaticalmente e também não conter erros tipográficos, pois a ferramenta é muito sensível a estes erros. Outro ponto, é que o dicionário interno do Palavras é genérico, portanto, vocabulário dependente de domínio pode não existir. Por exemplo, a palavra "celular" dentro do Palavras é relativa a uma célula (unidade estrutural de seres vivos) e não ao dispositivo eletrônico.

Em resumo, o Palavras pode auxiliar nos seguintes desafios no processamento de linguagem natural: identificação de palavras de sentimento; identificação de negações e intensificações de palavras de sentimento; co-referência de sujeitos observados, e; auxiliar na identificação do alvo do sentimento de uma determinada frase.

Neste trabalho, o Palavras foi utilizado para encontrar o sujeito e o predicado de uma determinada sentença, de modo que se possa encontrar o alvo do sentimento. No entanto, como o texto utilizado é conteúdo gerado por usuários, não foram obtidos bons resultados em sua utilização.

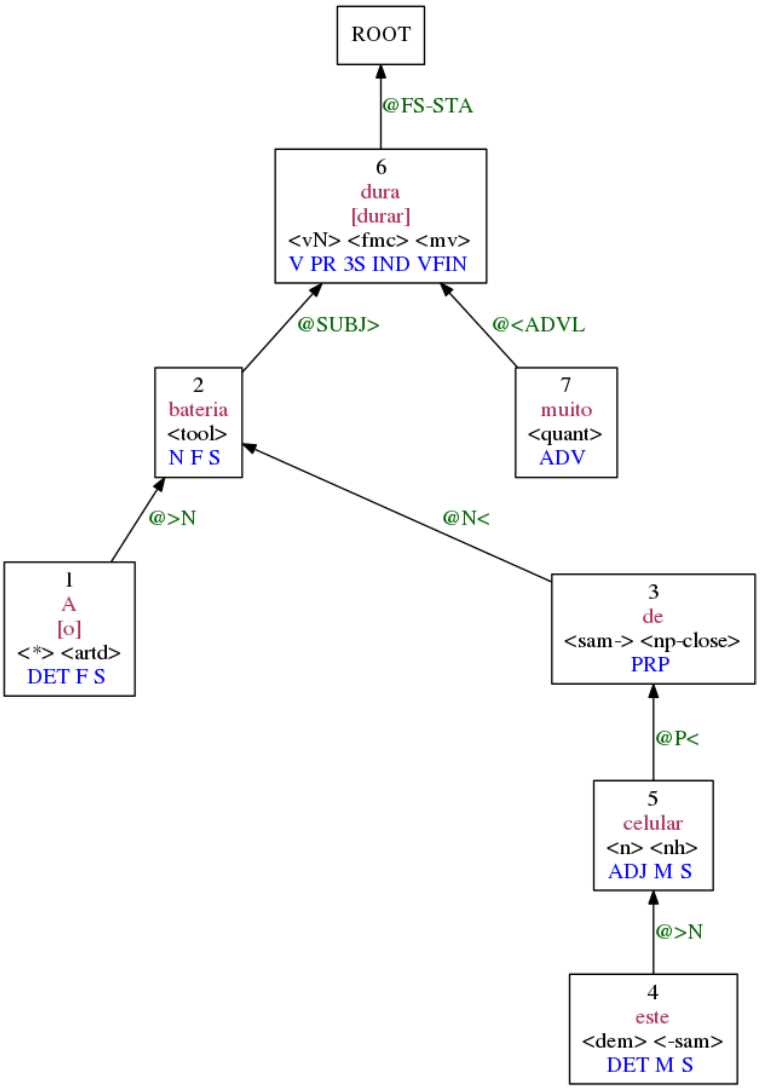

Figura 2.3 – Exemplo de árvore montada a partir da saída do Palavras. Texto de entrada: "A bateria deste celular dura muito".

2.7 Etapas da Mineração da Opinião

A mineração de opiniões pode ser caracterizada em termos de três grandes tarefas (TSYTSARAU; PALPANAS, 2012): a) identificar (tópicos, sentenças opinativas), b) classificar a polaridade do sentimento, e c) sumarizar. Este processo é esboçado na Figura 2.4.

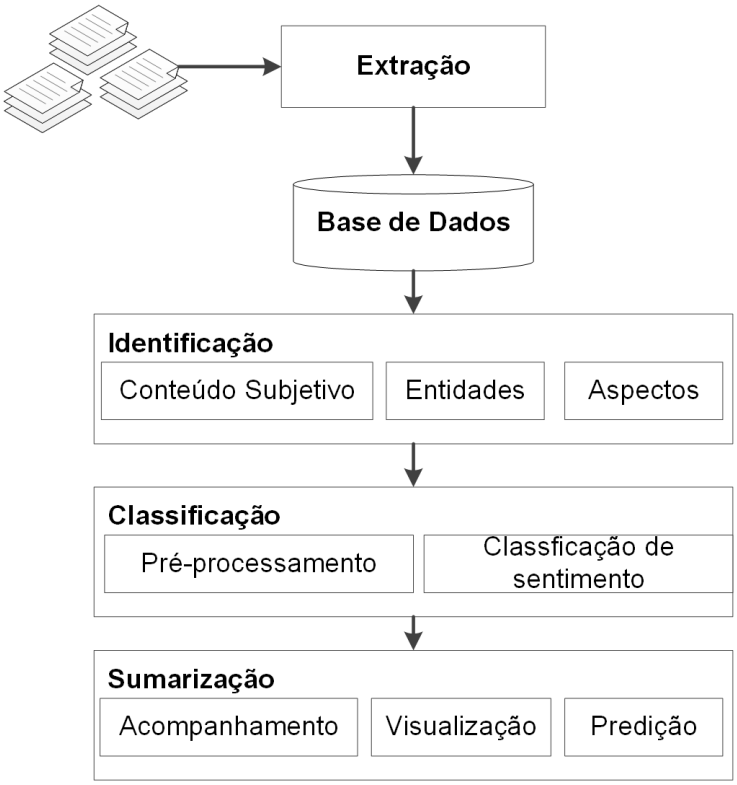

Figura 2.4 – Etapas da Mineração de Opiniões (Extraído de (BECKER; TUMITAN, 2013)).

2.7.1 Identificação

Dado um conjunto de textos extraídos de alguma fonte (e.g. jornais, redes sociais, plataformas de revisão de produtos/serviços), a etapa de *identificação* consiste em encontrar os tópicos existentes (e possivelmente seus aspectos) e possivelmente associá-los com o respectivo conteúdo subjetivo. A forma de identificar as entidades, aspectos e sentimento são dependentes da granularidade escolhida para análise (ver Seção 2.3) e os algoritmos utilizados podem ser distintos daqueles propostos para recuperação de documentos opinativos (PANG; LEE, 2008; TSYTSARAU; PALPANAS, 2012).

A complexidade da identificação do alvo da opinião depende em muito da mídia considerada, e de seu grau de estruturação. A aplicação mais frequente em mineração de opiniões é a de revisão de produtos e serviços, porque o alvo pode ser mais facilmente identificado. Assume-se que todo o documento refere-se a uma única entidade, o alvo da revisão, sendo que o desafio está em identificar os aspectos desta entidade, se a análise for nesta granularidade.

Já em jornais, blogs ou *posts*, não se conhece *a priori* as entidades envolvidas, podendo inclusive envolver muitas entidades na mesma porção de texto. Na situação mais simples, pode-se restringir a identificação a entidades pré-definidas, como a busca de celebridades, atletas, políticos ou marcas. Um dos problemas neste caso é resolver os problemas de coreferência e

de resolução de pronomes, já mencionados na Seção 2.4. Em mídias sociais, a coreferência pode ser um problema acentuado, pois as menções podem ser muito informais (apelidos, gírias com significado local ou temporal, *hashtags*, etc.). Se a identificação não for direcionada a alvos pré-definidos, pode-se ainda utilizar técnicas de identificação de entidades nomeadas da recuperação de informações (SARAWAGI, 2008; AGGARWAL; ZHAI, 2012).

Finalmente, esta tarefa pode envolver também o discernimento entre conteúdo ou sentenças com ou sem opinião, visando melhorar os resultados da próxima etapa. Isto é bastante comum quando o nível de análise é de granularidade menor. O critério utilizado para determinar o conteúdo de opinião é quase sempre a existência de palavras de sentimento (e.g. "Eu recomendo este filme") ou de classes de palavras candidatas a expressar sentimento (e.g. adjetivos).

Para a identificação das entidades foram utilizadas neste trabalho, expressões regulares baseadas nas entidades observadas. Como resultado obteve-se uma compilação de variações dos nomes usados para referenciar as entidades. A identificação de conteúdo subjetivo foi feita durante fase de classificação de polaridade.

2.7.2 Classificação da Polaridade

O problema de *classificação de sentimento*, também denominado *classificação de polaridade*, é frequentemente um problema de classificação binário: *positivo* ou *negativo*. No entanto, classes adicionais podem ser consideradas para que a análise seja mais robusta ou para aumentar o nível de detalhe dos resultados. Assim, estas classes podem ser desdobradas em classificações com diferentes graus de intensidade (e.g. muitoPositivo, moderadamentePositivo) ou em intervalos numéricos representando um grau de intensidade. Neste último caso, a divisão do sentimento está relacionada à capacidade de definir algum limiar para distinguir os níveis de sentimento.

Outra abordagem é considerar a categoria neutra, que engloba textos sem uma tendência clara quanto a sua polaridade ou simplesmente sem sentimento. Neste último caso, é a etapa de classificação de polaridade que tem como responsabilidade identificar textos sem sentimento de acordo com suas propriedades.

Para a classificação da polaridade, diferentes abordagens são propostas na literatura, as quais são discutidas com maiores detalhes na Seção 2.8. Cada técnica pode necessitar de operações de pré-processamento e transformação específicas, tais como reconhecimento de construtos sintáticos, reconhecimento de n-gramas, extração de atributos discriminastes, eliminação de termos irrelevantes, transformação em vetor de termos, etc.

Independente da abordagem empregada, a classificação da polaridade não é um problema trivial. Entre os principais desafios estão:

- o uso de palavras de sentimento pode ser enganoso, como discutido na Seção 2.4;
- a polaridade de alguns termos é dependente de contexto;
- muitos domínios são caracterizados pelo uso frequente de ironias ou sarcasmo, onde o sentido implícito é exatamente oposto ao sentimento expresso explicitamente. Outros domínios (e.g. debates políticos, críticas culturais) estabelecem uma opinião positiva por oposição a uma argumentação negativa (ou vice-versa) (BALAHUR; KOZAREVA; MONTOYO, 2009; PANG; LEE; VAITHYANATHAN, 2002; TURNEY, 2002);
- a opinião pode depender do observador. Por exemplo, a opinião representada na sentença "Um bom momento para as ações da Petrobras" é positiva para quem detém este tipo de ação, mas pode ser péssima para quem deixou de investir nelas;
- a polaridade de conteúdo subjetivo nem sempre é objeto de consenso. Por exemplo, em anotações feitas por humanos, dificilmente o consenso é maior que 75% (BRUCE; WIEBE, 1999; KU; LIANG; CHEN, 2006; WIEBE; WILSON; CARDIE, 2005; PANG; LEE; VAITHYANATHAN, 2002);
- a classificação é bastante dependente da extração das *features* do texto, a qual deve lidar com as várias questões da língua natural já discutidas na Seção 2.4.

2.7.3 Sumarização

Para poder identificar opinião média ou prevalecente de um grupo de pessoas sobre um determinado tópico/entidade, a opinião expressa por uma única pessoa não é suficiente, sendo necessário analisar uma grande quantidade de opiniões (LIU, 2012). É necessário a criação de métricas e sumários que quantifiquem a diversidade de opiniões encontradas a respeito um mesmo alvo. Este é o objetivo desta etapa, onde são criadas métricas que representem o sentimento geral, as quais podem ser visualizadas ou servir de entrada para outras aplicações.

O sentimento sumarizado também pode ser utilizado para diversas aplicações, como prever eleições (TUMASJAN et al., 2010), comportamento da bolsa de valores (BOLLEN; MAO; ZENG, 2011), arrecadação de bilheterias de filmes (ASUR; HUBERMAN, 2010), definição de preços (ARCHAK; GHOSE; IPEIROTIS, 2007), etc. No entanto, o sentimento puro (positivo ou negativo) não pode refletir de maneira correta o contexto analisado. Portanto, é importante criar métricas para representar o sentimento em relação ao alvo. Boa parte dos trabalhos na

área utilizam a média do sentimento ou a razão entre o sentimento positivo e negativo. Em certos casos, a predição pode ser feita somente com base na quantidade de menções às entidades, independente do sentimento sobre elas (e.g. (TUMASJAN et al., 2010)).

Outra forma de sumarização, comum em aplicações que extraem de mídias sociais o sentimento do público em geral sobre uma determinada entidade (e.g. uma marca, produto, político, celebridade), é apresentar o sentimento na forma de relógios, ou associá-lo a informações temporais ou geográficas. Normalmente este tipo de mídia reflete o que as pessoas pensam sobre o alvo, dado algum evento. Por exemplo, o lançamento de um novo produto terá impacto nas redes sociais, que expressarão reações a esse acontecimento através de posts, comentários, *tweets*, endossos, etc.

Neste trabalho, as métricas de sumarização são derivadas de combinações de totais de sentimento positivo e negativo em relação aos candidatos observados, assim como menções a eles. Também foram propostas métricas baseadas em explosões de expressão de sentimento. Todas as métricas propostas servem como entrada (atributos discriminantes) para o treinamento de um algoritmo de classificação preditivo.

2.8 Abordagens de Classificação de Polaridade

As abordagens de classificação podem ser divididas em quatro grandes grupos: a) léxicas, com o uso de dicionários de sentimentos; b) aprendizado de máquina, com o uso predominante de técnicas de classificação ou de regressão; c) estatísticas, que valem-se de técnicas para avaliar a coocorrência de termos, e d) semânticas, que definem a polaridade de palavras em função de sua proximidade semântica com outras de polaridade conhecidas. Técnicas destas diferentes abordagens podem ser combinadas para melhoria de resultados. Uma revisão sistemática descrita em (TSYTSARAU; PALPANAS, 2012) aponta uma predominância das duas primeiras abordagens, sem que nenhuma técnica se sobressaia em termos de desempenho. Neste trabalho foram utilizadas apenas as duas primeiras abordagens, portanto, apenas estas serão discutidas a seguir.

2.8.1 Abordagem Baseada em Dicionário

A abordagem baseada em *dicionário* é também denominada *léxica* ou *linguística*. O aspecto central desta abordagem é o uso de léxicos (dicionários) de sentimentos, que são com-

pilações de palavras ou expressões de sentimento associadas à respectiva polaridade.

Um dos métodos mais utilizados na abordagem linguística é o da coocorrência entre alvo e sentimento, que não leva em consideração nem a ordem dos termos dentro de um documento (*bag-of-words*), nem suas relações léxico-sintáticas. Para a classificação do sentimento em um texto, basta que exista uma palavra de sentimento, onde sua polaridade é dada por um léxico de sentimentos. Esse método é extensamente empregado para o atrelamento de um sentimento a uma entidade em uma sentença. Por exemplo, na sentença "o iPhone é muito bom", a polaridade positiva da palavra "bom" é associada à entidade "iPhone". O método por coocorrência apresenta bons resultados quando o nível de análise textual é de granularidade pequena, pois a palavra detentora do sentimento está próxima à entidade que qualifica. Sendo assim, este método é usualmente utilizado em análises de nível de sentença, cláusula ou até em documentos com poucos caracteres, como um *tweet*.

Quando aplicada em nível de maior granularidade, estabelece-se algum tipo de média sobre as palavras de sentimento encontradas. A Equação 2.1 (extraída de (TSYTSARAU; PALPANAS, 2012)) mostra uma função genérica de determinação de polaridade em um documento D, onde S_w representa a polaridade de uma palavra w em um dicionário. A agregação pode levar em conta funções de peso e de modificação. A função $peso()$ pode ser, por exemplo, alguma medida de distância entre a palavra de sentimento e o alvo, ou de importância da palavra no texto (e.g. frequência). A função $modificador()$ pode ser usada para tratar negações, palavras de intensidade (e.g. muito), etc. Esta função de agregação também pode ser estendida a sentenças, cujas cláusulas podem combinar diferentes palavras de sentimento.

$$\text{S(D)} = \frac{\sum_{w \in D} \; S_w \cdot peso(w) \cdot modificador(w)}{\sum peso(w)} \tag{2.1}$$

Existem métodos linguísticos mais complexos, como a utilização de *parsers* linguísticos, que têm como propósito analisar o texto e aumentar a qualidade da classificação com base em informações morfossintáticas ali presentes (e.g. sujeito, predicado, dependências, funções sintáticas, etc.). No entanto, ferramentas de processamento de linguagem natural são em sua maioria restritas a determinado idioma. Recursos para a língua portuguesa são escassos, quando comparada à língua inglesa, situação esta comum a outras línguas.

A composição básica de um léxico de sentimento é a palavra de sentimento com suas possíveis flexões (e.g. bonito, bonita, bonitos), sua respectiva polaridade expressa como uma categoria ou como um valor em uma escala. Muitos dicionários possuem adicionalmente: o lema e o *stem* de cada entrada; a marcação morfológica (POS); e o alvo do sentimento (predicado ou sujeito). A maioria dos léxicos existentes são dependentes de idioma e foram feitos estrita-

mente para a língua inglesa, como General Inquirer (STONE; DUNPHY; SMITH, 1966), OpinionFinder (WIEBE; WILSON; CARDIE, 2005), SentiWordNet (BACCIANELLA; ESULI; SEBASTIANI, 2010) e WordNetAffect (STRAPPARAVA; VALITUTTI, 2004). Já para a língua portuguesa estão disponíveis o OpLexicon (SOUZA et al., 2011), e o SentiLex-PT (SILVA; CARVALHO; SARMENTO, 2012), sendo o primeiro para português do Brasil e o último, para português de Portugal. A Tabela 2.1 mostra uma comparação entre os léxicos mencionados. Outro exemplo é o Linguistic Inquiry and Word Counts (LIWC) (PENNEBAKER et al., 2007), que é um software de análise de texto desenvolvido para avaliar os componentes estruturais, cognitivos e emocionais de amostras de texto, sendo que essa análise pode ser feita em vários idiomas disponibilizados na ferramenta.

Tabela 2.1 – Tabela comparativa de léxicos de sentimentos.

Dicionário	**Pos**	**Neg**	**PoS**	**Stem**	**Lema**	**Idioma**
General Inquirer	1.915	2.291	S	N	N	Inglês
OpinionFinder	2.718	4.912	S	S	N	Inglês
OpLexicon	8.675	14.469	S	N	N	Português
SentiLex-PT	82.347 entradas		S	N	S	Português
SentiWordNet	117.659 entradas		S	N	S	Inglês
WordNetAffect	7.661 entradas		S	S	S	Inglês

A maioria dos dicionários disponíveis são genéricos, ou seja, auxiliam na tarefa de classificação, independentemente do domínio dos textos sendo considerados. Entretanto, os melhores resultados obtidos na tarefa de classificação foram baseados em dicionários dependentes de contextos (HU; LIU, 2004), criados a partir de palavras semente e expandidos utilizando o WordNet (MILLER, 1995) ou tesauros. Esta abordagem também é classificada como *semântica* (TSYTSARAU; PALPANAS, 2012). Finalmente, léxicos são de pouca valia quando considerados em textos gerados em mídias informais (e.g. redes sociais, *tweets*), onde expressões regionais, gírias, abreviaturas típicas da internet, etc., são fartamente empregados.

Neste trabalho foi utilizado o dicionário de sentimento SentiLex-PT, pois é o mais completo dicionário para a língua portuguesa. Exemplos de entradas para este dicionário são mostradas na Figura 2.5. Algumas características deste dicionário podem ser vistas na Tabela 2.2. O SentiLex-PT possui 7.014 lemas e 82.347 formas flexionadas, onde cada entrada do dicionário possui a polaridade diferentes caso o alvo do sentimento seja o predicado ou o sujeito da sentença, podendo esta ser positiva, neutra ou negativa.

Tabela 2.2 – Conteúdo do SentiLex-PT em detalhes.

Categorial Gramatical	Lemas	Formas Flexionadas
Adjetivos	4.779	16.863
Substantivos	1.081	1.280
Verbos	489	29.504
Expressões Idiomáticas	666	34.700

```
aberração,aberração.PoS=N;FLEX=fs;TG=HUM:N0;POL:N0=-1;ANOT=MAN
aberrante,aberrante.PoS=Adj;FLEX=fs|ms;TG=HUM:N0;POL:N0=-1;ANOT=MAN
aberrantes,aberrante.PoS=Adj;FLEX=fp|mp;TG=HUM:N0;POL:N0=-1;ANOT=MAN
bonita,bonito.PoS=Adj;FLEX=fs;TG=HUM:N0;POL:N0=1;ANOT=MAN
bonitas,bonito.PoS=Adj;FLEX=fp;TG=HUM:N0;POL:N0=1;ANOT=MAN
boniteza,boniteza.PoS=N;FLEX=fs;TG=HUM:N0;POL:N0=1;ANOT=MAN
engole em seco,engolir em seco.PoS=IDIOM;Flex=Y2s|P2s|P4s|P3s;TG=HUM:N0;POL:N0=-1;ANOT=MAN
engolem em seco,engolir em seco.PoS=IDIOM;Flex=P4p|P3p;TG=HUM:N0;POL:N0=-1;ANOT=MAN
engoles em seco,engolir em seco.PoS=IDIOM;Flex=P2s;TG=HUM:N0;POL:N0=-1;ANOT=MAN
```

Figura 2.5 – Exemplo de entradas do dicionário de sentimento SentiLex-PT.

2.8.2 Abordagem Baseada em Aprendizado de Máquina

O objetivo principal das técnicas de aprendizado de máquina é descobrir automaticamente regras gerais em grandes conjuntos de dados, que permitam extrair informações implicitamente representadas. De modo geral, as técnicas de aprendizado de máquina podem ser divididas em dois tipos: aprendizado supervisionado e aprendizado não supervisionado (TAN; STEINBACH; KUMAR, 2006).

Na área de mineração de opiniões, nota-se um predomínio do uso de métodos supervisionados de aprendizagem, mais especificamente, *classificação* e *regressão*. Neste contexto, o problema de classificação é dividido em dois passos: (1) aprender um modelo de classificação sobre um corpus de treinamento previamente rotulado com as classes consideradas (e.g. positivo, negativo); e (2) prever a polaridade de novas porções de texto com base no modelo resultante. Dentre os algoritmos de classificação mais usados nesta área estão Support Vector Machine, Naïve Bayes, Maximum Entropy e algoritmos baseados em redes neurais (DAVE; LAWRENCE; PENNOCK, 2003; BOLLEN; MAO; ZENG, 2011; PANG; LEE; VAITHYANATHAN, 2002; PANG; LEE, 2008; TSYTSARAU; PALPANAS, 2012).

A qualidade do modelo preditivo resultante da etapa de aprendizagem é medida em termos de métricas como *acurácia* (capacidade do modelo de prever corretamente), *precisão* (número de instâncias previstas corretamente em uma dada classe), *revocação* (número de instâncias de uma dada classe previstas na classe correta), ou *medida-F* (média ponderada entre precisão e revocação). Alguns trabalhos obtêm precisões muito maiores na classificação da polaridade negativa, do que na positiva (SARMENTO et al., 2009; BALAHUR; KOZAREVA;

MONTOYO, 2009; TUMITAN; BECKER, 2013). Além das dificuldades próprias ao domínio, esta situação pode ser explicada pela dificuldade em tratar ironia e sarcasmo.

Os dados de treino para a classificação/regressão correspondem a um conjunto de instâncias caracterizadas por atributos. O rótulo é denominado atributo *alvo*, enquanto que os demais são designados como atributos discriminantes (*features*). O atributo alvo na classificação é discreto, enquanto que na regressão é numérico. Em termos de pré-processamento, é necessário extrair de cada porção de texto analisada, os atributos discriminantes relevantes para a tarefa de classificação e representá-las na forma de um vetor de termos.

Os tipos de atributos discriminantes mais frequentemente considerados são (PANG; LEE; VAITHYANATHAN, 2002; LIU, 2012; TSYTSARAU; PALPANAS, 2012):

- Palavras de sentimento: somente as palavras de sentimento de um corpus são utilizadas como atributo discriminante. Não existe ordem entre os termos, e estes são caracterizados de forma binária, isto é, presente ou ausente no texto;
- Termos e sua frequência: são usados n-gramas (de sentimento ou não), junto com sua frequência absoluta ou relativa (e.g. TF-IDF), como peso dos termos;
- *Part-of-Speech* (POS): as classes morfológicas das palavras também podem ser usadas, em complementação às palavras de sentimento ou termos;
- Dependência sintática: as dependências sintáticas entre as palavras podem ser utilizadas, com o intuito de auxiliar na definição do alvo e fonte do sentimento.

Uma das grandes limitações no uso de aprendizado supervisionado para definição de polaridade é a necessidade de dados rotulados para treino. O desempenho destes métodos é afetado não somente pela quantidade, mas igualmente pela qualidade dos dados de treino disponíveis. Ainda, cada conjunto de treino é fortemente vinculado ao seu domínio. Nos trabalhos envolvendo revisões de produto, a classificação dada pelos usuários na forma de notas ou estrelas é utilizada como o rótulo para o texto correspondente (PANG; LEE; VAITHYANATHAN, 2002; TURNEY, 2002). Tal facilidade não está disponível para outros domínios, implicando a necessidade de anotações manuais, as quais são trabalhosas e com alto teor de subjetividade. Nestes casos, as alternativas costumam ser os métodos léxicos, ou probabilísticos/semânticos, discutidos na próxima seção.

2.9 Anotação do Corpora

Existem trabalhos específicos para anotação de texto subjetivo (BRUCE; WIEBE, 1999; YU; HATZIVASSILOGLOU, 2003; WIEBE et al., 2004). O processo de anotação adotado no presente trabalho (ver Seção 5.2) foi baseado no esquema proposto em (WIEBE; WILSON; CARDIE, 2005), que pode ser divido em: a) seleção de corpus de treinamento de anotação, b) treinamento dos anotadores e c) anotação do corpus.

a) seleção de corpus de treinamento de anotação: primeiramente, os autores do trabalho definiram que a granularidade de anotação seria em nível de palavra e de sentença, e que cada porção de texto anotada conteria os seguintes elemento:

- *âncora do texto:* ponteiro para a porção de texto que foi anotada;
- *fonte:* quem expressou a opinião;
- *alvo:* o alvo da opinião expressa pela fonte;
- *propriedades:* propriedades quem envolvem intensidade do sentimento, significado, tipo de atitude, etc.

Para a criação de um *gold-standard*, os autores do trabalho selecionaram porções do corpus no qual o conteúdo subjetivo é explicito e não existe dúvida quanto a sua polaridade. Para verificar isto, todos os autores do trabalho concordaram quanto à polaridade do texto, e sua respectiva anotação.

b) treinamento dos anotadores: este *gold-standard* foi utilizado para treinar e orientar os anotadores. Os anotadores foram incentivados a solucionar suas dúvidas quanto à anotação, e sobre as possíveis diferenças de suas anotações e do *gold-standard*. Caso o anotador superasse esta fase de treinamento, o mesmo estava apto para anotar o restante do corpus.

Outra instrução dada aos anotadores é que eles deviam ser consistentes em suas anotações, não removendo as palavras do contexto do qual elas aparecem e que não realizem suposições de significados que pudessem existir. Também deviam se basear no que estava explicitamente escrito. Instruções semelhantes também foram dadas aos anotadores em (SARMENTO et al., 2009), pois ao se fazer isto, as opiniões dos anotadores não terão impacto em suas anotações.

c) anotação do corpus: uma vez o texto anotado, é avaliado o grau de concordância de anotação entre os anotadores, no qual foram utilizados três anotadores sem nenhuma experiência prévia em anotação de corpus. Caso houvesse consenso entre dois ou mais anotadores quanto a anotação, o texto anotado é mantido.

Uma alternativa para avaliação da concordância é utilizar a medida estatística Kappa de Cohen (CARLETTA, 1996), a qual ajusta o efeito de concordâncias que aconteceram apenas por chance nas proporções observadas. Trata-se, portanto, de uma medida mais robusta que o cálculo das porcentagens de concordâncias entre os anotadores.

2.10 Abordagens de Previsão

Esta seção irá discutir abordagens de previsão de futuro, no qual são divididas em três grupos: a) *regressão* ou *predição numérica*, do qual será explicado o método de regressão linear; b) *classificação*, e; c) *séries temporais*. A principal diferença entre as duas primeiras abordagens é que na classificação o valor a ser predito é uma classe (e.g. aumentar, diminuir, alto, baixo). Já na regressão o valor a ser predito é um valor numérico (e.g. valor de uma ação, lucro obtido) (BRAMER, 2013). A previsão utilizando séries temporais é amplamente usada para analisar mecanismos do mercado econômicos. Combinações destas técnicas podem ser utilizadas visando melhoria de resultados (ASUR; HUBERMAN, 2010; BOLLEN; MAO; ZENG, 2011). Em mineração de opiniões, existe uma maior predominância de abordagens que se utilizam regressão e classificação. Neste trabalho foi utilizada uma técnica de previsão baseada em classificação para prever variações de pesquisas públicas de intenção de voto, as classes alvos para serem preditas foram: “aumentou”, “diminuiu”, “inalterada”.

2.10.1 Regressão Linear

A regressão linear é um método estatístico de previsão baseado na premissa de que é possível estabelecer o comportamento sobre uma determinada variável a partir da relação do comportamento de uma outra variável cujo comportamento já se conhece (REIS, 2000). Por exemplo, pode-se estudar a relação entre o nível de escolaridade e a renda média de uma pessoa, ou seja, confirmar se maior nível de escolaridade de uma pessoa, maior é seu salário. A relação deste tipo entre duas variáveis é descrita matematicamente na Fórmula 2.2, onde:

$$\mathrm{Y} = \alpha + \beta X + \varepsilon \tag{2.2}$$

- Y é a variável explicada ou resultado (i.e. dependente);
- X é a variável explicativa ou preditora (i.e. independente);
- ε é uma variável residual que inclui outros fatores explicativos de y não incluídos em x,

bem como erros de medição ou ruído nos dados;

- α é a intercepção da reta com o eixo vertical e β o declive da reta. Ambos os valores α e β são constantes.

Uma melhor forma para se analisar os dados observados é através de uma representação gráfica. No eixo horizontal são representados os valores da variável independente X. Já no eixo vertical são representados os valores da variável Y. Os dados são então representados por pares ordenados (X, Y). Após isto, é ajustada uma reta aos dados observados com o propósito de encontrar uma relação entre eles.

Um dos métodos de ajuste de reta aos dados observado é o dos mínimos quadrados. O método dos mínimos quadrados permite um ajustamento de reta de forma que seja minimizado o somatório do quadrado das distâncias entres os valores observados e a reta ajustada. Um gráfico de dispersão de uma relação linear é exemplificado na Figura 2.6. Com a reta de regressão é possível prever o comportamento da variável dependente (Y) com base em valores conhecidos da variável explicativa (X).

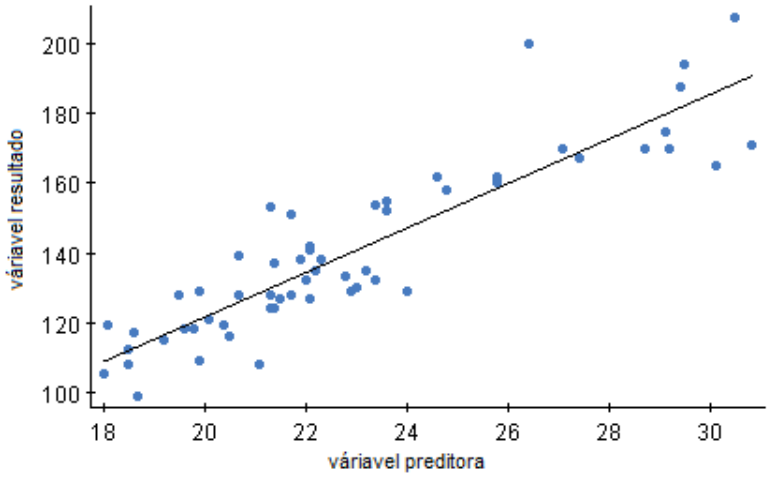

Figura 2.6 – Exemplo de gráfico de dispersão de uma relação linear positiva.

O modelo de previsão baseado em regressão linear dever ser utilizado quando espera-se que grande parte da variação da variável resultado seja explicada pela variável preditora, deste modo o modelo pode ser utilizado para obter valores de Y correspondentes a valores de X que não estavam entre os dados de análise.

2.10.2 Séries Temporais

Uma *série temporal* ou *série de tempo* é um conjunto de observações, onde cada uma das observações é registrada em um tempo específico (BROCKWELL; DAVIS, 2002). Formalmente, uma série de tempo pode ser definida por $X = \{x_t : t \in T\}$, onde x_t são as observações, e $T = [t_1, \ldots, t_n]$ o intervalo temporal. Existem dois tipos de séries temporais: séries temporais de tempo discreto e séries temporais de tempo continuo. No primeiro tipo de série temporal as observações são registradas em intervalos de tempo fixos. Já no último tipo, as observações são registradas através de algum intervalo de tempo, por exemplo, $T = [t_1, t_2]$. São exemplos de séries temporais: população de um país, valores de fechamentos diários do IBOVESPA, salários de uma pessoa em um ano.

Para se utilizar uma série temporal para previsão, é necessário identificar padrões não aleatórios na série temporal da variável observada. Existem quatro componentes padrões básicos em uma série temporal (KIRCHGÄSSNER; WOLTERS; HASSLER, 2012):

- **tendência:** desenvolvimento a longo prazo;
- **variações cíclicas ou ciclo:** componente cíclico com períodos superiores a um ano.
- **variações sazonais ou sazonalidade:** componente que contém flutuações inferiores a um ano;
- **variações irregulares:** componente que contêm todas as variações que não pertencem a uma tendência, ciclo ou sazonalidade.

O primeiro passo para se utilizar uma série temporal para previsão do futuro é utilizar os dados históricos desta série para criar um modelo matemático representativo do processo e, consequentemente, utilizar este modelo para realizar previsões. Durante o processo de análise, existem algumas combinações de padrões que são mais frequentes: apenas variações irregulares; apenas tendência e variações irregulares, e; apenas variações sazonais e irregulares. É importante ressaltar que pode existir qualquer outro tipo de combinação entre os componentes citados anteriormente.

No entanto, deve-se decidir como será a equação que relaciona os componentes com a variável observada, onde existem duas opções: o modelo aditivo e o modelo multiplicativo. No modelo aditivo, o valor da série será o resultado da soma dos valores dos componentes mencionados anteriormente. Já modelo multiplicativo os componentes são multiplicados. A Figura 2.7 mostra um exemplo de série temporal aditiva decomposta em tendência, sazonalidade e variações irregulares.

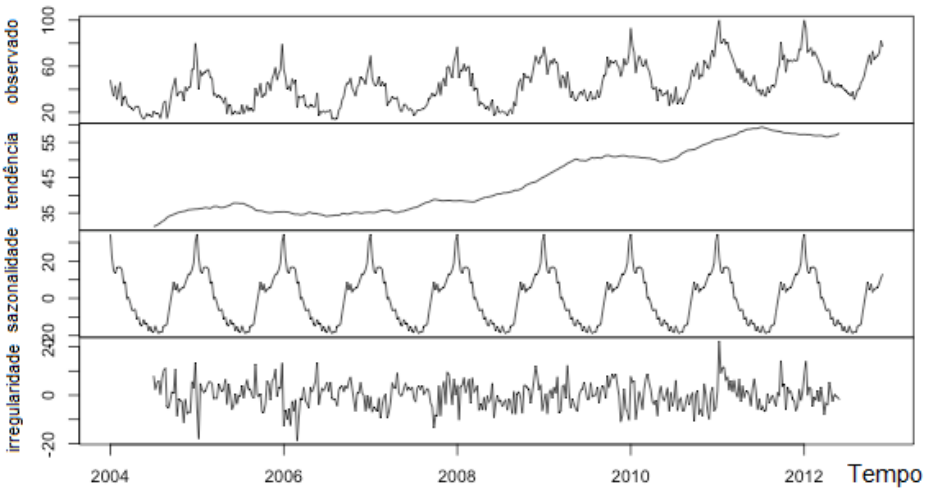

Figura 2.7 – Exemplo de série temporal decomposta no modelo aditivo.

Após o modelo matemático ser criado podem ser utilizadas diversas abordagens para previsão. A abordagem escolhida varia de acordo com o modelo adotado (e.g. variável constante, variável constante e tendência, e regressão por mínimos quadrados). São exemplos de métodos de previsão: média móvel, suavização exponencial, correção a priori, etc. (BROCKWELL; DAVIS, 2002).

3 TRABALHOS RELACIONADOS

Neste capítulo são apresentados trabalhos relevantes da área de mineração de opiniões, de forma a contextualizar nesta área de pesquisa a contribuição do trabalho proposto. Os trabalhos são segmentados da seguinte forma: mineração de opiniões em notícias, mineração de opiniões para a língua portuguesa, monitoramento de sentimento através do tempo, e previsão utilizando sentimento. Ao final do capítulo é apresentado um comparativo deles em relação ao trabalho proposto.

3.1 Mineração de Opiniões em Notícias

A maioria dos trabalhos em mineração de opiniões são concentrados em revisões de produtos, que possui como objetivo analisar o sentimento público sobre um respectivo produto ou suas respectivas características (DAVE; LAWRENCE; PENNOCK, 2003; HU; LIU, 2004; POPESCU; ETZIONI, 2005). Outro ramo na mineração de opiniões analisa opiniões em redes sociais (e.g. Twitter) sobre celebridades, equipes esportivas, empresas, etc (BERENDT; NAVIGLI, 2006; PAK, 2010; TSYTSARAU; PALPANAS, 2012). Por último, um domínio emergente e com menos trabalho, possui notícias como foco (KU; LIANG; CHEN, 2006; LLOYD; KECHAGIAS; SKIENA, 2005; GODBOLE; SRINIVASAIAH; SKIENA, 2007), que é um dos grandes meios de comunicação onde sentimento e opiniões são encontradas. Páginas de notícias comumente informam interesses de figuras políticas, decisões e ideias, o que é uma excelente maneira de acompanhar figuras políticas, através do sentimento expresso nos comentários sobre as respectivas notícias.

Godbole *et al.* em (GODBOLE; SRINIVASAIAH; SKIENA, 2007) têm como objetivo desenvolver um sistema de análise de sentimentos em notícias e blogs que monitore o sentimento do público geral em relação a determinadas entidades, como pessoas, locais ou marcas. Assume-se que as entidades analisadas possuem características singulares, tais como atletas, celebridades, políticos, criminosos, etc, e que as opiniões emitidas devem ser interpretadas dentro de cada contexto. O método propõe uma forma algorítmica de construir dicionários de sentimentos voltados a cada contexto e métricas para medir o sentimento expresso sobre essas entidades. Os autores ainda fazem uma comparação entre o conteúdo de blogs e jornais para as mesmas entidades, concluindo que o sentimento relacionado a certas entidades pode diferenciar-se em notícias e blogs, devido ao viés do meio de publicação. As métricas propostas nesse trabalho, que são baseadas em totais de sentimento, serviram de inspiração para as

métricas propostas neste presente trabalho, mas até onde sabemos, os atributos discriminantes baseados em explosões de sentimento não foram propostos anteriormente.

3.2 Mineração de Opiniões para a Língua Portuguesa

Existem diversos recursos para a mineração de opiniões para a língua inglesa, como dicionário de sentimentos, ferramentas para processamento de linguagem natural (PLN), e corpora anotados. Já para outras línguas, como o português, os recursos são escassos. Como discutido na Seção 2.8.1, o mais completo dicionário de sentimento para o português é o SentiLex-PT (SILVA; CARVALHO; SARMENTO, 2012), no qual contém palavras de sentimento e expressões idiomáticas para português de Portugal. O Palavras (Seção 2.6) é a ferramenta de processamento de linguagem natural mais completa.

Levando em consideração esta deficiência, Silva *et al.* desenvolveram em (SILVA et al., 2009) um sistema de mineração de opiniões que combina recursos de dicionário de sentimentos, padrões léxico-sintáticos, aprendizado de máquina, e ainda ontologias, para realizar a correta classificação do sentimento de conteúdo político em português. Apesar de muitas das dificuldades enfrentadas neste trabalho se assemelharem com as nossas (e.g. conteúdo gerado por usuário), o objetivo do presente trabalho não é criar ou otimizar um método específico de classificação sentimento. Somente necessitamos selecionar um que seja mais adequado ao nosso contexto.

Avanco e Nunes em (AVANCO; NUNES, 2014) analisam o desempenho de três diferentes dicionários de sentimentos em português: SentiLex-PT, OpLexicon e LIWC. Os autores utilizam um algoritmo de classificação baseado em mudanças de polaridade (negação e intensificação) para classificar revisões de produtos em um site de compras. Como gold-standard, é utilizada a recomendação (positivo) ou não-recomendação (negativo) de um determinado produto. Os experimentos realizados testam cada dicionário de forma isolada, bem como a combinação dos três dicionários. Quando cada dicionário de sentimento é experimentado de forma isolada, os melhores resultados são obtidos com o dicionário SentiLex-PT, e os piores com o OPLexicon. Já a combinação dos três dicionários apenas piorou os resultados. Este trabalho mostra, de certa forma, que o SentiLex-PT é um dos melhores dicionários de sentimento disponíveis para o português, sendo, portanto, adotado adotado no presente trabalho.

Um estudo que analisa a polaridade de *tweets* em português usando aprendizado de máquina supervisionado é apresentado em (ALVES et al., 2014). O estudo analisa *tweets* sobre a copa das confederações de 2013, onde a análise foi dividida em duas etapas: identificação

de *tweets* com conteúdo subjetivo, e; classificação do sentimento destes *tweets* em positivo ou negativo. Os autores compararam dois algoritmos de classificação diferentes para cada etapa: SVM e Naive-Bayes. Como corpus de teste e treinamento foram utilizados tanto *tweets* anotados manualmente, quanto *tweets* anotados automaticamente com base na presença de emoticons. O rótulo positivo foi atribuído com base na presença de *emoticons* felizes (e.g. ":-)", ":D"), enquanto que o rótulo negativo, na de *emoticons* tristes (e.g. ":-(", ":("). O algoritmo SVM permitiu atingir um melhor desempenho, tanto na identificação de *tweets* com conteúdo subjetivo, como na classificação da polaridade. Os autores deste trabalho afirmam que seus resultados foram sempre superiores quando os *tweets* anotados automaticamente foram utilizados para treinamento dos classificadores. Embora lide com um corpus em português e conteúdo gerado por usuário, o domínio dos textos é diferente: futebol. Além disso, *tweets* possuem características particulares que os diferenciam de comentários de jornais, tais como tamanho e vocabulário (e.g. hashtags, abreviações). Portanto, o corpus de treinamento não pode ser reutilizado no contexto do presente trabalho (política), pois atributos discriminantes como palavras de sentimento, expressões idiomáticas, gírias, são altamente dependente do contexto original. Ainda, as técnicas e algoritmos de classificação utilizados podem não obter um desempenho semelhante, devido a esta mudança de contexto.

Um dos únicos trabalhos que abordam conteúdo gerado por usuários relacionados a notícias é (SARMENTO et al., 2009), no qual os autores criaram um conjunto de padrões léxico-sintáticos para identificar a polaridade de sentenças. A vantagem de se utilizar padrões léxico-sintáticos é que quando as sentenças casam com o padrão, é possível obter a polaridade com um alto nível de precisão. Todas as sentenças utilizadas neste trabalho são de se comentários relacionados a notícias políticas, mas o objetivo dos autores é criar um corpus de referência para mineração de opiniões na área política.

As técnicas utilizadas para classificação de conteúdo em língua portuguesa são diversas. Por exemplo, em (CHAVES et al., 2012), foi apresentado e utilizado um algoritmo baseado em ontologias e uma lista de adjetivos para a classificação de sentimento. Já em (NASCIMENTO et al., 2012), foi utilizado aprendizado de máquina. Ambos os trabalhos sofreram as mesmas dificuldades encontradas no presente trabalho, tais como falta de: dicionários de sentimento, corpora anotado, ferramentas para processamento de linguagem natural, etc. Estas dificuldades são comumente encontradas em trabalhos de mineração de opiniões que não são direcionados a língua inglesa.

3.3 Previsão Utilizando Sentimento

Muitos trabalhos têm abordado previsão com base em sentimento, principalmente usando o Twitter. Um estudo de caso é relatado em (ASUR; HUBERMAN, 2010), onde a popularidade de filmes pré-definidas no Twitter é usada para prever a sua bilheteria, usando tanto o sentimento, como o volume de *tweets*. Os autores comparam dois tipos de previsão: quantitativo (quantidade de *tweets* e *retweets* contendo URL's de material promocional sobre o filme), e baseado em opinião expressa nos *tweets*. A segunda implica a necessidade de classificação da polaridade dos *tweets*, que foi realizada utilizando o classificador DynamicLMClassifier). Eles consideram duas dezenas de filmes, séries de tempo de quase três meses, e regressão linear para a previsão do sentimento. Eles concluem que a previsão é mais afetada pelo número de menções, mas o sentimento pode ser usado em combinação para aumentar ligeiramente a precisão.

O sentimento do Twitter também é usado para prever o movimento do mercado de ações em (BOLLEN; MAO; ZENG, 2011). Diversas longas séries temporais diárias de emoção e sentimento são consideradas para a previsão. Além de regressão linear, também foi utilizado um método baseado em redes neurais fuzzy (*self-organizing fuzzy neural network* - SOFNN) para correlacionar as séries temporais de sentimento, com a série temporal Down Jones Industrial Average (DIJA), e desenvolver um modelo preditivo. Os autores ressaltam a importância da suavização e o atraso entre os eventos e o movimento das ações. Também afirmam que a polaridade do sentimento não é um bom indicador e, portanto, realizam experimentos com humor (e.g. alegria, raiva, calma) que, para casos específicos, produz melhores resultados.

O trabalho que mais se assemelha ao nosso em termos de objetivo é o estudo de caso relatado em (O'CONNOR et al., 2010), em que o sentimento expresso em *tweets* que contenham menções aos candidatos são correlacionados com indicadores externos de confiança dos consumidores (índice de sentimento do consumidor e índice econômico Gallup) e opinião política (aprovação do trabalho presidencial e intenção de voto). Usando séries temporais diárias, eles tentam prever as pesquisas através de regressão linear, mas obtêm mal resultados para todos os indicadores. Nenhuma das técnicas pode ser aplicada a dados esparsos, e, por conseguinte, não são adequados para o nosso estudo de caso.

Ainda no contexto político, um estudo (TUMASJAN et al., 2010) mostra que resultado final de uma eleição presidencial na Alemanha poderia ser previsto usando *tweets* menções a partidos políticos que estavam concorrendo. Eles examinam o sentimento dos *tweets*, mas concluem que a simples menção aos partidos políticos se correlaciona fortemente com sua res-

pectiva proporção de votos nos resultados eleitorais. O uso do Twitter como uma importante fonte de previsão de resultados eleitorais é questionado em (SCHOEN et al., 2013), com o principal argumento de que o conteúdo do Twitter pode não ser suficientemente representativo da população-alvo ou ser usado para previsão.

3.4 Considerações Finais

Esta seção tem como objetivo sumarizar as características que diferenciam o presente trabalho dos demais trabalhos apresentados neste capítulo. Estes aspectos serão discutidos a seguir:

- **Contribuições para mineração de opiniões:** nosso trabalho complementa estes trabalhos, tendo em conta uma fonte diferente de opiniões: comentários de jornais. Nós desenvolvemos uma abordagem para analisar o sentimento que eles contêm e obter atributos discriminantes baseados em sentimento para a previsão;
- **Proposta de atributos discriminantes baseados em sentimento:** as métricas de sumarização propostas são inspiradas por métricas anteriormente sugeridas em obras como (GODBOLE; SRINIVASAIAH; SKIENA, 2007; BOLLEN; MAO; ZENG, 2011), mas com o melhor de nosso conhecimento, os atributos discriminantes baseado em manifestação de sentimento não foram propostos antes;
- **Previsão:** do ponto de vista de previsão baseada em séries de tempo de sentimento, este trabalho se diferencia dos citados acima nas seguintes características: as variáveis a serem previstas são esparsas, pois são de pesquisas de intenção de voto que são publicadas infrequentemente; foi lidado com comentários gerados por usuários escritos em português brasileiro como reações a notícias políticas; foram criados atributos discriminantes baseados em sentimento que leva em consideração picos de sentimento no período analisado.

4 DESCRIÇÃO DO ESTUDO DE CASO

O propósito deste capítulo é apresentar em linhas gerais o estudo de caso desenvolvido no presente trabalho, descrevendo seus principais componentes, e também suas principais características.

4.1 Objetivos

O objetivo deste estudo de caso é verificar se o sentimento expresso sobre candidatos a eleições em comentários de notícias online pode ser utilizado na previsão da intenção de voto. A abordagem proposta dividiu o estudo de caso em dois aspectos:

- A mineração de opiniões sobre candidatos políticos, considerando a fonte de dados citada e suas características, tais como português do Brasil, linguagem não estruturada, expressões idiomáticas, gírias, etc.;
- A representação do sentimento em atributos discriminantes e o estudo de sua influência na previsão da variação de intenção de voto através de um modelo preditivo.

Para este último aspecto foram propostos atributos discriminantes baseados no sentimento extraído dos comentários de notícias, levando em consideração o problema das séries temporais esparsas. Uma visão geral do estudo de caso é mostrada na Figura 4.1.

4.2 Fontes de Dados

Neste estudo de caso foram utilizadas duas fontes de dados: comentários de notícias políticas online e resultados de pesquisas de intenção de voto.

As notícias e seus respectivos comentários foram extraídos da seção de política do jornal online Folha de São Paulo, que possui audiência nacional. Os dados analisados são referentes ao mês que precede a data de 3 eleições (primeiro turno) brasileiras: eleição governamental de 2010, eleição presidencial de 2010 e eleição municipal de 2012. Para cada eleição, foram analisados somente os candidatos com maior intenção de voto.

Como pesquisas eleitorais de intenção de voto, foram utilizadas as publicadas pelo instituto de pesquisa Datafolha[1], uma das mais tradicionais empresas de pesquisas de opinião

[1] http://datafolha.folha.uol.com.br/

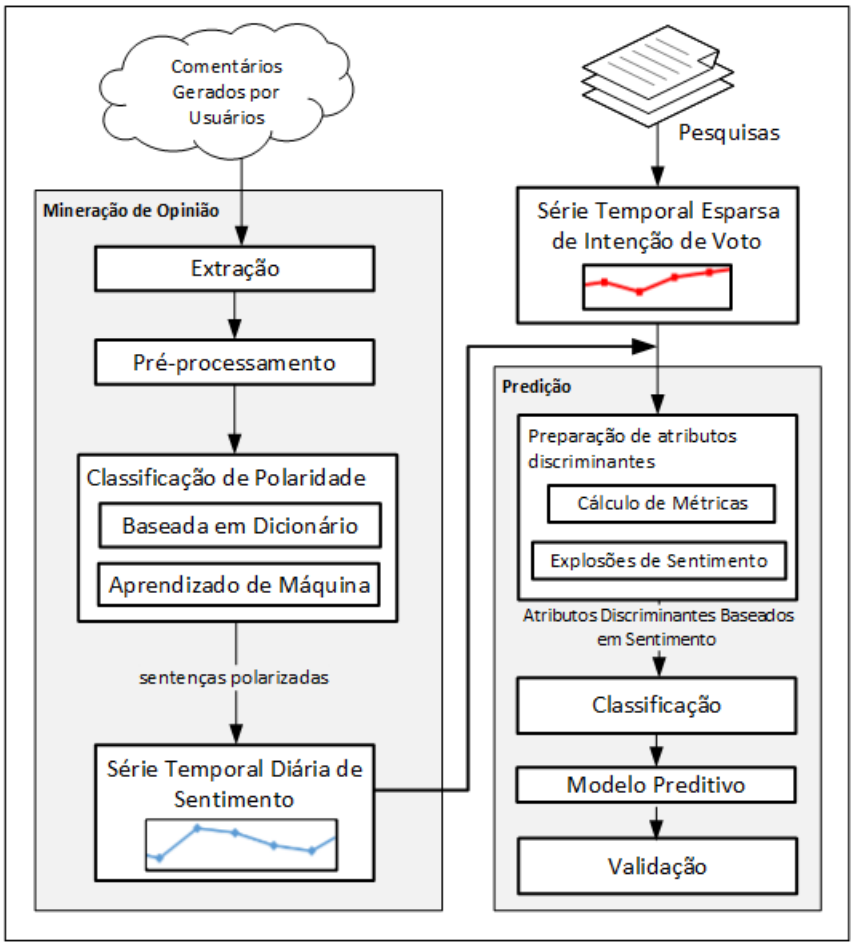

Figura 4.1 – Visão geral da abordagem proposta.

pública. Como já mencionado, uma das características das pesquisas eleitorais de intenção de voto brasileiras, é que o período entre a publicação de duas pesquisas varia enormemente entre 8 dias (começo do período observado) até 3 dias (perto da data da eleição) (DATAFOLHA, 2010b; DATAFOLHA, 2010a; DATAFOLHA, 2012a).

4.3 Características do Corpus

Nesta seção são apresentadas as principais características do corpus utilizado no presente trabalho. O capítulo contém exemplos reais de comentários que ilustram os principais problemas enfrentados. Os comentários apresentados no capítulo foram transcritos sem qualquer modificação da fonte original e não apresentam de forma alguma a opinião do autor deste trabalho. Além disso, procuramos utilizar exemplos com todos os diversos candidatos analisados.

A maioria dos problemas que o corpus apresenta é devido ao conteúdo ser gerado por usuários, o que resulta em uma grande presença de ruídos. Algumas das características presente

no corpus são: a) *spams*, conteúdo duplicado ou quase vazio; b) problemas no uso da linguagem; c) palavras de sentimento dependentes de domínio; d) opinião sobre múltiplos alvos; e) uso de caracteres especiais, e; f) ironia/sarcasmo. Estas características serão detalhadas a seguir.

- ***Spams:*** um dos problemas observados é a elevada presença de comentários duplicados. O primeiro motivo pelo qual isto ocorre é que durante a publicação do comentário, o autor aperta o botão de submeter repetidamente, resultando em comentários repetidos seguidos. Esta situação pode ser verificada através da data e hora da postagem dos comentários, que são muito próximos. O segundo motivo é a presença de *spam*, onde os autores propositalmente postam comentários com o mesmo conteúdo várias vezes, até mesmo em notícias diferentes, a fim de criticar ou enaltecer um candidato.
- **Problemas no uso da linguagem:** outra característica refere-se ao mau uso da linguagem escrita, que inclui erros de português, linguagem não-estruturada, linguagem chula, gírias, internetês, etc. Por exemplo, veja o comentário abaixo.

  > *"HADÁRDITURDU TRAIRDU PELA MILITANÇA ! ÓRBRA DI PAÍNHU "CONFÚRCIO ELEVARDU A (-1)"DISISPERAARDU !* "Tá ruçu manu ! Hadárditurdu num zadianta ! Nóis vai di russu manu da Ziguardaardi, Zonestidaardi"= FESTA NA ZZZZ ZONA ! = RUSSUMANZORRA !* SANTIN RUSSOMANZORRA PLORO VESTRA NEGLIGENTIA !*

  Neste comentário, praticamente todas as palavras do comentário possuem erros tipográficos, onde não existe qualquer tipo de estrutura léxica-sintática, e a linguagem utilizada não segue a norma culta da língua portuguesa. É importante ressaltar que este exemplo de comentário é um caso extremo, já que nem todos os comentários analisados neste trabalho estão tão mal escritos. Comentários com ruídos podem dificultar que palavras de sentimento e menções às entidades analisadas sejam encontradas.
- **Palavras de sentimento dependentes de domínio:** uma outra característica também comum em conteúdo gerado usuários são palavras de sentimento dependentes de domínio. Em nosso contexto, são palavras de sentimento relacionadas ao contexto político brasileiro. Veja o exemplo de comentário abaixo:

  > *"Meu voto e Haddad no segundo turno chega de PSDB em São Paulo chega...!!!! chega da **privataria tucana**. Todos falam do **mensalão**, entranto se fosse na época do FHC isso seria abafado totalemente..."*

  O autor do comentário utiliza os termos "privataria tucana" e "mensalão", os quais são termos usados para designar escândalos políticos relacionados aos dois partidos. Para a resolução deste problema, é necessário a disponibilidade de dicionários de sentimento especializados, ou corpora anotados especificamente para o domínio analisado.
- **Opinião sobre múltiplos alvos:** os comentários utilizados não necessariamente expressam a opinião sobre um único alvo, sendo que em alguns comentários existem múltiplas

menções de sujeitos observados. Veja o comentário abaixo:

> *"O ESQUEMA era o seguinte: "Eles" sabiam que provavelmente o **Serra** iria para o segundo turno. Então colocaram o **Russomanno** lá em cima nas pesquisas para induzir os eleitores a votarem nele e não no **Haddad**. Conforme se aproximava a eleição, foram ajustando as pesquisas para não ficar tão evidente. o velho "esquema" de manipulação que sempre fizeram e que hoje em dia não dá tão certo quanto antigamente. Mas alguns ainda acreditam. No 2º turno tentarão fazer outras para eleger o **Serra**".*

A primeira sentença expressa opinião sobre o Serra, enquanto que na segunda sentença a opinião é sobre o candidato Russomanno e Haddad. Já na última sentença, o candidato Serra volta a ser referenciado. Quando existem múltiplas menções em uma mesma sentença, surge a necessidade de se encontrar o alvo de cada opinião existente.

- **Uso de caracteres especiais:** Os comentários utilizados neste trabalharam também apresentam o uso de caracteres especiais para mascarar palavras de sentimento e menções de entidades. Veja um exemplo de comentário onde isto ocorre:

  > *"Ai tem coisa! Apoiar um **b@ndid0**, **s@f@d0**, não é paixão partidária, é interesse financeiro mesmo. Só mesmo um povo desqualificado, **pil@ntr@** apoia um político corrupto, racista, **fraud@dor**, vendido que só esta livre por causa da idade."*

  Isto ocorre pois existe uma moderação de comentários que contenham calúnia, difamação, injúria ou ofensa feita pelo site do jornal. Portanto, o uso destes caracteres dificulta que o comentário seja removido futuramente pelos moderadores. O uso destes caracteres especiais também dificulta que palavras de sentimento (principalmente de polaridade negativa) sejam encontradas.
- **Ironia/Sarcasmo:** outra característica do corpus é a presença de ironia e/ou sarcasmo. Veja o exemplo de comentário escrito abaixo:

  > *"Assistindo hoje a propaganda do serra na tv, cheguei a pensar se não foi ele quem fundou São Paulo. Tudo aqui foi feito por ele! Antes do serra, São Paulo era só mato. Depois do Serra, São Paulo virou um paraíso infinitamento melhor do que Amsterdã. Claro que o Kassab ajudou, né"*

  O uso de ironia/sarcasmo é muito comum em conteúdo gerado por usuários relacionados a política (LIU, 2012), pois os autores dos comentários usam palavras que aparentemente valorizam os políticos e seus partidos, mas possuem, na verdade, a finalidade de desvalorizá-las. A presença de ironia/sarcasmo dificulta a classificação da polaridade do texto por um classificador automático. Além disso, a ironia/sarcasmo é um dos maiores desafios em mineração de opinião.

4.4 Séries Temporais

O processo de predição de variação de intenção de voto possui como entrada duas séries temporais: uma série temporal diária de sentimento e uma série temporal de intenção de voto.

Sejam:

- $T = [t_1, \ldots, t_n]$ e $K = [k_1, \ldots, k_m]$ índices temporais, onde $K \subset T$ e $k_1 = t_1$ e $k_m = t_n$;
- $E = [e_1, \ldots, e_i]$ o conjunto de candidatos observados.

A série temporal de sentimento é resultado do processo de mineração de opiniões que foi aplicado aos comentários de notícias. Formalmente, esta série pode ser definida como $V = \{v_{jt} : t \in T, j \in E\}$, onde v_{jt} é uma quadrupla $< e_j, pos_{jt}, neg_{jt}, m_{jt} >$, e:

- e_j é uma entidade, i.e., um candidato observado;
- pos_{jt} é o total de sentimento positivo em relação a entidade e_j no tempo t;
- neg_{jt} é o total de sentimento negativo em relação a entidade e_j no tempo t;
- m_{jt} é o número total de menções em relação a entidade e_j no tempo t, independentemente do sentimento.

A série temporal de intenção de voto é extraída pesquisas eleitorais, e é considerada esparsa pelos poucos elementos que contém, particularmente quando comparada à série temporal de sentimento, que é diária. Aproximadamente doze pesquisas são publicadas por ano, e o intervalo de tempo entre duas pesquisas varia enormemente, de dias a meses. Formalmente, esta série pode ser definida como $P = \{p_{jt} : t \in K, j \in E\}$, onde p_{jt} é uma tupla $< e_j, int_{jt} >$, e:

- e_j é uma entidade;
- int_{jt} é a intenção de voto de uma entidade e_j no tempo t.

4.5 Mineração de Opiniões

O processo de mineração de opiniões desenvolvido neste estudo de caso é uma parte essencial da nossa abordagem, pois a incorreta polarização do sentimento pode afetar a predição. Nesta seção, são discutidas as principais etapas do processo e técnicas aplicadas no nosso contexto, i.e., conteúdo gerado por usuário em português expresso em forma de comentários como reações de notícias, e os desafios enfrentados. Foram experimentadas duas abordagens de classificação de sentimento: baseada em dicionário e em aprendizado de máquina. No entanto, outras técnicas de mineração de opiniões podem ser aplicadas para se obter as séries temporais de sentimento. Maiores detalhes sobre as técnicas de mineração de opiniões empregadas e os desafios enfrentados neste estudo de caso são apresentados no Capítulo 5.

4.5.1 Extração

Este primeiro passo tem como objetivo extrair automaticamente conteúdo gerado por usuário da web, criando uma base de dados com conteúdo subjetivo. A base de dados precisa conter uma quantidade considerável de comentários, pois comentários com ruído podem ser identificados e removidos no próximo passo.

4.5.2 Pré-processamento

Este passo é responsável por: a) tratar dados ruidosos, b) quebrar comentários em sentenças, c) identificar sentenças com menções para candidatos, e d) fazer as transfor-mações necessárias de acordo com cada técnica de classificação de polaridade.

Como já discutido na Seção 4.3, a maior parte do ruído presente nos dados é devida às seguintes razões: a) usuários frequentemente não submetem os comentários devidamente; b) uso extensivo de expressões informais; e c) uso incorreto da língua Portuguesa. As principais tarefas realizadas para remover ruído foram: a) identificação e remoção de comentários não significativos (em branco, duplicados, excessivamente curtos); b) identificação e correção de palavras mascaradas por caracteres especiais; c) normalização de palavras pela remoção de toda acentuação; d) identificação de gírias, expressões regionais e idiomáticas, principalmente por afetarem o desempenho do método de classificação de polaridade baseado em dicionário); e) identificação de menções para candidatos.

A análise de subjetividade foi feita em nível de sentença, pois frequentemente cada comentário expressa opiniões sobre mais de um candidato, considerados como o alvo de cada opinião. Foram selecionas sentenças com menções explicitas a candidatos, utilizando variações em torno de seus nomes. Foram identificadas e unificadas todas as menções para os nomes de candidatos, apelidos, e todo tipo de expressão que se referia ao nome do candidato.

4.5.3 Classificação de Polaridade

O objetivo desse passo é polarizar cada sentença, e atribuir esta polaridade do sentimento ao candidato alvo. Devido aos desafios da linguagem alvo, comparamos dois métodos de classificação de polaridade. Para a abordagem baseada em dicionário, o desafio foi a falta de um bom dicionário de sentimento para português brasileiro. Por outro lado, a abordagem

com aprendizado de máquina depende da anotação de corpus para obtenção de resultados de qualidade.

Independentemente da abordagem de polarização, o resultado deste passo é um conjunto de sentenças polarizadas, associada com seu respectivo candidato e tempo. Formalmente, o conjunto de sentenças é definido como $S = \{s_i : i \in \mathbb{N}\}$, onde s_i é uma quádrupla $< texto_i, pol_i, e_i, t_i >$, e:

- $texto_i$ é a sentença pré-processada;
- pol_i é a polaridade do sentimento em relação a entidade e_i no tempo t_i;
- e_i é a entidade mencionada na sentença $(e_i \in E)$;
- t_i é o tempo quando o comentário ao qual a sentença pertence foi escrito.

No método baseado em dicionário, o dicionário de sentimento SentiLex-PT (SILVA; CARVALHO; SARMENTO, 2012) foi usado para polarizar as sentenças em positivo, negativo ou neutro (1, -1 e 0). Para compensar suas limitações em relação ao português brasileiro e vocabulário informal dependente de domínio, foi criado um dicionário especializado contendo palavras de sentimento do cenário político brasileiro, gírias, expressões idiomáticas e regionais (TUMITAN; BECKER, 2013). Nesta abordagem, para cada palavra da sentença, é procurado em ambos dicionários se a palavra tem uma polaridade associada. Após isto, a polaridade de cada uma das palavras na sentença é agregada, onde os termos positivos são somados, e os termos negativos subtraídos.

Na segunda abordagem, um algoritmo chamado Sequential Minimal Optimization (SMO) foi utilizado para treinar um classificador Support Vector Machine (SVM) (SCHöLKOPF; BURGES; SMOLA, 1999). O SVM foi treinado para classificar a polaridade das sentenças em duas classes (positivo e negativo). Foram testadas diferentes preparações de dados, mas a que apresentou melhores resultados foi o uso de unigramas das sentenças com sua contagem de palavras. Também foi aplicada transformação TF-IDF (frequência do termo–inverso da frequência nos documentos) ao texto analisado. Outros classificadores também foram testados (ver Apêndice B, Seção B.1), mas eles não superaram o desempenho do classificador SVM.

Para validar ou treinar o classificador, um conjunto de sentenças selecionadas aleatoriamente foram manualmente anotadas. Devido ao alto nível de subjetividade, somente sentenças com no mínimo duas concordâncias foram consideradas.

4.6 Predição de Variação de Intenção de Voto

Este processo tem como objetivo o desenvolvimento de um modelo que, dado um conjunto de atributos discriminantes extraídos das séries temporais de sentimento, pode prever a variação de intenção de voto. O problema foi simplificado como a predição de classes discretas de variação de intenção de voto (i.e., aumentou, diminuiu, inalterado), a fim de lidar com o nosso cenário de dados esparsos. O processo de predição é composto de dois passos: a preparação dos atributos discriminantes, e o aprendizado do modelo preditivo através de classificação baseada em aprendizado de máquina. Os detalhes desta etapa do estudo de caso são mostrados no Capítulo 6.

4.6.1 Preparação de Atributos Discriminantes

As duas séries temporais V e P são transformadas em atributos discriminantes para prover dados de treinamento para um algoritmo de classificação. Para quaisquer dois pontos consecutivos de dados $t_i, t_k \in K$ ($t_i < t_k$), um registro é preparado contendo atributos discriminantes baseados em sentimento derivados de V, juntamente com a variação de intenção de voto discretizada extraída de P.

No estudo de caso, o classificador foi treinado com dois tipos de atributos baseados em sentimento: a) *métricas de sumarização* que agregam de várias formas o sentimento positivo e negativo em relação aos candidatos, e b) *explosões de expressão de sentimento* em relação aos candidatos. Estes atributos discriminantes foram preparados de forma que pudessem representar o efeito *cumulativo* do sentimento (desde o começo da campanha política), e o efeito a *curto-prazo* (desde a última pesquisa eleitoral publicada).

4.6.2 Classificação Variação de Intenção de Voto e Validação

Neste passo, os dados preparados são usados como dados de teste/treinamento para o algoritmo de classificação. Foram desenvolvidos vários experimentos para verificar qual tipo de atributo discriminante baseado em sentimento apresenta o melhor comportamento preditivo, usando diferentes algoritmos. Na etapa de *Validação* foram usados dados da intenção de voto do segundo turno das eleições como dados de teste para o algoritmo de classificação.

5 ESTUDO DE CASO: MINERAÇÃO DE OPINIÕES

O propósito deste capítulo é detalhar e apresentar os resultados do estudo de caso desenvolvido neste trabalho referente a mineração de opiniões, onde são discutidas as principais etapas do processo, as técnicas aplicadas e os desafios enfrentados.

5.1 Base de Dados

A base de dados é composta de comentários de notícias on-line sobre política. As notícias foram extraídas de forma automática da seção de política (chamada Poder) da Folha online, um dos jornais mais populares no Brasil. A base de dados contém notícias e seus respectivos comentários referentes a 3 eleições. Os dados foram coletados durante o mês que precede a data da eleição (primeiro turno), o período quando as pesquisas de opinião pública são publicadas mais frequentes. Para cada eleição, nós selecionamos os candidatos com maior intenção de voto (e portanto, os mais comentados). Os candidatos são descritos abaixo, juntamente com seus respectivos partidos políticos:

- **Eleição Governamental de 2010**: Aloízio Mercadante (PT) e Geraldo Alckmin (PSDB);
- **Eleição Presidencial de 2010**: Dilma Rousseff (PT), José Serra (PSDB) e Marina Silva (PV);
- **Eleição Municipal de 2012**: Celso Russomanno (DEM), Fernando Haddad (PT) e José Serra (PSDB).

Foram utilizadas o mesmo conjunto de sentenças para analisar o sentimento dos candidatos da eleição Presidencial e Governamental de 2010. Estas eleições envolvem candidatos do mesmo partido político e, portanto, comentários e notícias frequentemente referem-se a ambos. Foi utilizado o alvo da opinião na sentença para distinguir uma eleição da outra (e.g. uma menção para a Dilma refere-se a eleição Presidencial). O perfil dos dados é descrito na Tabela 5.1. No restante deste capítulo, iremos referir a este conjunto de dados como base de dados de 2010 e 2012.

Tabela 5.1 – Perfil dos dados da base de dados do primeiro turno das eleições analisada.

	Eleição de 2010		**Eleição de 2012**	
	Bruto	**Pré-processado**	**Bruto**	**Pré-processado**
Número de notícias	2.232	1.763	583	340
Número de comentários	225.217	190.975	36.108	25.115
Média de comentários por notícia (DP)	98,6 (±235,6)	86,1 (±206,5)	61,9 (±142,5)	44,1 (±92,5)
Número de sentenças	-	673.146	-	79.752
Média de sentenças por comentário (DP)	-	3,1 (±2,1)	-	3,2 (±2,3)
Comentários muito curtos	5.148	0	7.185	0
Comentários quase duplicados	29.094	0	3.808	0
Período	01/09/2010 até 03/10/2010		01/09/2012 até 07/10/2012	
Entidades	5 candidatos		3 candidatos	

5.2 *Gold-Standard* de Sentimento

Foi construído um *gold-standard* para avaliar o desempenho da classificação de sentimento baseada em dicionário e para usar como corpus de treinamento no método baseado em aprendizado de máquina. Nós selecionamos aleatoriamente 1.000 e 600 sentenças da base de dados das eleições de 2010 e 2012, respectivamente. Para cada base de dados, foram utilizados 3 anotadores com graduação em ciência da computação e sem prévia experiência de anotação de corpus. Eles foram instruídos a basear suas anotações no conteúdo que estava explicitamente escrito, desconsiderando qualquer tipo de suposição a respeito de políticos ou partidos políticos (SARMENTO et al., 2009), para que suas visões políticas não interfiram em seus julgamentos. O método de anotação descrito na Seção 2.9 foi adotado. Apenas sentenças com ao menos duas concordâncias foram mantidas, representando 92,7% das sentenças anotadas para 2010 e 97% para as eleições de 2012. A concordância entre os anotadores é mostrada na Tabela 5.2 e o resultado do processo de anotação para ambos os conjuntos de sentenças é mostrado na Tabela 5.3.

Embora as instruções para anotações fossem as mesmas, notamos algumas diferenças entre os dois processos de anotação, pois apenas um dos anotadores era comum nos dois processos. Comparado com o de 2012, o conjunto de sentenças das eleições de 2010 revelaram uma proporção mais alta de sentenças anotadas como positiva, e um significante maior número de sentenças neutras, o que pode ter influenciado os resultados. Durante o processo de anotação, também foram identificadas expressões regionais e idiomáticas, apelidos e palavras de sentimento informais, os quais foram incluídos no dicionário especializado.

Tabela 5.2 – Concordância entre anotadores das eleições de 2010 e 2012.

Eleição	**2010**			**2012**		
Anotadores	**A vs. B**	**B vs. C**	**C vs. A**	**A vs. B**	**B vs. C**	**C vs. A**
Porcentagem	59,60%	49,40%	48,90%	75,66%	73,33%	74%
Todos concordam	33,10%			63%		
Todos discordam	7,30%			3%		

Tabela 5.3 – Resultado do processo de anotação para ambas as bases de dados analisadas.

Polaridade	**Eleição de 2010**	**Eleição de 2012**
Positiva	154	72
Negativa	356	482
Neutra	417	28
Todos discordaram	73	18
Total	1.000	600

5.3 Processo de Mineração de Opiniões

5.3.1 Extração dos Dados

Para extrair os dados utilizados neste trabalho foi utilizada como principal ferramenta o Google Reader[1]. O Google Reader, ao receber um RSS (*Rich Site Summary*), indexa todas as entradas desse sumário. Portanto, foi adicionado ao Google Reader um RSS da seção Poder do site da Folha online. Após isto, foi possível rastrear todas as notícias desta seção do jornal através de seu endereço da Internet.

Uma vez obtido o endereço da notícia, o processo de extração ocorreu em dois passos. O primeiro passo corresponde à extração do conteúdo da notícia, dividido em: título, cidade da notícia, conteúdo, data e hora, e autor. No segundo passo da extração são extraídos os comentários da notícia, onde cada comentário possui as seguintes informações: data e hora, autor e conteúdo do comentário.

Este trabalho explorou como fonte de comentários a Folha de São Paulo. Também foram realizadas extração de comentários e notícias de outras fontes, entre elas, G1, Terra Notícias, Uol e Zero Hora. No entanto, estas fontes de notícias possuíam poucos comentários e nos comentários existentes, haviam poucas menções aos candidatos políticos observados. Portanto, estas fontes não foram utilizadas neste estudo de caso.

[1] http://www.google.com/reader - descontinuado

5.3.2 Pré-processamento

Esta seção detalha como os problemas descritos na Seção 4.3 foram tratados. Embora os tratamentos aplicados aos corpora serão apresentados de forma linear, este processo é altamente interativo, no qual retornos a passos anteriores são necessários para melhorar os resultados.

Quanto ao ruído existente, foi notada a existência de comentários duplicados ou quase duplicados (i.e., pequenas modificações). Este problema é reconhecido em mineração de opiniões como *spam* de opinião (JINDAL; LIU, 2008). Para verificar a extensão deste problema, para cada base de dados das eleições foi utilizada a Similaridade de Cosseno para obter uma medida de similaridade entre todos os pares de comentários. A similaridade entre pares de comentários para ambas as eleições é mostrada na Figura 5.1. Foi detectado uma quantidade elevada de comentários com no mínimo 85% de similaridade, correspondendo a 29.054 (12%) comentários na base de dados de 2010 e 3.808 (10%) comentários na base de dados de 2012. Assim, tomou-se a decisão de removê-los, com o propósito de remover o viés dos resultados.

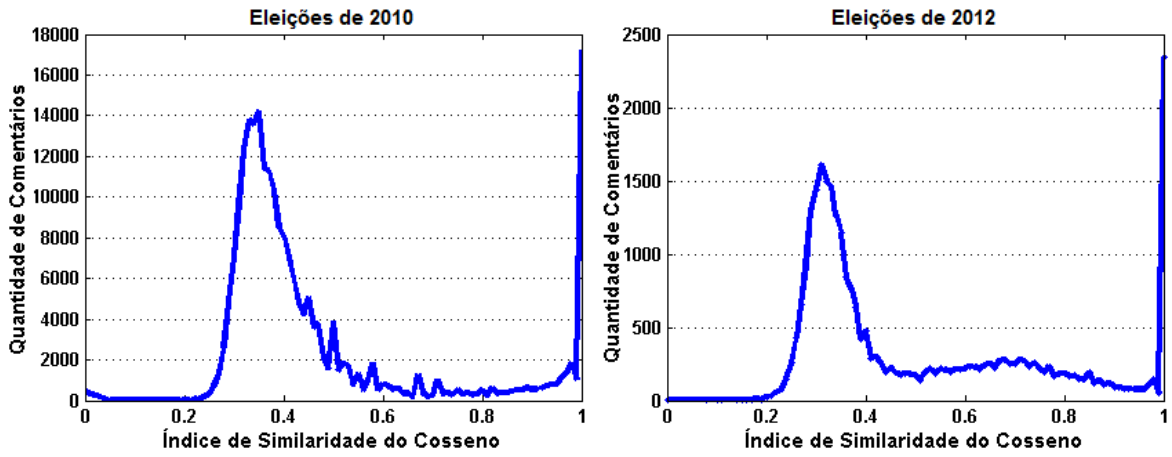

Figura 5.1 – Similaridade entre pares de comentários para a base de dados de 2010 (esquerda) e para a base dados de 2012 (direita).

Também foram removidos comentários excessivamente curtos (até 3 palavras), porque geralmente não possuem conteúdo relevante. Estes comentários representavam 2,2% (5.148 comentários) e 19,9% (7.185 comentários) das bases de dados de 2010 e 2012, respectivamente.

Outro ruído importante foram palavras de baixo calão mascaradas pelo uso de caracteres especiais. Estas palavras são mascaradas devido à moderação de conteúdo feita pelo jornal. Para resolver isto, nós apenas substituímos caracteres especiais manualmente pré-selecionados por sua letra correspondente (idi0t@ $\rightarrow$ idiota). Ao se fazer isto, pode ter ser sido introduzidos

erros que eram previamente inexistentes, mas uma análise manual superficial revelou que, na maioria dos casos, as palavras mascaradas correspondem a palavras de sentimento negativas.

Para tratar a falta ou mal-uso de acentuação, foram removidas todas as letras acentuadas tanto do texto das sentenças, quanto dos dicionários de sentimento.

Para identificar menções aos candidatos observados, foi compilado um conjunto de possíveis termos usados para referenciá-los, utilizando expressões regulares baseadas em variações de seus nomes. Por exemplo, para o candidato José Serra, nós encontramos "zehserra", "serrinha", "serremos", "vampserra". Algumas dessas menções possuem sentimento (e.g. malhaddad, vampserra, Dilmais). Estes termos, quando identificados, foram adicionados a um dicionário de sentimento especifico de domínio. Além disso, para encontrar vocabulário dependente de domínio para a abordagem baseada em dicionário, foram manualmente analisadas as 1.000 palavras mais frequentes que não foram encontradas no SentiLex-PT.

Foram desenvolvidos vários experimentos, incluindo a remoção de *stop words*, *stemming*, e o tratamento de negação utilizando uma janela de proximidade (ver Seção 5.3.3.2). Nenhuma dessas ações apresentou em bons resultados. A respeito da remoção de *stop words*, na abordagem baseada em dicionário, esta dificultou o reconhecimento de expressões idiomáticas (e.g. "Ele é o cara"). Já na abordagem de aprendizado de máquina, a remoção de *stop words* diminuiu significativamente a precisão.

Finalmente, para quebrar os comentários em sentenças foi utilizado uma plataforma de processamento de linguagem natural chamada NLTK[2], usando um módulo específico para Português (*punkt*).

Utilizando o Palavras, foram desenvolvidos experimentos para quebrar sentenças em cláusulas, tratamento de negações e o descobrimento do verdadeiro alvo das palavras de sentimento. No entanto, não apresentou bons resultados devido ao excesso de erros sintáticos e de estrutura, bem como uso de linguagem informal. Resultados obtidos com o Palavras são mostrados no Apêndice B.

5.3.3 Classificação de Sentimento Baseada em Dicionário

5.3.3.1 Descrição

Neste trabalho, foi primeiramente experimentado o método baseado em dicionário, pois este não necessita de uma grande quantidade de corpus anotado. Apenas é necessário um con-

[2]http://nltk.org

junto de sentenças (*gold-standard*) para validar o método.

A classificação baseada em dicionário é baseada na coocorrência de palavras de sentimento e menções dos candidatos observados, onde os termos positivos são adicionados e os negativos são subtraídos. A polaridade dos termos da sentença é dada pelo dicionário de sentimento. Veja a sentença:

*"**Serra** é um **ótimo** candidato, **melhor** que ele nunca vi "*

Nesta sentença existe a menção do candidato José Serra (*"Serra"*) e a presença das palavras de sentimento *"ótimo"* e *"melhor"*, ambas com sentimento positivo (+1). Portanto, os valores das classes das palavras de sentimento são somados (+2), resultando na polaridade da sentença (positiva) associada a menção do candidato observado.

Um dos problemas desta abordagem é que o resultado da classificação de sentimento pode ser neutra, se houver a mesma quantidade de palavras positivas e negativas na sentença. Veja a sentença:

*"A Dilma **perdeu** sua **credibilidade**"*

O termo "perdeu" possui polaridade negativa (-1), enquanto "credibilidade" possui polaridade positiva (+1). Consequentemente, o resultado da polarização desta sentença é erroneamente neutro. Apesar desta deficiência, este método não exige a utilização de analisadores sintáticos ou morfológicos.

5.3.3.2 Resultados

Os experimentos com a abordagem baseada em dicionário envolveram apenas as eleições de 2012 (TUMITAN; BECKER, 2013). Os experimentos levaram em consideração o *gold-standard* da eleição de 2012, ou seja, o conjunto de 600 sentenças anotadas, onde 72 sentenças foram anotadas como positiva, e 482 sentenças como negativa. Foram desenvolvidos diferentes experimentos para classificar o sentimento destas sentenças utilizando a abordagem baseada em dicionário. Em cada variação do experimento foi realizado uma tentativa de tratamento de advérbios de negação (e.g. não, nem) baseada em uma janela de distância de três palavras entre o termo de sentimento e um termo de negação. Os resultados das variações das abordagens de classificação são mostrados na Tabela 5.4 usando as medidas de acurácia, precisão, revocação e medida-F, no qual não são mostrados os resultados para a classe neutra.

Tabela 5.4 – Acurácia (A), Precisão (P), Revocação (R) e medida-F (F) para cada variação do processo de classificação e suas respectivas abordagens.

Variação	Abordagem	A(%)	Polaridade	P (%)	R (%)	F(%)
Baseline	Com Negação	34,11	Positiva	18,35	25,64	21,39
			Negativa	89,53	35,48	50,82
	Sem Negação	35,54	Positiva	17,35	21,79	19,32
			Negativa	89,22	37,76	53,06
Dicionário Especializado	Com Negação	42,68	Positiva	24,87	62,82	35,64
			Negativa	89,62	39,42	54,76
	Sem Negação	43,21	Positiva	26,02	65,38	37,23
			Negativa	90,52	39,63	55,12
Stemming	Com Negação	34,82	Positiva	21,28	38,46	27,40
			Negativa	89,19	34,23	49,48
	Sem Negação	37,32	Positiva	21,01	32,05	25,38
			Negativa	87,62	38,17	**53,18**
Sem Acentuação	Com Negação	52,14	Positiva	**26,99**	56,41	36,51
			Negativa	89,86	51,45	65,44
	Sem Negação	52,14	Positiva	**26,99**	56,41	36,51
			Negativa	90,18	51,45	**65,52**

- **Baseline:** consiste na aplicação do método baseado em coocorrência, sem nenhum pré-processamento especial além da tokenização. Nesta abordagem, foi obtida uma boa precisão para a classificação de sentenças negativas. No entanto, a precisão para sentenças positivas e revocação para ambas sentenças positivas e negativas não são satisfatórias. Portanto, os próximos experimentos desenvolvidos tiveram como propósito melhorar a classificação de sentimento de sentenças positivas e a revocação.
- **Dicionário Especializado:** observou-se no experimento *baseline* que um dos problemas foi que o léxico Sentilex-PT não capturava corretamente termos do português do Brasil, bem como gírias, expressões idiomáticas ou termos dependentes do contexto. Para suprir esta deficiência, foram manualmente analisadas as mil palavras mais frequentes do corpus que não estavam no SentiLex-PT. Como resultado, foram selecionadas e adicionadas ao dicionário de sentimento 268 novas palavras e expressões idiomáticas. A maioria das palavras encontradas são gírias ou expressões idiomáticas que fazem referência aos candidatos observados. Por exemplo, foram encontradas palavras como “baixaria”, “privataria”, “mensalão”,“bolsa-esmola”. Também foram encontrados e adicionados a este dicionário variações dos nomes dos candidatos que denotam sentimento, por exemplo, “Dilmalvadeza”, “Bandilma”, “Motoserra”, “VampSerra”,“Malhaddad”. Um outro tipo de menção existente é composto de apelidos dados aos candidatos, que somente puderam ser extraídos através desta análise manual, são exemplos deste tipo de menção: “Mr. Burns”, “Nosferatu”, “Molusco”, que são referências a um dos candidatos. Apesar do

SentiLex-PT conter expressões idiomáticas, foram encontradas novas expressões que não estavam presentes no dicionário, como: “farinha do mesmo saco”, “põe a mão na massa”, “vá se catar”. Com esta modificação, os resultados para classe positiva melhoraram significativamente. Como consequência dos bons resultados desta variação, o dicionário especializado foi adotado nas variações de experimentos abaixo descritos.

- ***Stemming:*** observou-se também que algumas flexões de palavras de sentimento não estavam presentes no SentiLex-PT, o que prejudicava a revocação da classificação de sentimento. Portanto, nesta variação, além do dicionário especializado, foi aplicado um *stemmer* nas sentenças analisadas e no léxico sentimento utilizado. O *stemmer* utilizado é chamado RSLP, incluso no NLTK (ORENGO; HUYCK, 2001). Este *stemmer* foi desenvolvido especificamente para o idioma Português, descartando a necessidade de traduzir o texto. Embora com resultados superiores ao *baseline*, o uso de *stemming* piorou os resultados obtidos com o uso do léxico especializado.
- **Sem Acentuação:** observou-se também uma grande quantidade de palavras nos comentários sem qualquer acentuação, ou com acentuação incorreta. Portanto, foi removida a acentuação dos comentários e do dicionário de sentimento. Como resultado, a revocação para sentenças negativas foi melhorada significantemente, mantendo a mesma precisão para ambas as classes. Então, de acordo com acurácia, e medida-F, o resultado deste experimento constitui uma melhora significativa comparadas a todas as tentativas anteriores. Este então é a melhor configuração encontrada para a abordagem baseada em dicionário.

De forma geral, a abordagem de classificação de dicionário não obteve um bom desempenho na classificação de sentenças positivas. Alguns dos problemas encontrados foram resolvidos até um certo ponto. Conteúdo gerado por usuários possui muitos erros tipográficos, e portanto, tanto o SentiLex-PT, quanto o dicionário especializado podem conter a palavra de sentimento, mas o termo não será reconhecido. A remoção da acentuação tentou minimizar este problema. Também foi observada a importância de palavras de sentimento relacionadas a um contexto específico (KU; LIANG; CHEN, 2006; GODBOLE; SRINIVASAIAH; SKIENA, 2007), com muitas gírias e expressões idiomáticas.

Como já mencionado, um problema importante é que muitas sentenças expressam opiniões comparativas (e.g. *“O candidato X é melhor que o candidato Y, porque ele está menos envolvido com corrupção”*), onde a polaridade negativa da palavra “corrupção” será associada a ambos os candidatos. Consequentemente, não foi possível identificar que a preferência do autor do comentário pelo candidato X.

Por fim, a existência de ironia e sarcasmo em conteúdo gerado por usuário e especialmente de domínio político, é um problema de difícil solução, sendo então um dos grandes desafios em mineração de opiniões (CARVALHO et al., 2009; SARMENTO et al., 2009). Como consequência deste problema, o classificador polarizou erroneamente sentenças positivas como negativas (e.g. "Que maravilha! Vou continuar pagando vários pedágios em São Paulo!"). Estes problemas são comumente encontrados em trabalhos de mineração de opiniões (LIU, 2012).

5.3.4 Classificação de Sentimento Baseado em Aprendizado de Máquina

5.3.4.1 Descrição

Em função dos maus resultados obtidos na abordagem baseada em dicionário, principalmente em classificar sentenças positivas, foi utilizada uma segunda abordagem utilizando aprendizado de máquina para classificar as sentenças analisada. Para esta abordagem, foi utilizado o *gold-standard* de sentimento composto pelas 1.000 sentenças da eleição de 2010, e as 600 sentenças da eleição de 2012. Os algoritmos utilizados nessa abordagem estão disponíveis no ambiente Weka (HALL et al., 2009). A classificação utilizando aprendizado de máquina pode ser dividida em dois passos: a) preparação dos atributos discriminantes; b) submissão ao classificador. Estes passos serão detalhados a seguir.

Os atributos discriminantes submetidos ao algoritmo de classificação foram resultado dos seguintes procedimentos:

- Foram utilizados todos os termos das sentenças analisadas em um modelo *bag-of-words* (i.e., não foi considerada a ordem entre os termos).
- Foram extraídos unigramas dos textos, juntamente com sua frequência relativa (TF-IDF) como peso dos termos.

Além destas preparações, outras foram testadas, mas não apresentaram bons resultados, sendo portanto desconsideradas na apresentação dos resultados. Entre as variações que apresentaram baixo desempenho estão o uso de *stemming*, a remoção de *stop words*, e a adoção de bigramas e trigramas. Artifícios como classe morfológica, dependência sintática, seleção de atributos discriminantes *feature selection* não foram utilizados.

No segundo passo, o algoritmo Sequential Minimal Optimization (SMO) foi utilizado para treinar um classificador Support Vector Machine (SVM). O SVM foi treinado para classificar as sentenças em positivo e negativo, ou seja, a classe neutra não foi considerada durante

esta abordagem. Além do SMO, foram testados outros algoritmos de classificação disponíveis no Weka (e.g. Naive Bayes, J48, Logistic), que apresentaram resultados foram inferiores.

5.3.4.2 Resultados

Foi comparado o desempenho da abordagem de aprendizado de máquina e da baseada em dicionário usando medidas padrões (acurácia, precisão, revocação e medida-F). Para os experimentos com SVM, foram desenvolvidas duas variações: a) uso de ambas as sentenças de 2010 e 2012 com *cross-validation* de 10 *folds*, e; b) uso do conjunto de sentenças de 2010 como um conjunto de treinamento, e o de 2012 como um conjunto de teste. Como já mencionado, variações de pré-processamento e algoritmos foram testadas, sendo que apenas os melhores resultados são mostrados na Tabela 5.5.

Tabela 5.5 – Resultados dos experimentos comparando a abordagem baseada em dicionário e aprendizado de máquina, de acordo com Acurácia (A), Precisão (P), Revocação (R) e medida-F (F).

Eleição	**Abordagem**	**A(%)**	**Polaridade**	**P (%)**	**R (%)**	**F(%)**
2010	Baseada em Dicionário	50,39	Positiva	29,39	62,99	40,08
			Negativa	54,79	44,94	49,38
	SVM	**81,37**	Positiva	**70,63**	**65,58**	**68,01**
			Negativa	**85,56**	**88,20**	**86,86**
2012	Baseada em Dicionário	52,14	Positiva	26,99	**56,41**	**36,51**
			Negativa	**90,18**	51,45	65,52
	SVM	**83,24**	Positiva	**27,59**	10,53	15,24
			Negativa	86,45	**95,38**	**90,70**
	SVM (2010 como dados de treinamento)	77,40	Positiva	25,56	30,26	27,71
			Negativa	87,98	85,27	86,61

A abordagem baseada em dicionário apresentou quase o mesmo desempenho para ambas as bases de dados de 2010 e 2012, mantendo aproximadamente a mesma precisão, revocação e medida-F. No entanto, a precisão para sentenças negativas das eleições de 2012 (90.18%) foi consideravelmente mais alta, quando comparada com as sentenças de 2010 (54.79%).

Entre os algoritmos testados, o SVM consistentemente apresentou o melhor desempenho. No conjunto de sentenças da Eleição de 2010, a acurácia foi de 81.37%. Comparado com a abordagem baseada em dicionário para a mesma base de dados, a precisão, revocação e valor-F foram melhores para ambas a classes (positiva e negativa). Para as eleições de 2012, a acurácia do SVM com *cross-validation* (83.24%) também foi melhor, devido principalmente aos resultados para a classe negativa.

Quando considerado o conjunto de sentenças de 2010 como conjunto de treinamento, e o de 2012 como conjunto de teste, a acurácia do SVM diminui significativamente (77.4%). Uma das possíveis explicações para isto é a existência de um superajustamento. Para investigar esta hipótese, foi aplicada uma técnica de seleção de atributos discriminantes baseada em *Information Gain* para ambos os conjuntos de teste e treinamento. Por exemplo, nas eleições presidenciais de 2010 havia várias referências a um dos candidatos por ser do sexo feminino, e seus respectivos atributos discriminantes foram decisivos para a classe positiva. Nas eleições de 2012, o vocabulário se refere ao passado de cada candidato, e seus envolvimentos com escândalos.

Consistentemente, foram enfrentados resultados ruins para a classe positiva com todos os métodos. Uma das razões, é que as sentenças positivas são realmente escassas na base de dados, portanto, ambos o treinamento e validação usando o *gold-standard* podem conter viés. Além disso, os algoritmos de classificação utilizados podem ter um desempenho inferior ou sofrer um superajustamento devido a este desbalanceamento. Uma possível solução para este problema, seria o uso de técnicas de balanceamento de base dados, como por exemplo, o Synthetic Minority Oversampling Technique (SMOTE) (CHAWLA et al., 2002), que permite o aumento de instâncias que são minorias, que em nosso caso são sentenças com polaridade positiva. Outra explicação para o desempenho ruim na classificação de sentenças positivas é o uso constante de ironia nos comentários.

5.3.5 Considerações

Observa-se que a maior parte do esforço em relação à classificação de sentimento neste trabalho concentrou-se no método baseado em dicionário, pois uma de suas grandes vantagens é não necessitar de um corpus anotado como pré-condição de aplicação. A utilização deste método limita-se à procura de palavras de sentimento em um léxico de sentimentos. Portanto, insistimos bastante neste tipo de abordagem, o que resultou em vários experimentos desenvolvidos neste sentido: uso do Palavras; desenvolvimento de um dicionário especializado; *stemming*; diversas iterações de pré-processamento, etc. Apesar deste imenso esforço, os resultados não foram satisfatórios, o que levou à investigação e adoção da abordagem baseada em aprendizado de máquina.

Por ter apresentado os melhores resultados, os experimentos de predição de variação de intenção de votos descritos no próximo capítulo foram desenvolvidos usando o sentimento da abordagem SVM com *cross-validation*. Contudo, ao se avaliar os resultados do método baseado

em aprendizado de máquina, observa-se que seus resultados são atrelados aos dados rotulados. Apesar dos resultados com SVM serem melhores, o desempenho do experimento que utiliza os dados da eleição de 2010 como dado de treinamento de 2012 como teste, possivelmente seja mais realista.

6 ESTUDO DE CASO: PREDIÇÃO DE VARIAÇÃO DE INTENÇÃO DE VOTO

O propósito deste capítulo é detalhar e apresentar os principais passos e resultados do estudo de caso desenvolvido referente a predição de intenção de voto.

6.1 Pesquisas de Opinião Pública

Foram utilizadas pesquisas de opinião pública de intenção de voto publicadas pela Datafolha [1], umas das mais tradicionais companhias de pesquisa, no qual vem apresentando resultados bem consistentes. Todas as pesquisas (DATAFOLHA, 2010b; DATAFOLHA, 2010a; DATAFOLHA, 2012a) correspondem ao primeiro turno de sua respectiva eleição. As datas das pesquisas de cada eleição diferem uma das outras. Nós utilizamos como primeiro ponto de dado em todas as séries temporais de intenção de voto, os números disponíveis a partir da pesquisa precedente ao dia 1 de setembro. Os períodos entre as pesquisas variam entre 8 dias (começo do período observado) até 3 dias (perto da data da eleição). Esta irregularidade ilustra a necessidade da abordagem proposta nesse trabalho para a previsão de variação de intenção de voto. Os dados de intenção de votos do primeiro turno da Eleição Governamental de 2010, Eleição Presidencial de 2010, Eleição Municipal de 2012 são detalhados no Apêndice A deste documento, nas tabelas A.1, A.2 e A.4, respectivamente.

6.2 Atributos Discriminantes Preditivos

Dois atributos discriminantes baseados em sentimento são propostos para treinar um classificador. O primeiro tipo de aspecto consiste em métricas que sumarizam o sentimento positivo e negativo com respeito a entidades, assim como o volume de menção relacionado a eles. Este tipo de atributo são variações de métricas que são propostas em trabalhos como (GODBOLE; SRINIVASAIAH; SKIENA, 2007; ASUR; HUBERMAN, 2010; BOLLEN; MAO; ZENG, 2011).

O segundo tipo de atributo discriminante representa explosões de sentimento no período (i.e. mais sentimento que o normal), indicando que as pessoas estão escrevendo mais comentários opinativos. Picos revelam reações exaltadas a notícias. Por exemplo, picos negativos podem significar reações a um novo escândalo, enquanto que picos positivos podem revelar

[1] http://datafolha.folha.uol.com.br

apoio a alguma declaração ou ação de campanha.

Nós preparamos outros tipos de atributos para representar tanto o efeito cumulativo (desde o começo da campanha), quanto o efeito a curto prazo (desde a última pesquisa que foi publicada). A primeira preparação traduz a estratégia geral de campanha, enquanto que a última representa o curso de ações tomadas em resposta do resultado da última pesquisa.

Então, dados dois momentos $t_i, t_k \in T$, $t_i < t_k$, para cada atributo discriminante foram calculadas duas variações, considerando diferentes intervalos de tempo $[t_i, \ldots, t_k]$, a seguir:

- **cumulativo**: esta variação leva em consideração todo o sentimento expresso desde o começo do período observado, i.e. $[t_1, \ldots, t_k]$.
- **curto prazo:** esta variação leva em consideração todo o sentimento expresso entre duas pesquisas consecutivas, i.e. $k = i + 1$.

6.2.1 Métricas de Sumarização

A Tabela 6.1 descreve sete diferentes métricas propostas para sumarizar o sentimento através do tempo. Além das métricas de sentimento, também foi incluída uma métrica para sumarizar menções para os candidatos (métrica s7). Nas fórmulas de 1 à 7, E corresponde ao conjunto de candidatos da eleição, $j \in E$ corresponde ao candidato, e $Q \subseteq T$ refere-se a o período de tempo no qual o sentimento é sumarizado. Todas as métricas são representadas por razões com o propósito de normalizar os dados em relação as três eleições, e torná-los comparáveis. De fato, os volumes de comentários são diferentes de acordo com a audiência das notícias, ou seja, a população da cidade de São Paulo (eleição municipal), estado de São Paulo (eleição governamental) e brasileiros (eleição presidencial). Em (TUMITAN; BECKER, 2013) foi realizada uma análise da correlação entre estas métricas e a variação da intenção de voto para cada candidato das eleições de 2012. Contudo, nenhum padrão consistente foi encontrado.

6.2.2 Explosões de Sentimento

Este tipo de atributo discriminante tem como objetivo sumarizar e identificar picos de sentimento positivo e negativo ao longo de um período de tempo, i.e. mais sentimento que o normal. Este atributo foi criado baseado na suposição que explosões de sentimento expressam reações a eventos que podem influenciar na intenção de votos, e portanto, pode ser usado para

Tabela 6.1 – Descrição da métrica de sumarização

Descrição	**Métrica**
Razão do sentimento positivo em relação a uma entidade sobre o sentimento negativo da mesma entidade	$$\text{s1}_{jQ} = \frac{\sum\limits_{t \in Q} pos_{jt}}{\sum\limits_{t \in Q} neg_{jt}} \quad (6.1)$$
Razão do sentimento positivo em relação a uma entidade sobre o sentimento total da mesma entidade	$$\text{s2}_{jQ} = \frac{\sum\limits_{t \in Q} pos_{jt}}{\sum\limits_{t \in Q} (pos_{jt} + neg_{jt})} \quad (6.2)$$
Razão do sentimento negativo em relação a uma entidade sobre o sentimento total da mesma entidade	$$\text{s3}_{jQ} = \frac{\sum\limits_{t \in Q} \text{neg}_{jt}}{\sum\limits_{t \in Q} (pos_{jt} + neg_{jt})} \quad (6.3)$$
Razão da diferença entre sentimento positivo e negativo em relação a uma entidade sobre o sentimento total da mesma entidade	$$\text{s4}_{jQ} = \frac{\sum\limits_{t \in Q} (pos_{jt} - neg_{jt})}{\sum\limits_{t \in Q} (pos_{jt} + neg_{jt})} \quad (6.4)$$
Razão do sentimento positivo em relação a uma entidade sobre o sentimento positivo total (em relação a todas as entidades)	$$\text{s5}_{jQ} = \frac{\sum\limits_{t \in Q} pos_{jt}}{\sum\limits_{c \in E} \sum\limits_{t \in Q} pos_{ct}} \quad (6.5)$$
Razão do sentimento negativo em relação a uma entidade sobre o sentimento negativo total (em relação a todas as entidades)	$$\text{s6}_{jQ} = \frac{\sum\limits_{t \in Q} neg_{jt}}{\sum\limits_{c \in E} \sum\limits_{t \in Q} neg_{ct}} \quad (6.6)$$
Razão das menções em relação a uma entidade sobre o total de menções (em relação a todas as entidades)	$$\text{s7}_{jQ} = \frac{\sum\limits_{j \in Q} m_{jt}}{\sum\limits_{c \in E} \sum\limits_{t \in Q} m_{ct}} \quad (6.7)$$

prever sua variação. Por exemplo, se durante um determinado período existem explosões de sentimento negativo (e.g. reação a um escândalo), a intenção de voto para um candidato alvo deste sentimento pode diminuir.

Para identificar picos, foi adotado o processo de quantização descrito em (ROMANI et al., 2013), que tem como intenção detectar evento importantes em uma série temporal produzida por dados de sensores. A técnica identifica “picos”, “vales” e “plateaus” em uma série temporal. Dado um limiar, picos correspondem a valores que são considerados muito maiores

que o esperado, enquanto que vales correspondem a valores em uma série temporal que são muito menores do que o esperado.

Considerando a série temporal V resultante do processo de mineração de opiniões, foram preparadas três séries temporais para cada entidade $j \in E$:

- $POS_j = \{pos_{jt} : t \in T, j \in E\}$, *i.e.*, sentimento positivo;
- $NEG_j = \{neg_{jt} : t \in T, j \in E\}$, *i.e.*, sentimento negativo;
- $R_j = \{\frac{pos_{jt}}{neg_{jt}} : t \in T, j \in E\}$, *i.e.*, razão de sentimento positivo sobre sentimento negativo.

Aplicando a técnica de quantização anteriormente mencionada, foram identificados os picos e vales para cada série temporal. A Figura 6.1 exemplifica os picos e vales para uma série temporal da razão de sentimento. Foram definidos limiares diferentes para cada tipo de série temporal, pelo fato de existirem um maior número de sentenças negativas do que positivas. Os limiares foram definidos empiricamente.

Finalmente, foram preparadas quatro métricas para representar os picos de sentimento expressos, calculados de acordo com as variações cumulativa e curto prazo:

- ExplosõesPositivas: dado POS_j, se no período considerado existir ao menos um pico de sentimento positivo;
- ExplosõesNegativas: dado NEG_j, se no período considerado existir ao menos um pico de sentimento negativo;
- ExplosõesDelta: dado POS_j e NEG_j, a diferença entre o número de picos e vales identificados em ambas séries temporais;
- RazãoExplosões: dado R_j, se existe uma predominância de montanhas ou vales no período observado da série temporal;

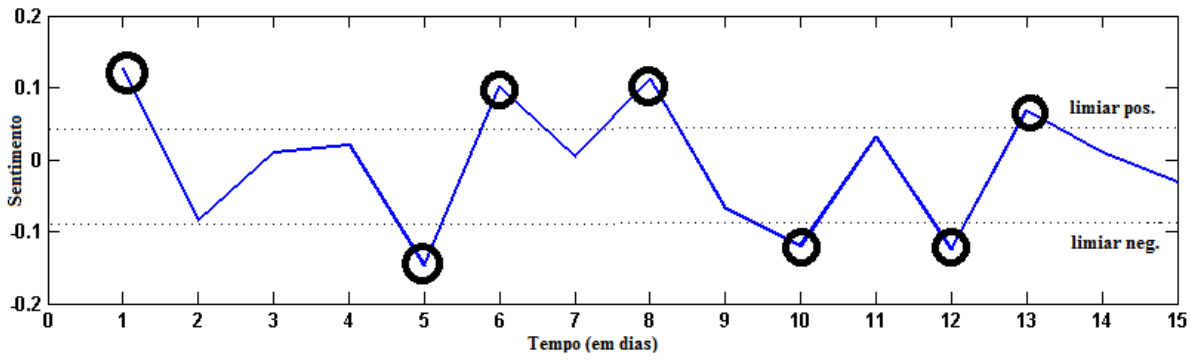

Figura 6.1 – Série temporal da razão do sentimento com picos e vales.

6.3 Experimentos

6.3.1 Descrição

Esta seção descreve os resultados dos experimentos envolvendo o desenvolvimento de um modelo de predição de variação de intenção de voto utilizando atributos baseados em sentimento.

As variações de intenção de voto extraídas das pesquisas de opinião pública são usadas como classe alvo para serem preditas, e os atributos discriminantes baseados em sentimento são preparados como descrito na Seção 6.2. Levando em consideração as variações entre duas pesquisas consecutivas para cada candidato, o conjunto de treinamento contém 51 registros para a classe "aumentou", 16 para a classes "diminuiu", e 10 para a "inalterada".

Foram experimentados diferentes algoritmos de classificação. A lista completa dos algoritmos testados é mostrada na Seção B.2 do Apêndice B deste documento. Nesta seção serão reportados apenas os algoritmos que apresentaram três melhores resultados: OneR (HOLTE, 1993), Naive Bayes (JOHN; LANGLEY, 1995) e Multinomial Logistic, refenciado no texto apenas como Logistic (CESSIE; HOUWELINGEN, 1992). Nos experimentos, a predição foi tratada tanto como um problema ternário (i.e. aumentou, diminuiu, e inalterada), quanto como um problema binário (i.e. aumentou ou diminuiu). Para este último caso, todos os registros com a classe "inalterada" foram descartados.

O objetivo desses experimentos foi avaliar o poder de predição de cada atributo discriminante proposto. Portanto, cada tipo de atributo discriminante foi submetido como um atributo preditivo isoladamente. Seu uso em conjunto também foi testado. Além disso, também foram realizados experimentos utilizando os dados do segundo turno de cada eleição como conjuntos de teste. Os experimentos relacionados ao segundo turno são reportados na Seção 6.3.3.

6.3.2 Resultados

A Tabela 6.2 mostra o resultado de cada métrica de sumarização, usando as preparações de curto prazo (CP) e cumulativas (C). A seguir serão discutidos os resultados de cada classificador para este tipo de métrica.

- **OneR**: para a preparação cumulativa, entre os classificadores analisados, o OneR foi o

que obteve melhor desempenho preditivo. Este desempenho foi obtido para a métrica s6 (Fórmula 6.6), ou seja, a razão de sentimento negativo em relação a uma entidade sobre o total de sentimento negativo de todas as entidades. Esta métrica produziu os melhores resultados para ambos os problemas de duas e três classes (70,73% e 54,90% de acurácia) na preparação cumulativa.

- **Logistic**: com o classificador Logistic, os melhores resultados foram atingidos com a preparação de curto prazo, na mesma métrica s6. Obteve-se uma acurácia de 64,41% para o problema de duas classes, e aproximadamente 51% para a variação de três classes.
- **Naive Bayes**: neste classificador nenhuma das métricas observadas se destacou com bons resultados. Além disto, não apresentou de forma geral melhor desempenho que os outros dois classificadores, tanto para preparação cumulativa, quanto para a preparação de curto prazo.

De forma geral, a métrica s6 (Fórmula 6.6) foi a que repetiu os melhores resultados entre os classificadores testados, considerando a preparação cumulativa e de curto prazo, bem como para ambos os problemas de duas e três classes. Em todos os classificadores, quando todos os atributos discriminantes foram submetidos ao algoritmo de classificação de forma combinada, ou seja, as seis métricas juntas, o desempenho se manteve ou piorou em relação ao melhor resultado obtido com métricas isoladas.

Os experimentos com os atributos discriminantes baseados em explosões de expressão de sentimento são apresentados na Tabela 6.3. Em todos os classificadores observados, para a variação cumulativa, o atributo preditivo ExplosõesDelta apresentou o melhor desempenho, onde foi atingida uma acurácia de 65,85% para ambos os classificadores OneR e Naive Bayes para o problema de duas classes. O classificador Logistic, para esta mesma métrica e preparação, obteve um resultado semelhante, com 63,41% de acurácia. Ainda para a preparação cumulativa, o melhor resultado para o problema de três classes foi obtido pelo classificador Naive Bayes (50,98% de acurácia).

Na preparação de curto prazo, o melhor resultado para o problema de duas classes foi uma acurácia de 60,97%. No entanto, este resultado foi obtido com métricas diferentes em cada classificador: ExplosõesDelta, para o classificador OneR; ExplosõesPositivas, para o classificador Naive Bayes; ExplosõesPositivas e ExplosõesDelta, para o classificador Logistic. Observa-se então que estes dois tipos de atributos discriminantes (ExplosõesPositivas e ExplosõesDelta) foram os que obtiveram melhor comportamento preditivo para a preparação de curto prazo.

Tabela 6.2 – Acurácia dos atributos discriminantes baseados nas métricas de curto prazo (CP) e cumulativas (C).

		Classificador					
		OneR		**Naive Bayes**		**Logistic**	
Atributo	**# classes**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**
s1 (Fórmula 6.1)	2 classes	43,90	48,78	53,65	58,53	60,97	**60,97**
	3 classes	35,29	41,17	**45,09**	37,25	39,21	43,13
s2 (Fórmula 6.2)	2 classes	43,90	48,78	53,65	**60,97**	63,41	**60,97**
	3 classes	35,29	41,17	41,17	37,25	37,25	39,21
s3 (Fórmula 6.3)	2 classes	43,90	53,65	56,09	**60,97**	63,41	**60,97**
	3 classes	43,13	47,05	41,17	37,25	37,25	39,21
s4 (Fórmula 6.4)	2 classes	53,65	58,53	53,65	56,09	58,53	**60,97**
	3 classes	39,21	33,33	43,13	**45,09**	45,09	**47,05**
s5 (Fórmula 6.5)	2 classes	53,65	56,09	**58,53**	51,21	58,53	51,21
	3 classes	45,09	43,13	43,13	43,13	43,13	43,13
s6 (Fórmula 6.6)	2 classes	**56,09**	**70,73**	48,78	53,65	**64,41**	58,53
	3 classes	**45,09**	**54,90**	**45,09**	**45,09**	**50,98**	**47,05**
s7 (Fórmula 6.7)	2 classes	43,90	39,02	46,34	51,21	53,65	53,65
	3 classes	41,17	43,13	43,13	41,17	43,13	43,13
Combinados	2 classes	51,21		46,34		56,09	
	3 classes	39,21		41,17		**47,05**	

Levando em consideração todos os resultados das Tabelas 6.2 e 6.2, o desempenho da classificação foi substancialmente melhor ao prever somente duas classes. O melhor resultado para predição de três classes foi de 54.90% utilizando o classificador OneR, comparado com 70.74% para classes binárias. A melhor precisão sempre foi obtida para a classe “aumentou”.

Entre os classificadores reportados, em geral, o OneR foi o que obteve os melhores resultados, tanto para os atributos discriminantes baseados em métricas (métrica s6) de sumarização, como para os atributos baseados em explosões de sentimento (ExplosõesDelta).

Portanto, em seguida, foram submetidos a cada classificador todos os atributos discriminantes em conjunto, ou seja, as métricas s1-s7 e as baseadas em explosões de sentimento, para a preparação de curto prazo e cumulativa. Para o classificador OneR, foi observado um mal desempenho, mostrando 51,21% e 37,25% de acurácia em predizer duas e três classes, respectivamente. O classificador Naive Bayes também mostrou um mal desempenho, com 41,46% de acurácia em predizer duas classes e 41,17% em três classes. Por último, os melhores resultados para esta combinação ficaram para o classificador Logistic com uma acurácia de 56,09% e 47,05% para a previsão duas e três classes, respectivamente. Resultados adicionais com todos os atributos discriminantes em conjuntos quanto a todos algoritmos de classificação testados são mostrados na Seção B.3 do Apêndice B deste trabalho.

Tabela 6.3 – Acurácia dos atributos discriminantes baseados nas métricas de explosão de sentimento de curto prazo (CP) e cumulativas (C).

		Classificador					
		OneR		**Naive Bayes**		**Logistic**	
Atributo	**# classes**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**
RazãoExplosões	2 classes	60,97	56,09	56,09	48,78	56,09	48,78
	3 classes	49,01	41,17	**49,01**	39,21	**50,98**	41,17
ExplosõesPositivas	2 classes	53,65	53,65	**60,97**	58,53	**60,97**	60,97
	3 classes	41,17	37,25	41,17	47,05	41,17	**49,01**
ExplosõesNegativas	2 classes	60,97	60,97	53,65	53,65	53,65	56,09
	3 classes	49,01	41,17	43,13	41,17	43,13	41,17
ExplosõesDelta	2 classes	**60,97**	**65,85**	58,53	**65,85**	**60,97**	**63,41**
	3 classes	**49,01**	**41,17**	45,09	**50,98**	43,13	**49,01**
Combinados	2 classes	43,90		53,65		48,78	
	3 classes	37,25		39,21		41,17	

6.3.3 Avaliação da Predição

Após analisarmos o poder preditivo de cada atributo discriminante proposto, foram realizados novos experimentos com dados do segundo turno da Eleição Presidencial de 2010 e da Eleição Municipal de 2012 como uma forma de validação da abordagem proposta. Para isto, os dados do segundo de turno foram utilizados para criar um conjunto de teste para o modelo preditivo, e os dados do primeiro turno foram utilizados como conjunto de treinamento. As eleições de segundo turno são realizadas aproximadamente um mês depois da eleição do primeiro turno.

A Eleição Governamental de 2010 não foi analisada, pois não houve segundo turno para esta eleição. Os candidatos analisados para o segundo turno das eleições são descritos abaixo, juntamente com seus respectivos partidos políticos:

- **Eleição Presidencial de 2010**: Dilma Rousseff (PT) e José Serra (PSDB);
- **Eleição Municipal de 2012**: Fernando Hadddad (PT) e José Serra (PSDB).

O perfil de dados referente ao segundo turno é detalhado na Tabela 6.4. Todo o processo de mineração de opiniões referente à análise do primeiro turno (Capítulo 5) foi igualmente aplicados aos dados do segundo turno. Os dados de intenção de votos do segundo turno da eleição de 2010 e 2012 são detalhados no Apêndice A deste documento, na Tabela A.3 e na Tabela A.5, respectivamente.

Tabela 6.4 – Perfil dos dados da base de dados do segundo turno das eleições analisada.

	Eleição de 2010 (2º turno)		**Eleição de 2012** (2º turno)	
	Bruto	**Pré-processado**	**Bruto**	**Pré-processado**
Número de notícias	1.407	1.182	1.213	849
Número de comentários	274.586	233.457	41.410	38.743
Média de comentários por notícia (DP)	195,15 (±366,15)	165,92 (±313,25)	34,13 (±90,55)	31,93 (±84,86)
Número de sentenças	-	720.569	-	103.547
Média de sentenças por comentários (DP)	-	3,08 (±2,06)	-	2,67 (±1,74)
Comentários muito curtos	5.375	0	904	0
Comentários quase duplicados	35.754	0	1.763	0
Período	04/10/2010 até 30/10/2010		07/10/2012 até 27/10/2012	
Entidades	2 candidatos		2 candidatos	

Primeiramente utilizamos os dados do primeiro turno de cada eleição para prever as variações de intenção de voto do segundo turno de sua respectiva eleição. Após isto, foram combinados os dados do primeiro turno de ambas as eleições como conjunto de treinamento, e os dados do segundo turno também de ambas as eleições como conjunto de teste. Para esta etapa, somente foi utilizado o classificador OneR, pois este foi o que obteve os melhores resultados com os dados do primeiro turno. Cada atributo discriminante foi submetido individualmente ao classificador, assim como foi feito nos experimentos referentes ao primeiro turno (Seção 6.3.2).

Nos experimentos com dados do segundo turno a classe "inalterada" não foi utilizada, pois para as eleições de 2012 não houveram dados com esta característica, e para as eleições de 2010 houve apenas um único registro no segundo turno. Para o segundo turno das eleições de 2010 existem sete registros para a classe "aumentou" e quatro registros para a classe "diminuiu". Para as eleições de 2012, são nove registros para a classe "aumentou" e três para a classe "diminuiu".

A Tabela 6.5 mostra os resultados de cada métrica de sumarização, e a Tabela 6.6 mostra os resultados dos atributos discriminantes baseados em explosões de expressão de sentimento. Os resultados destes experimentos para ambos os tipos de atributos discriminantes serão discutidos a seguir.

- **Eleição de 2010:** nesta eleição, praticamente todos os atributos discriminantes baseados em sumarização de sentimentos tiveram resultados semelhantes, onde foi atingida uma acurácia de 63,63%, tanto para a variação de curto prazo, como para a variação de longo prazo. Este mesmo comportamento também ocorreu para as métricas baseadas em

explosões de sentimento desta eleição.

- **Eleição de 2012:** na preparação de curto prazo desta eleição, utilizando os atributos baseados em sumarização de sentimento, atingiu-se um acurácia de 83,33% através das métricas s1, s2 e s3. Na preparação cumulativa, os melhores resultados foram utilizando as métricas s4, s5 e s7 (75% de acurácia). A acurácia de todos os atributos baseados em explosões de sentimento (em ambas as preparações) também foi de 75%.
- **Dados das eleições combinadas:** nesta variação foi obtida uma acurácia de 73,91% para a preparação de curto prazo, utilizando as métricas s1 e s2. Na preparação cumulativa, o melhor resultado foi utilizando a métrica s6, com uma acurácia de 78,26%. Para os atributos baseados em explosão de sentimento foi obtida uma acurácia de 69,56% em todos os atributos experimentados.

Finalmente, foi experimentada a submissão de todos os atributos e variações combinadas ao classificador simultaneamente, quando o observou-se que o desempenho se manteve ou piorou, mostrando uma acurácia de 45,45% para e eleição de 2010, 83,33% para a eleição de 2012, e 78,26% para a variação onde os dados das eleições foram combinados.

De forma geral, os resultados de classificação para os dados do segundo turno foram significativamente melhores quando comparados a utilização apenas dos dados do primeiro turno. A melhor acurácia nestes experimentos foi de 83,33%, contra 70,73% do primeiro turno. Os resultados do segundo turno devem ser observados com precaução, pois observou-se um viés nos resultados principalmente para a eleição de 2012, que foram afetados pela pequena quantidade de registros da classe "diminuiu". Portanto, quando os dados das eleições foram combinados houve um melhor balanceamento do conjunto de dados, o que de certa forma ajudou a minimizar este viés.

Tabela 6.5 – Acurácia dos atributos discriminantes baseados nas métricas de curto prazo (CP) e cumulativas (C) utilizando dados do segundo turno como conjunto de teste.

		Eleição					
		2010		**2012**		**Combinadas**	
Atributo	**# classes**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**
s1 (Fórmula 6.1)	2 classes	**63,63**	**63,63**	**83,33**	66,66	73,91	65,21
s2 (Fórmula 6.2)	2 classes	**63,63**	**63,63**	**83,33**	66,66	73,91	65,21
s3 (Fórmula 6.3)	2 classes	**63,63**	**63,63**	**83,33**	66,66	60,86	65,21
s4 (Fórmula 6.4)	2 classes	**63,63**	**63,63**	75,00	**75,00**	69,56	69,56
s5 (Fórmula 6.5)	2 classes	**63,63**	**63,63**	75,00	**75,00**	69,56	65,21
s6 (Fórmula 6.6)	2 classes	36,36	36,36	75,00	58,33	53,52	**78,26**
s7 (Fórmula 6.7)	2 classes	**63,63**	36,36	75,00	**75,00**	65,21	65,21
Combinados	2 classes	45,45		**83,33**		**78,26**	

Tabela 6.6 – Acurácia dos atributos discriminantes baseados nas métricas de explosão de sentimento de curto prazo (CP) e cumulativas (C) utilizando dados do segundo turno como conjunto de teste.

		Eleição					
		2010		**2012**		**Combinadas**	
Atributo	**# classes**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**	**CP (%)**	**C(%)**
RazãoExplosões	2 classes	45,45	54,54	75,00	75,00	69,56	69,56
ExplosõesPositivas	2 classes	63,63	63,63	75,00	75,00	69,56	69,56
ExplosõesNegativas	2 classes	63,63	63,63	75,00	75,00	69,56	69,56
ExplosõesDelta	2 classes	63,63	63,63	75,00	75,00	69,56	69,56
Combinados	2 classes	63,63		75,00		69,56	

6.3.4 Considerações

Todas métricas propostas no presente trabalho têm como objetivo representar uma dimensão do sentimento geral dos leitores sobre os candidatos analisados nas pesquisas eleitorais no mundo real. Por exemplo, a métrica s4 avalia se dentre os comentários feitos sobre um mesmo candidato, há uma tendência positiva ou negativa. O mesmo para as métricas baseadas em explosões de sentimento, que tentam capturar a percepção quanto a ações positivas e/ou escândalos que ocorreram durante a campanha eleitoral. No entanto, as métricas propostas ainda podem ser falhas nesta tarefa de retratar o sentimento geral destes leitores.

Nota-se também que, à exceção da métrica de menções (s7), todas as demais métricas de sumarização ou explosão de sentimentos são variações de relações entre sentimentos positivos e negativos. Portanto, não se pode assumir que o valor de um determinado atributo discriminante gerado a partir de uma métrica, seja independente do valor de qualquer outro atributo discriminante. Esta relação entre as métricas pode ter um impacto no poder preditivo de cada métrica e também no algoritmo de classificação utilizado, afetando consequentemente o desempenho da predição da variação de intenção de voto em geral.

No tocante aos algoritmos de classificação, resultados bastante variados foram obtidos levando em conta o atributo discriminante e a respectiva forma de preparação, sem revelar um padrão.

7 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho, foi examinado se o sentimento extraído de conteúdo gerado por usuários como reação a notícias políticas pode ser usado para prever variações de intenção de votos em pesquisas de opinião pública. Apesar do problema não ser novo, as caraterísticas que diferenciam o estudo de caso apresentado nesse trabalho foram: a fonte de opinião (comentário de notícias); comentários em português do Brasil, para o qual os recursos são escassos; e as variáveis a serem previstas eram séries temporais esparsas, pois são correspondentes a pesquisas de intenção de voto que são publicadas infrequentemente e irregularmente. Foi comparado o desempenho de duas abordagens de classificação de sentimento em comentários gerados por usuários (baseada em dicionário, e em aprendizado de máquina). Também foram propostos dois tipos de atributos discriminantes para representar a sumarização de sentimento e explosões de expressão de sentimento, onde foram preparados seguindo duas variações (efeito cumulativo e a curto prazo). A irregularidade dos dados a serem previstos foi abordada através da variação de intenção de votos de múltiplas eleições.

Em relação ao processo de mineração de opiniões, os resultados obtidos neste estudo de caso ainda não são satisfatórios. Na abordagem baseada em dicionário, a melhor acurácia observada foi de 52,14%. A indisponibilidade de bons dicionários de sentimento é a principal desvantagem desta abordagem. Além disso, o processo de aquisição de vocabulário específico de domínio é trabalhoso e sujeito a erros. Já na abordagem baseada em aprendizado de máquina, os resultados foram melhores, onde foi obtido uma acurácia de aproximadamente 83%. Nesta abordagem, o custo do processo de anotação de dados para dados de treinamento gera um impedimento para abordagem de aprendizado de máquina em situações reais. É necessário também melhorar o processo de análise de sentimentos por meio do tratamento de menções indiretas (e.g. pronomes), cláusulas, coreferência, negação e ironia. Além disso, técnicas de balanceamento de base de dados podem ser melhor estudadas para minimizar o problema da classificação de polaridade de sentenças positivas.

Também deve-se levar em consideração que podem existir questões quanto ao *gold-standard* utilizado neste trabalho, devido aos grupos distintos de anotadores. Apesar das instruções dadas aos grupos serem as mesmas, ainda podem existir diferenças no processo de anotação. O uso de uma métrica mais robusta para resolução de divergências, tais como Kappa, poderia ser investigada para minimizar este problema.

A abordagem utilizada neste trabalho trata a tarefa de previsão da variação da intenção de voto como um problema de classificação. Esta abordagem possui limitações quando compara-

dos aos métodos tradicionais (i.e., regressão, séries temporais), pois o resultado da predição não será um valor discreto, o que pode ser mais relevante quando o valor a ser predito são intenções de voto. No entanto, o método aplicado permitiu superar a desvantagem em se lidar com dados esparsos. Em relação à predição de variação de intenção de voto, foi atingida uma acurácia de 70% para o problema de classe binária (i.e. "aumentou", "diminuiu"), principalmente baseada no sentimento negativo, que foi detectado com uma confiança significativa. Diferentemente de outros trabalhos, menções para os candidatos revelaram pouco poder preditivo, comparado aos atributos discriminantes baseados em sentimento.

Em termos de representatividade, o uso exclusivo de comentários de jornais pode ser tão limitado quanto o uso do Twitter. Portanto, este trabalho deve ser tratado como um passo em direção a um *framework* mais genérico que é capaz de analisar comportamentos e executar previsões baseadas em sentimento. Os experimentos revelaram um comportamento de expressão de opinião totalmente diferente comparado com o Twitter, e possivelmente, eles representam uma população diferente (TUMITAN; BECKER, 2013). Ainda, estes resultados devem ser tratados com precaução, , pois o sentimento intrínseco nos comentários precisa refletir o sentimento geral público, caso contrário os resultados podem ser influenciados pelo grupo composto dos autores de comentários, que pode ser limitado por uma região demográfica ou uma classe social específica.

O trabalho resultou nos experimentos desenvolvidos, os quais podem guiar outros trabalhos futuros. Os resultados do trabalho também foram publicados em conferências nacionais e internacionais:

1. TUMITAN, D.; BECKER, K. Tracking Sentiment Evolution on User-Generated Content: a case study on the brazilian political scene. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS. Anais. . . [S.l.: s.n.], 2013. p.139–144.
2. BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: conceitos, aplicações e desafios. Simpósio Brasileiro de Banco de Dados, [S.l.], p.27–52, 2013.
3. TUMITAN, D.; BECKER, K. Sentiment-Based Features for Predicting Election Polls:a case study on the brazilian scenario. In: WEB INTELLIGENCE (WI) AND INTELLIGENT AGENT TECHNOLOGIES (IAT), 2014 IEEE/WIC/ACM INTERNATIONAL-JOINT CONFERENCES ON.Anais. . .[S.l.: s.n.], 2014. v.2, p.126–133.

Como trabalhos futuros, pode-se utilizar das eleições presidenciais e governamentais de 2014 para validar nossos resultados. Além disso, a abordagem proposta pode ser experimentada com outros indicadores governamentais esparsos, como popularidade ou aprovação de governo,

dados de censo por área crítica (e.g. saúde, educação), etc. Também é possível desenvolver mecanismos para integrar o sentimento expresso em várias mídias, cada uma com sua forma de expressão e representatividade. Alguns dos problemas que ainda precisam ser resolvidos são: quais técnicas são apropriadas para cada mídia, e seu tipo de comportamento de expressão de sentimento; como descobrir a representatividade da população que está interagindo através da mídia, e em que proporção a opinião deve ser levada em consideração na predição, entre outros. Outra importante linha de trabalho é visar as reações a notícias além dos comentários diretos, como sua repercussão no Facebook ou Twitter.

ApêndiceA PESQUISAS ELEITORAIS DE INTENÇÃO DE VOTO

Tabela A.1 – Intenção de voto do primeiro turno das eleições governamentais de 2010 da cidade São Paulo (Extraído de (DATAFOLHA, 2010a)).

	03/09	09/09	14/09	22/09	29/09	02/10
Geraldo Alckmin (PSDB)	50	49	51	51	49	50
Aloizio Mercadante (PT)	24	23	23	23	27	26

Tabela A.2 – Intenção de voto do primeiro turno das eleições presidenciais de 2010 (Extraído de (DATAFOLHA, 2010b)).

	03/09	09/09	15/09	22/09	27/09	28/09	02/10
Dilma (PT)	50	50	51	49	46	47	47
José Serra (PSDB)	28	27	27	28	28	28	29
Marina Silva (PV)	10	11	11	13	14	14	16

Tabela A.3 – Intenção de voto do segundo turno das eleições presidenciais de 2010 (Extraído de (DATAFOLHA, 2010c)).

	08/10	14/10	21/10	26/10	28/10	30/10
Dilma (PT)	48	47	50	49	50	51
José Serra (PSDB)	41	41	40	38	40	41

Tabela A.4 – Intenção de voto do primeiro turno das eleições municipais de 2012 da cidade São Paulo (Extraído de (DATAFOLHA, 2012a)).

	04/09	11/09	19/09	27/09	03/10	06/10
José Serra (PSDB)	21	20	21	22	23	24
Celso Russomanno (PRB)	35	32	35	30	25	23
Fernando Haddad (PT)	16	17	15	18	19	20

Tabela A.5 – Intenção de voto do segundo turno das eleições municipais de 2012 da cidade São Paulo (Extraído de (DATAFOLHA, 2012b)).

	10/10	18/10	24/10	27/10
Fernando Haddad (PT)	56	60	60	58
José Serra (PSDB)	44	40	40	42

ApêndiceB RESULTADOS ADICIONAIS DE EXPERIMENTOS

B.1 Resultados do Uso do Palavras para Classificação de Sentimento

Utilizando o Palavras, foram desenvolvidos experimentos para quebrar sentenças em cláusulas, tratamento de negações e o descobrimento do verdadeiro alvo das palavras de sentimento. No entanto, não foram obtidos bons resultados devido ao excesso de erros sintáticos e de estrutura, bem como uso de linguagem informal. Os resultados da classificação de sentimentos baseada em dicionário com o apoio do Palavras são mostrados na Tabela B.1.

Tabela B.1 – Acurácia (A), *Micro-Average* (Mi-A), *Macro-Average* (Ma-A), Precisão (P), Revocação (R) e medida-F (F) de experimentos de classificação de sentimento tratando o texto com o apoio do Palavras.

A(%)	**Mi-A(%)**	**Ma-A(%)**	**Polaridade**	**P (%)**	**R (%)**	**F(%)**
45,18	53,38	49,28	Positiva	30,43	62,82	41,00
			Negativa	89,87	42,32	57,55

B.2 Algoritmos Usados em Experimentos Previsão de Variação de Intenções de Voto

Os algoritmos de classificação testados nos experimentos de predição da variação de intenção de votos foram implementações disponíveis no Weka, a saber: Bayes Network, Naive Bayes, Naive Bayes Updateable, Logistic, Multilayer Perceptron, SMO, Simple Logistic, IBk, K*, LWL, AdaBoost M1, Attribute Selected Classifier, Bagging, CVParameterSelection, Classification Via Regression, Filtered Classifier, LogitBoost, MultiScheme, Random Committee, Random Subspace, Stacking, Vote, Decision Table, JRip, OneR, PART decision list, ZeroR, Decision Stump, J48, LMT (logistic model trees), REPTree (Fast decision tree learner), Random Forest, Random Tree.

B.3 Resultados Complementares da Previsão de Variação de Intenções de Voto Baseada em Métricas de Sumarização

Em relação à predição da variação de intenção de voto baseadas em métricas de sumarização (i.e. s1-s7), também foram submetidos a cada algoritmo testado o conjunto de todos os

atributos discriminantes, tanto para a preparação de curto prazo, quanto cumulativa. Os resultados considerando as preparações de curto prazo para a variação de três classes e duas classes são mostrados na Tabela B.2 e na Tabela B.3, respectivamente. Já os resultados considerando a preparação cumulativa quanto a variação de três classes são mostrados na Tabela B.4, enquanto que a Tabela B.5 mostra para a variação de duas classes.

Tabela B.2: Precisão (P), Revocação (R) e medida-F (F) de diversos algoritmos de classificação para previsão de variação de intenção de votos considerando as métricas de sumarização combinadas de curto prazo (problema ternário).

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Bayes Network	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Naive Bayes	Diminuiu	0,21	0,31	0,25
	Inalterada	0,4	0,22	0,28
	Aumentou	0,45	0,5	0,48
Naive Bayes Updateable	Diminuiu	0,21	0,31	0,25
	Inalterada	0,4	0,22	0,28
	Aumentou	0,45	0,5	0,48
Logistic	Diminuiu	0,22	0,15	0,18
	Inalterada	0,38	0,33	0,35
	Aumentou	0,35	0,45	0,39
Multilayer Perceptron	Diminuiu	0	0	0
	Inalterada	0,28	0,280	0,28
	Aumentou	0,32	0,45	0,38
SMO	Diminuiu	0	0	0
	Inalterada	0,33	0,06	0,1
	Aumentou	0,38	0,9	0,53
Simple Logistic	Diminuiu	0	0	0
	Inalterada	0,32	0,33	0,32
	Aumentou	0,4	0,6	0,48

Continua na próxima página

Tabela B.2 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
IBk	Diminuiu	0,25	0,23	0,24
	Inalterada	0,19	0,17	0,18
	Aumentou	0,35	0,4	0,37
K*	Diminuiu	0,27	0,23	0,25
	Inalterada	0,19	0,17	0,18
	Aumentou	0,38	0,45	0,41
LWL	Diminuiu	0	0	0
	Inalterada	0,36	0,5	0,42
	Aumentou	0,45	0,5	0,48
AdaBoost M1	Diminuiu	0,25	0,31	0,28
	Inalterada	0,35	0,5	0,41
	Aumentou	0,44	0,2	0,28
Attribute Selected Classifier	Diminuiu	0	0	0
	Inalterada	0,28	0,28	0,28
	Aumentou	0,37	0,55	0,44
Bagging	Diminuiu	0,33	0,31	0,32
	Inalterada	0,47	0,39	0,42
	Aumentou	0,42	0,5	0,45
CVParameterSelection	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Classification Via Regression	Diminuiu	0	0	0
	Inalterada	0,3	0,17	0,21
	Aumentou	0,36	0,7	0,47
Filtered Classifier	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
LogitBoost	Diminuiu	0,36	0,38	0,37
	Inalterada	0,45	0,28	0,34
	Aumentou	0,46	0,6	0,52

Continua na próxima página

Tabela B.2 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
MultiScheme	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Random Committee	Diminuiu	0,23	0,23	0,23
	Inalterada	0,25	0,22	0,24
	Aumentou	0,45	0,5	0,48
Random Subspace	Diminuiu	0,18	0,15	0,17
	Inalterada	0,33	0,44	0,38
	Aumentou	0,56	0,45	0,5
Stacking	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Vote	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Decision Table	Diminuiu	0	0	0
	Inalterada	0,2	0,06	0,09
	Aumentou	0,39	0,9	0,55
JRip	Diminuiu	0	0	0
	Inalterada	0,38	0,28	0,32
	Aumentou	0,39	0,75	0,52
OneR	Diminuiu	0,39	0,54	0,45
	Inalterada	0,33	0,28	0,3
	Aumentou	0,5	0,45	0,47
PART decision list	Diminuiu	0,2	0,08	0,11
	Inalterada	0,39	0,5	0,44
	Aumentou	0,35	0,4	0,37
ZeroR	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56

Continua na próxima página

Tabela B.2 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Decision Stump	Diminuiu	0,25	0,31	0,28
	Inalterada	0,35	0,5	0,41
	Aumentou	0,44	0,2	0,28
J48	Diminuiu	0,17	0,08	0,11
	Inalterada	0,33	0,39	0,36
	Aumentou	0,33	0,4	0,36
LMT	Diminuiu	0	0	0
	Inalterada	0,35	0,44	0,39
	Aumentou	0,32	0,35	0,33
REPTree	Diminuiu	0,12	0,08	0,1
	Inalterada	0,35	0,44	0,39
	Aumentou	0,45	0,45	0,45
Random Forest	Diminuiu	0,14	0,15	0,15
	Inalterada	0,28	0,22	0,25
	Aumentou	0,43	0,5	0,47
Random Tree	Diminuiu	0,3	0,23	0,26
	Inalterada	0,38	0,33	0,35
	Aumentou	0,52	0,65	0,57

Tabela B.3: Precisão (P), Revocação (R) e medida-F (F) de diversos algoritmos de classificação para previsão de variação de intenção de votos considerando as métricas de sumarização combinadas de curto prazo (problema binário).

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Bayes Network	Dimiuiu	0,33	0,08	0,12
	Aumentou	0,6	0,9	0,72
Naive Bayes	Dimiuiu	0,42	0,38	0,4
	Aumentou	0,62	0,65	0,63

Continua na próxima página

Tabela B.3 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Naive Bayes Updateable	Dimiuiu	0,42	0,38	0,4
	Aumentou	0,62	0,65	0,63
Logistic	Dimiuiu	0,2	0,15	0,17
	Aumentou	0,52	0,6	0,56
Multilayer Perceptron	Dimiuiu	0,47	0,54	0,5
	Aumentou	0,67	0,6	0,63
SMO	Dimiuiu	0	0	0
	Aumentou	0,61	1	0,75
Simple Logistic	Dimiuiu	0,25	0,15	0,19
	Aumentou	0,56	0,7	0,62
IBk	Dimiuiu	0,5	0,54	0,52
	Aumentou	0,68	0,65	0,67
K*	Dimiuiu	0,46	0,46	0,46
	Aumentou	0,65	0,65	0,65
LWL	Dimiuiu	0,45	0,69	0,55
	Aumentou	0,69	0,45	0,55
AdaBoost M1	Dimiuiu	0,47	0,54	0,5
	Aumentou	0,67	0,6	0,63
Attribute Selected Classifier	Dimiuiu	0,25	0,08	0,12
	Aumentou	0,59	0,85	0,69
Bagging	Dimiuiu	0,4	0,31	0,35
	Aumentou	0,61	0,7	0,65
CVParameterSelection	Dimiuiu	0	0	0
	Aumentou	0,61	1	0,75
Classification Via Regression	Dimiuiu	0	0	0
	Aumentou	0,55	0,8	0,65
Filtered Classifier	Dimiuiu	0,33	0,08	0,12
	Aumentou	0,6	0,9	0,72
LogitBoost	Dimiuiu	0,54	0,54	0,54
	Aumentou	0,7	0,7	0,7

Continua na próxima página

Tabela B.3 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
MultiScheme	Dimiuiu	0	0	0
	Aumentou	0,61	1	0,75
Random Committee	Dimiuiu	0,33	0,31	0,32
	Aumentou	0,56	0,6	0,59
Random Subspace	Dimiuiu	0,67	0,15	0,25
	Aumentou	0,63	0,95	0,76
Stacking	Dimiuiu	0	0	0
	Aumentou	0,61	1	0,75
Vote	Dimiuiu	0	0	0
	Aumentou	0,61	1	0,75
Decision Table	Dimiuiu	0,33	0,08	0,12
	Aumentou	0,6	0,9	0,72
JRip	Dimiuiu	0,33	0,08	0,12
	Aumentou	0,6	0,9	0,72
OneR	Dimiuiu	0,57	0,54	0,56
	Aumentou	0,71	0,75	0,73
PART decision list	Dimiuiu	0,25	0,08	0,12
	Aumentou	0,59	0,85	0,69
ZeroR	Dimiuiu	0	0	0
	Aumentou	0,61	1	0,75
Decision Stump	Dimiuiu	0,4	0,77	0,53
	Aumentou	0,62	0,25	0,36
J48	Dimiuiu	0,25	0,08	0,12
	Aumentou	0,59	0,85	0,69
LMT	Dimiuiu	0,28	0,31	0,3
	Aumentou	0,53	0,5	0,51
REPTree	Dimiuiu	0,4	0,15	0,22
	Aumentou	0,61	0,85	0,71
Random Forest	Dimiuiu	0,46	0,46	0,46
	Aumentou	0,65	0,65	0,65

Continua na próxima página

Tabela B.3 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Random Tree	Dimiuiu	0,33	0,23	0,27
	Aumentou	0,57	0,7	0,64

Tabela B.4: Precisão (P), Revocação (R) e medida-F (F) de diversos algoritmos de classificação para previsão de variação de intenção de votos considerando as métricas de sumarização combinadas cumulativas (problema ternário).

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Bayes Network	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,4	1	0,56
Naive Bayes	Diminuiu	0,36	0,31	0,33
	Inalterada	0,64	0,39	0,48
	Aumentou	0,45	0,65	0,53
Naive Bayes Updateable	Diminuiu	0,36	0,31	0,33
	Inalterada	0,64	0,39	0,48
	Aumentou	0,45	0,65	0,53
Logistic	Diminuiu	0,25	0,15	0,19
	Inalterada	0,42	0,44	0,43
	Aumentou	0,33	0,4	0,36
Multilayer Perceptron	Diminuiu	0,38	0,46	0,41
	Inalterada	0,31	0,28	0,28
	Aumentou	0,32	0,3	0,31
SMO	Diminuiu	0	0	0
	Inalterada	0,67	0,11	0,19
	Aumentou	0,4	0,95	0,56
Simple Logistic	Diminuiu	0,1	0,08	0,09
	Inalterada	0,5	0,22	0,31

Continua na próxima página

Tabela B.4 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
	Aumentou	0,39	0,65	0,49
IBk	Diminuiu	0,23	0,23	0,23
	Inalterada	0,39	0,39	0,39
	Aumentou	0,4	0,4	0,4
K*	Diminuiu	0,21	0,23	0,22
	Inalterada	0,38	0,28	0,32
	Aumentou	0,38	0,45	0,41
LWL	Diminuiu	0	0	0
	Inalterada	0,64	0,39	0,48
	Aumentou	0,45	0,85	0,59
AdaBoost M1	Diminuiu	0	0	0
	Inalterada	0,86	0,33	0,48
	Aumentou	0,45	1	0,62
Attribute Selected Classifier	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Bagging	Diminuiu	0,38	0,38	0,38
	Inalterada	0,67	0,56	0,61
	Aumentou	0,61	0,7	0,65
CVParameterSelection	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Classification Via Regression	Diminuiu	0,25	0,08	0,12
	Inalterada	0,5	0,06	0,1
	Aumentou	0,4	0,9	0,55
Filtered Classifier	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,4	1	0,56
LogitBoost	Diminuiu	0,28	0,31	0,3
	Inalterada	0,5	0,39	0,44

Continua na próxima página

Tabela B.4 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
	Aumentou	0,43	0,5	0,47
MultiScheme	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Random Committee	Diminuiu	0,36	0,38	0,37
	Inalterada	0,5	0,5	0,5
	Aumentou	0,37	0,35	0,36
Random Subspace	Diminuiu	0,4	0,31	0,35
	Inalterada	0,6	0,5	0,55
	Aumentou	0,46	0,6	0,52
Stacking	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Vote	Diminuiu	0	0	0
	Inalterada	0	0	0
	Aumentou	0,39	1	0,56
Decision Table	Diminuiu	0	0	0
	Inalterada	0,17	0,06	0,08
	Aumentou	0,4	0,9	0,55
JRip	Diminuiu	0	0	0
	Inalterada	0,6	0,33	0,43
	Aumentou	0,44	0,9	0,59
OneR	Diminuiu	0,46	0,46	0,46
	Inalterada	0,55	0,33	0,41
	Aumentou	0,52	0,7	0,6
PART decision list	Diminuiu	0,25	0,15	0,19
	Inalterada	0,56	0,28	0,37
	Aumentou	0,38	0,65	0,48
ZeroR	Diminuiu	0	0	0
	Inalterada	0	0	0

Continua na próxima página

Tabela B.4 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
	Aumentou	0,39	1	0,56
Decision Stump	Diminuiu	0	0	0
	Inalterada	0,86	0,33	0,48
	Aumentou	0,45	1	0,62
J48	Diminuiu	0,33	0,31	0,32
	Inalterada	0,6	0,33	0,43
	Aumentou	0,45	0,65	0,53
LMT	Diminuiu	0	0	0
	Inalterada	0,5	0,28	0,36
	Aumentou	0,38	0,6	0,46
REPTree	Diminuiu	0,33	0,15	0,21
	Inalterada	0,45	0,28	0,34
	Aumentou	0,41	0,7	0,52
Random Forest	Diminuiu	0,37	0,54	0,44
	Inalterada	0,47	0,44	0,46
	Aumentou	0,53	0,4	0,46
Random Tree	Diminuiu	0,35	0,46	0,4
	Inalterada	0,53	0,56	0,54
	Aumentou	0,4	0,3	0,34

Tabela B.5: Precisão (P), Revocação (R) e medida-F (F) de diversos algoritmos de classificação para previsão de variação de intenção de votos considerando as métricas de sumarização combinadas cumulativas (problema binário).

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
Bayes Network	Diminuiu	0	0	0
	Aumentou	0,57	0,9	0,71
Naive Bayes	Diminuiu	0,5	0,38	0,43

Continua na próxima página

Tabela B.5 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
	Aumentou	0,65	0,75	0,7
Naive Bayes Updateable	Diminuiu	0,5	0,38	0,43
	Aumentou	0,65	0,75	0,7
Logistic	Diminuiu	0,22	0,15	0,18
	Aumentou	0,54	0,65	0,59
Multilayer Perceptron	Diminuiu	0,31	0,31	0,31
	Aumentou	0,55	0,55	0,55
SMO	Diminuiu	0	0	0
	Aumentou	0,61	1	0,75
Simple Logistic	Diminuiu	0	0	0
	Aumentou	0,52	0,7	0,6
IBk	Diminuiu	0,33	0,23	0,27
	Aumentou	0,57	0,7	0,64
K*	Diminuiu	0,38	0,38	0,38
	Aumentou	0,6	0,6	0,6
LWL	Diminuiu	0,42	0,38	0,4
	Aumentou	0,62	0,65	0,63
AdaBoost M1	Diminuiu	0,62	0,62	0,62
	Aumentou	0,75	0,75	0,75
Attribute Selected Classifier	Diminuiu	0	0	0
	Aumentou	0,57	0,9	0,71
Bagging	Diminuiu	0,6	0,23	0,33
	Aumentou	0,64	0,9	0,75
CVParameterSelection	Diminuiu	0	0	0
	Aumentou	0,61	1	0,75
Classification Via Regression	Diminuiu	0	0	0
	Aumentou	0,57	0,9	0,71
Filtered Classifier	Diminuiu	0	0	0
	Aumentou	0,57	0,9	0,71
LogitBoost	Diminuiu	0,6	0,46	0,52

Continua na próxima página

Tabela B.5 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
	Aumentou	0,7	0,8	0,74
MultiScheme	Diminuiu	0	0	0
	Aumentou	0,61	1	0,75
Random Committee	Diminuiu	0,38	0,38	0,38
	Aumentou	0,6	0,6	0,6
Random Subspace	Diminuiu	1	0,15	0,27
	Aumentou	0,65	1	0,78
Stacking	Diminuiu	0	0	0
	Aumentou	0,61	1	0,75
Vote	Diminuiu	0	0	0
	Aumentou	0,61	1	0,75
Decision Table	Diminuiu	0	0	0
	Aumentou	0,57	0,9	0,71
JRip	Diminuiu	0,5	0,15	0,24
	Aumentou	0,62	0,9	0,73
OneR	Diminuiu	0,5	0,31	0,38
	Aumentou	0,64	0,8	0,71
PART decision list	Diminuiu	0,5	0,08	0,13
	Aumentou	0,61	0,95	0,75
ZeroR	Diminuiu	0	0	0
	Aumentou	0,61	1	0,75
Decision Stump	Diminuiu	0,33	0,38	0,36
	Aumentou	0,56	0,5	0,53
J48	Diminuiu	0,5	0,08	0,13
	Aumentou	0,61	0,95	0,75
LMT	Diminuiu	0	0	0
	Aumentou	0,5	0,65	0,56
REPTree	Diminuiu	0,75	0,23	0,35
	Aumentou	0,66	0,95	0,78
Random Forest	Diminuiu	0,4	0,31	0,35

Continua na próxima página

Tabela B.5 – Continuação da página anterior

Algoritmo	**Classe**	**P (%)**	**R (%)**	**F(%)**
	Aumentou	0,61	0,7	0,65
Random Tree	Diminuiu	0,4	0,46	0,43
	Aumentou	0,61	0,55	0,57

REFERÊNCIAS

AGGARWAL, C. C.; ZHAI, C. X. **Mining Text Data**. [S.l.]: Springer Publishing Company, Incorporated, 2012. ISBN 1461432227, 9781461432227.

ALVES, A. L. F. et al. A comparison of svm versus naive-bayes techniques for sentiment analysis in tweets: A case study with the 2013 fifa confederations cup. In: **Proceedings of the 20th Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: ACM, 2014. (WebMedia '14), p. 123–130. ISBN 978-1-4503-3230-9. Available from Internet: <http://doi.acm.org/10.1145/2664551.2664561>.

ARCHAK, N.; GHOSE, A.; IPEIROTIS, P. G. Show me the money!: Deriving the pricing power of product features by mining consumer reviews. In: **Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2007. (KDD '07), p. 56–65. ISBN 978-1-59593-609-7. Available from Internet: <http://doi.acm.org/10.1145/1281192.1281202>.

ASUR, S.; HUBERMAN, B. A. Predicting the future with social media. In: **Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01**. Washington, DC, USA: IEEE Computer Society, 2010. (WI-IAT '10), p. 492–499. ISBN 978-0-7695-4191-4. Available from Internet: <http://dx.doi.org/10.1109/WI-IAT.2010.63>.

AVANCO, L. V.; NUNES, M. das G. V. Lexicon-based sentiment analysis for reviews of products in brazilian portuguese. In: **2014 Brazilian Conference on Intelligent Systems, BRACIS 2014, Sao Paulo, Brazil, October 18-22, 2014**. [s.n.], 2014. p. 277–281. Available from Internet: <http://dx.doi.org/10.1109/BRACIS.2014.57>.

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: CHAIR), N. C. C. et al. (Ed.). **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)**. Valletta, Malta: European Language Resources Association (ELRA), 2010. ISBN 2-9517408-6-7.

BALAHUR, A.; KOZAREVA, Z.; MONTOYO, A. Determining the polarity and source of opinions expressed in political debates. In: **Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing**. Berlin, Heidelberg: Springer-Verlag, 2009, (CICLing '09). p. 468–480. ISBN 978-3-642-00381-3. Available from Internet: <http://dx.doi.org/10.1007/978-3-642-00382-0_38>.

BALAHUR, A. et al. Opinion mining on newspaper quotations. In: **Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 03**. Washington, DC, USA: IEEE Computer Society, 2009. (WI-IAT '09), p. 523–526. ISBN 978-0-7695-3801-3. Available from Internet: <http://dx.doi.org/10.1109/WI-IAT.2009.340>.

BALAHUR, A. et al. Sentiment analysis in the news. In: CHAIR), N. C. C. et al. (Ed.). **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)**. Valletta, Malta: European Language Resources Association (ELRA), 2010. ISBN 2-9517408-6-7.

BECKER, K.; TUMITAN, D. Introdução à mineração de opiniões: Conceitos, aplicações e desafios. **Simpósio Brasileiro de Banco de Dados**, Anais do 28º Simpósion Brasileiro de Banco de Dados, p. 27–52, 2013.

BERENDT, B.; NAVIGLI, R. Finding your way through blogspace: Using semantics for cross-domain blog analysis. In: **Computational Approaches to Analyzing Weblogs, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-03, Stanford, California, USA, March 27-29, 2006**. [s.n.], 2006. p. 1–8. Available from Internet: <http://www.aaai.org/Library/Symposia/Spring/2006/ss06-03-001.php>.

BICK, E. **The Parsing System "Palavras". Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**. Århus: University of Arhus, 2000.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, Elsevier, v. 2, n. 1, p. 1–8, 2011.

BOLLEN, J.; PEPE, A.; MAO, H. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. **CoRR**, abs/0911.1583, 2009.

BRAMER, M. A. **Principles of Data Mining, Second Edition**. [S.l.]: Springer, 2013. I-XIV, 1-440 p. (Undergraduate Topics in Computer Science). ISBN 978-1-4471-4883-8, 978-1-4471-4884-5.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series and forecasting**. [S.l.]: Taylor & Francis, 2002.

BRUCE, R. F.; WIEBE, J. M. Recognizing subjectivity: A case study in manual tagging. **Nat. Lang. Eng.**, Cambridge University Press, New York, NY, USA, v. 5, n. 2, p. 187–205, jun. 1999. ISSN 1351-3249. Available from Internet: <http://dx.doi.org/10.1017/S1351324999002181>.

CARLETTA, J. Assessing agreement on classification tasks: The kappa statistic. **Comput. Linguist.**, MIT Press, Cambridge, MA, USA, v. 22, n. 2, p. 249–254, jun. 1996. ISSN 0891-2017. Available from Internet: <http://dl.acm.org/citation.cfm?id=230386.230390>.

CARVALHO, P. et al. Clues for detecting irony in user-generated contents: Oh...!! it's "so easy";-). In: **Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion**. New York, NY, USA: ACM, 2009. (TSA '09), p. 53–56. ISBN 978-1-60558-805-6. Available from Internet: <http://doi.acm.org/10.1145/1651461.1651471>.

CESSIE, L. S.; HOUWELINGEN, J. C. van. Ridge Estimators in Logistic Regression. **Applied Statistics**, v. 41(1), p. 191–201, 1992.

CHAVES, M. et al. Pirpo: An algorithm to deal with polarity in portuguese online reviews from the accommodation sector. In: BOUMA, G. et al. (Ed.). **Natural Language Processing and Information Systems**. Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7337). p. 296–301. ISBN 978-3-642-31177-2. Available from Internet: <http://dx.doi.org/10.1007/978-3-642-31178-9_37>.

CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. **J. Artif. Int. Res.**, AI Access Foundation, USA, v. 16, n. 1, p. 321–357, jun. 2002. ISSN 1076-9757. Available from Internet: <http://dl.acm.org/citation.cfm?id=1622407.1622416>.

DATAFOLHA. **Intenção de voto para governador de São Paulo**. 2010. <http://media.folha.uol.com.br/datafolha/2013/05/02/intvoto_gov_sp_02102010.pdf>. Retrieved January 02, 2014.

DATAFOLHA. **Intenção de voto para presidente da República**. 2010. <http://media.folha.uol.com.br/datafolha/2013/05/02/intvoto_pres_02102010.pdf>. Retrieved January 02, 2014.

DATAFOLHA. **Intenção de voto para presidente da República 2º turno**. 2010. <http://media.folha.uol.com.br/datafolha/2013/05/02/intvoto_pres_30102010.pdf>. Retrieved January 02, 2014.

DATAFOLHA. **Intenção de voto para prefeito de São Paulo**. 2012. <http://media.folha.uol.com.br/datafolha/2013/05/02/int_voto_pref_sp_05102012.pdf>. Retrieved November 19, 2012.

DATAFOLHA. **Intenção de voto para prefeito de São Paulo 2º turno**. 2012. <http://media.folha.uol.com.br/datafolha/2013/05/02/int_voto_pref_sp_28102012.pdf>. Retrieved January 02, 2014.

DAVE, K.; LAWRENCE, S.; PENNOCK, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: **Proceedings of the 12th International Conference on World Wide Web**. New York, NY, USA: ACM, 2003. (WWW '03), p. 519–528. ISBN 1-58113-680-3. Available from Internet: <http://doi.acm.org/10.1145/775152.775226>.

GHANI, R. et al. Text mining for product attribute extraction. **SIGKDD Explor. Newsl.**, ACM, New York, NY, USA, v. 8, n. 1, p. 41–48, jun. 2006. ISSN 1931-0145. Available from Internet: <http://doi.acm.org/10.1145/1147234.1147241>.

GILBERT, E.; KARAHALIOS, K. **Widespread Worry and the Stock Market**. 2010. Available from Internet: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1513>.

GODBOLE, N.; SRINIVASAIAH, M.; SKIENA, S. Large-scale sentiment analysis for news and blogs. In: **Proceedings of the International Conference on Weblogs and Social Media (ICWSM)**. [S.l.: s.n.], 2007. v. 2.

GUERRA, P. H. C. et al. From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In: **Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2011. (KDD '11), p. 150–158. ISBN 978-1-4503-0813-7. Available from Internet: <http://doi.acm.org/10.1145/2020408.2020438>.

HALL, M. et al. The weka data mining software: An update. **SIGKDD Explor. Newsl.**, ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, nov. 2009. ISSN 1931-0145. Available from Internet: <http://doi.acm.org/10.1145/1656274.1656278>.

HOLTE, R. C. Very simple classification rules perform well on most commonly used datasets. **Mach. Learn.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 11, n. 1, p. 63–90, abr. 1993. ISSN 0885-6125. Available from Internet: <http://dx.doi.org/10.1023/A:1022631118932>.

HU, M.; LIU, B. Mining and summarizing customer reviews. In: **Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2004. (KDD '04), p. 168–177. ISBN 1-58113-888-1. Available from Internet: <http://doi.acm.org/10.1145/1014052.1014073>.

JINDAL, N.; LIU, B. Opinion spam and analysis. In: **Proceedings of the 2008 International Conference on Web Search and Data Mining**. New York, NY, USA: ACM, 2008. (WSDM '08), p. 219–230. ISBN 978-1-59593-927-2. Available from Internet: <http://doi.acm.org/10.1145/1341531.1341560>.

JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. In: **Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (UAI'95), p. 338–345. ISBN 1-55860-385-9. Available from Internet: <http://dl.acm.org/citation.cfm?id=2074158.2074196>.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing (2Nd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2009. ISBN 0131873210.

KIRCHGÄSSNER, G.; WOLTERS, J.; HASSLER, U. **Introduction to modern time series analysis**. [S.l.]: Springer, 2012.

KU, L.; LIANG, Y.; CHEN, H. Opinion extraction, summarization and tracking in news and blog corpora. In: **Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs**. [S.l.: s.n.], 2006.

LIU, B. Sentiment analysis and subjectivity. In: INDURKHYA, N.; DAMERAU, F. J. (Ed.). **Handbook of Natural Language Processing, Second Edition**. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2010. ISBN 978-1420085921.

LIU, B. **Sentiment Analysis and Opinion Mining**. Morgan & Claypool, 2012. (Synthesis digital library of engineering and computer science). ISBN 9781608458844. Available from Internet: <https://books.google.com.br/books?id=Gt8g72e6MuEC>.

LIU, Y. et al. Arsa: A sentiment-aware model for predicting sales performance using blogs. In: **Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**. New York, NY, USA: ACM, 2007. (SIGIR '07), p. 607–614. ISBN 978-1-59593-597-7. Available from Internet: <http://doi.acm.org/10.1145/1277741.1277845>.

LLOYD, L.; KECHAGIAS, D.; SKIENA, S. Lydia: A system for large-scale news analysis. In: CONSENS, M.; NAVARRO, G. (Ed.). **String Processing and Information Retrieval**. Springer Berlin Heidelberg, 2005, (Lecture Notes in Computer Science, v. 3772). p. 161–166. ISBN 978-3-540-29740-6. Available from Internet: <http://dx.doi.org/10.1007/11575832_18>.

MILLER, G. A. Wordnet: A lexical database for english. **Commun. ACM**, ACM, New York, NY, USA, v. 38, n. 11, p. 39–41, nov. 1995. ISSN 0001-0782. Available from Internet: <http://doi.acm.org/10.1145/219717.219748>.

NASCIMENTO, P. et al. Análise de sentimento de tweets com foco em notícias. **SBC**, 2012.

O'CONNOR, B. et al. From tweets to polls: Linking text sentiment to public opinion time series. In: **Proceedings of the Fourth International Conference on Weblogs and Social**

Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010. [s.n.], 2010. Available from Internet: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1536>.

ORENGO, V.; HUYCK, C. A stemming algorithm for the portuguese language. In: **String Processing and Information Retrieval, 2001. SPIRE 2001. Proceedings.Eighth International Symposium on**. [S.l.: s.n.], 2001. p. 186–193.

PAICE, C. D. Another stemmer. **SIGIR Forum**, ACM, New York, NY, USA, v. 24, n. 3, p. 56–61, nov. 1990. ISSN 0163-5840. Available from Internet: <http://doi.acm.org/10.1145/101306.101310>.

PAK, P. P. A. Twitter as a corpus for sentiment analysis and opinion mining. In: CHAIR), N. C. C. et al. (Ed.). **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)**. Valletta, Malta: European Language Resources Association (ELRA), 2010. ISBN 2-9517408-6-7.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Found. Trends Inf. Retr.**, Now Publishers Inc., Hanover, MA, USA, v. 2, n. 1-2, p. 1–135, jan. 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: **Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. (EMNLP '02), p. 79–86. Available from Internet: <http://dx.doi.org/10.3115/1118693.1118704>.

PENNEBAKER, J. W. et al. The development and psychometric properties of liwc2007. **Austin, TX, LIWC. Net**, 2007.

POPESCU, A.-M.; ETZIONI, O. Extracting product features and opinions from reviews. In: **Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, (HLT '05). p. 339–346. Available from Internet: <http://dx.doi.org/10.3115/1220575.1220618>.

REIS, E. **Estatística descritiva**. [S.l.]: Edições Sílabo, 2000. ISBN 9789726184768.

ROMANI, L. et al. A new time series mining approach applied to multitemporal remote sensing imagery. **Geoscience and Remote Sensing, IEEE Transactions on**, v. 51, n. 1, p. 140–150, Jan 2013. ISSN 0196-2892.

SARAWAGI, S. Information extraction. **Found. Trends databases**, Now Publishers Inc., Hanover, MA, USA, v. 1, n. 3, p. 261–377, mar. 2008. ISSN 1931-7883. Available from Internet: <http://dx.doi.org/10.1561/1900000003>.

SARMENTO, L. et al. Automatic creation of a reference corpus for political opinion mining in user-generated content. In: **Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion**. New York, NY, USA: ACM, 2009. (TSA '09), p. 29–36. ISBN 978-1-60558-805-6. Available from Internet: <http://doi.acm.org/10.1145/1651461.1651468>.

SCHMID, H. Probabilistic part-of-speech tagging using decision trees. In: **International Conference on New Methods in Language Processing**. Manchester, UK: [s.n.], 1994. p. 44–49.

SCHOEN, H. et al. The power of prediction with social media. **Internet Research**, v. 23, n. 5, p. 528–543, 2013.

SCHöLKOPF, B.; BURGES, C. J. C.; SMOLA, A. J. (Ed.). **Advances in Kernel Methods: Support Vector Learning**. Cambridge, MA, USA: MIT Press, 1999. ISBN 0-262-19416-3.

SILVA, M. J.; CARVALHO, P.; SARMENTO, L. Building a sentiment lexicon for social judgement mining. In: **Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language**. Berlin, Heidelberg: Springer-Verlag, 2012. (PROPOR'12), p. 218–228. ISBN 978-3-642-28884-5. Available from Internet: <http://dx.doi.org/10.1007/978-3-642-28885-2_25>.

SILVA, M. J. et al. The design of OPTIMISM, an opinion mining system for portuguese politics. In: **New Trends in Artificial Intelligence: Proceedings of EPIA 2009 - Fourteenth Portuguese Conference on Artificial Intelligence**. Universidade de Aveiro, 2009. Available from Internet: <http://epia2009.web.ua.pt/onlineEdition/565.pdf>.

SOUZA, M. et al. Construction of a portuguese opinion lexicon from multiple resources. In: **In 8th Brazilian Symposium in Information and Human Language Technology - STIL, Mato Grosso**. [S.l.: s.n.], 2011.

STONE, P. J.; DUNPHY, D. C.; SMITH, M. S. The general inquirer: A computer approach to content analysis. **Journal of Regional Science**, MIT press, 1966.

STRAPPARAVA, C.; VALITUTTI, A. Wordnet affect: an affective extension of wordnet. In: **In Proceedings of the 4th International Conference on Language Resources and Evaluation**. [S.l.: s.n.], 2004. v. 4, p. 1083–1086.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. [S.l.]: Addison Wesley, 2006. ISBN 0321321367.

THET, T. T.; NA, J.-C.; KHOO, C. S. Aspect-based sentiment analysis of movie reviews on discussion boards. **J. Inf. Sci.**, Sage Publications, Inc., Thousand Oaks, CA, USA, v. 36, n. 6, p. 823–848, dec. 2010. ISSN 0165-5515. Available from Internet: <http://dx.doi.org/10.1177/0165551510388123>.

TOUTANOVA, K. et al. Feature-rich part-of-speech tagging with a cyclic dependency network. In: **Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (NAACL '03), p. 173–180. Available from Internet: <http://dx.doi.org/10.3115/1073445.1073478>.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Min. Knowl. Discov.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 3, p. 478–514, may 2012. ISSN 1384-5810. Available from Internet: <http://dx.doi.org/10.1007/s10618-011-0238-6>.

TUMASJAN, A. et al. **Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment**. 2010. Available from Internet: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>.

TUMITAN, D.; BECKER, K. Tracking sentiment evolution on user-generated content: A case study on the brazilian political scene. In: PROCEEDINGS OF THE 28 SIMPÓSION BRASILEIRO DE BANCO DE DADOS. **Simpósio Brasileiro de Banco de Dados**. [S.l.], 2013. p. 139–144.

TUMITAN, D.; BECKER, K. Sentiment-based features for predicting election polls: A case study on the brazilian scenario. In: **Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on**. [S.l.: s.n.], 2014. v. 2, p. 126–133.

TURNEY, P. D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: **Proceedings of the 40th Annual Meeting on Association for Computational Linguistics**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. (ACL '02), p. 417–424. Available from Internet: <http://dx.doi.org/10.3115/1073083.1073153>.

WIEBE, J. et al. Learning subjective language. **Comput. Linguist.**, MIT Press, Cambridge, MA, USA, v. 30, n. 3, p. 277–308, sep. 2004. ISSN 0891-2017. Available from Internet: <http://dx.doi.org/10.1162/0891201041850885>.

WIEBE, J.; WILSON, T.; CARDIE, C. Annotating expressions of opinions and emotions in language. **Language Resources and Evaluation**, Kluwer Academic Publishers, v. 39, n. 2-3, p. 165–210, 2005. ISSN 1574-020X. Available from Internet: <http://dx.doi.org/10.1007/s10579-005-7880-9>.

YU, H.; HATZIVASSILOGLOU, V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: **Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (EMNLP '03), p. 129–136. Available from Internet: <http://dx.doi.org/10.3115/1119355.1119372>.