

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

MARCELO DOS SANTOS DIAS

**Detecção Não Supervisionada de  
Posicionamento em Textos de Tweets**

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

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Karin Becker

Porto Alegre  
2017

## CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Dias, Marcelo dos Santos

Detecção Não Supervisionada de Posicionamento em Textos de Tweets / Marcelo dos Santos Dias. – Porto Alegre: PPGC da UFRGS, 2017.

85 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2017. Orientador: Karin Becker.

1. Análise de Sentimento. 2. Detecção de Posicionamento. 3. Rotulação Automática. 4. Twitter. I. Becker, Karin. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitor: Prof<sup>a</sup>. Jane Fraga Tutikian

Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves

Diretor do Instituto de Informática: Prof<sup>a</sup>. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. João Luiz Dihl Comba

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

## **AGRADECIMENTOS**

Agradeço à minha esposa, Clarissa, pelo incentivo e apoio de sempre.

Agradeço à minha orientadora, Karin, pela orientação incansável, paciência e dedicação profissional.

Ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande do Sul, pela oportunidade de realizar o mestrado. Aos professores, por transmitir de forma especial seus ensinamentos. Aos meus colegas, por partilhar seus conhecimentos e experiências.

## RESUMO

Detecção de posicionamento é a tarefa de automaticamente identificar se o autor de um texto é favorável, contrário, ou nem favorável e nem contrário a uma dada proposição ou alvo. Com o amplo uso do Twitter como plataforma para expressar opiniões e posicionamentos, a análise automatizada deste conteúdo torna-se de grande valia para empresas, organizações e figuras públicas. Em geral, os trabalhos que exploram tal tarefa adotam abordagens supervisionadas ou semi-supervisionadas. O presente trabalho propõe e avalia um processo não supervisionado de detecção de posicionamento em textos de tweets que tem como entrada apenas o alvo e um conjunto de tweets a rotular e é baseado em uma abordagem híbrida composta por 2 etapas: a) rotulação automática de tweets baseada em um conjunto de heurísticas e b) classificação complementar baseada em aprendizado supervisionado de máquina. A proposta tem êxito quando aplicada a figuras públicas, superando o estado-da-arte. Além disso, são avaliadas alternativas no intuito de melhorar seu desempenho quando aplicada a outros domínios, revelando a possibilidade de se empregar estratégias tais como o uso de alvos e perfis somente dependendo das características de cada domínio.

**Palavras-chave:** Análise de Sentimento. Detecção de Posicionamento. Rotulação Automática. Twitter.

## Unsupervised Stance Detection in Texts of Tweets

### ABSTRACT

Stance Detection is the task of automatically identifying if the author of a text is in favor of the given target, against the given target, or whether neither inference is likely. With the wide use of Twitter as a platform to express opinions and stances, the automatic analysis of this content becomes of high regard for companies, organizations and public figures. In general, works that explore such task adopt supervised or semi-supervised approaches. The present work proposes and evaluates a non-supervised process to detect stance in texts of tweets that has as entry only the target and a set of tweets to classify and is based on a hybrid approach composed by 2 stages: a) automatic labelling of tweets based on a set of heuristics and b) complementary classification based on supervised machine learning. The proposal succeeds when applied to public figures, overcoming the state-of-the-art. Beyond that, some alternatives are evaluated with the intention of increasing the performance when applied to other domains, revealing the possibility of use of strategies such as using seed targets and profiles depending on each domain characteristics.

**Keywords:** Sentiment Analysis, Stance Detection, Automatic labelling, Twitter.

## LISTA DE FIGURAS

Figura 4.1 Fluxo do Processo Não Supervisionado de Detecção de Posicionamento em Tweets. ....	36
Figura 5.1 Cobertura das Regras x $F1_{Média}$ .....	53
Figura 5.2 Cobertura: Processo completo (1-RH-RS-SVM) x Apenas regras de sentimento (3-RS-SVM).....	54
Figura 5.3 Cobertura das Regras.....	56
Figura 5.4 Precisão das Regras (Média Ponderada): Este Trabalho (1-RH-RS-SVM) X WIC.....	57
Figura 5.5 Cobertura das Regras: Este Trabalho (1-RH-RS-SVM) X WIC.....	58
Figura 5.6 Precisão das Regras por Posicionamento .....	59
Figura 5.7 Precisão por Regra.....	59
Figura 5.8 Precisão da Tarefa de Detecção de Polaridade por Alvo .....	62
Figura A.1 Pcesso de Detecção Semi-supervisionada de Posicionamento em Tweets.....	82

## LISTA DE TABELAS

Tabela 1.1	Exemplos de formas de expressar posicionamento .....	13
Tabela 2.1	Exemplos de formas de expressar posicionamento .....	20
Tabela 3.1	Competidores do Semeval melhor colocados em cada Subtarefa .....	27
Tabela 3.2	Corpora Disponibilizados pelos organizadores do Semeval 2016.....	27
Tabela 3.3	Distribuição dos Corpora por Posicionamento e Polaridade .....	28
Tabela 3.4	Trabalhos de abordagem semi-supervisionada para detecção de posicionamento.	31
Tabela 4.1	Lista de regras por ordem de aplicação .....	37
Tabela 5.1	Experimento comparado aos baselines para a subtarefa B.....	49
Tabela 5.2	Experimento comparado aos Baselines para os corpora da subtarefa A .....	50
Tabela 5.3	Medida-F por Alvo X Classe.....	52
Tabela 5.4	Dicionários de hashtags de posicionamento por Alvo.....	55
Tabela 5.5	Resultados por Estapa e Alvo .....	63
Tabela 5.6	Instâncias de Treino do Modelo Preditivo .....	63
Tabela 5.7	Experimentos com Variações do Processo .....	64
Tabela 5.8	Componentes Utilizados nas Variações do Processo.....	65
Tabela 5.9	Resultados dos Experimentos.....	66
Tabela 5.10	Regras de Sentimento Contrário.....	69
Tabela 5.11	Perfis Semente utilizados para os experimentos 6-PSH-RH-RS-SVM e 7- RH-RS-PST-SVM.....	69
Tabela 5.12	Alvos Semente utilizados para o experimento 5-AS-RH-RS-RC-SVM .....	70
Tabela A.1	Regras para Rotulação Automática .....	84

## LISTA DE ABREVIATURAS E SIGLAS

AA	Atributos Adicionais
API	Application Programming Interface
AS	Alvos Semente
EMB	Embeddings
F1	Medida-F
POL	Detecção de Posicionamento por Polaridade
pp	pontos percentuais
PSH	Perfis Semente para Identificação de Hashtags
PST	Perfis Semente para Enriquecimento do Corpus de Treino
RC	Regras de Sentimento para Alvo Contrário
RF	Random Forest
RH	Regras baseadas em Hashtags
RS	Regras baseadas em Sentimento
SemEval International Workshop on Semantic Evaluation	
SL	Simple Logistic
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency
WIC	Web Intelligence Conference

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>11</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>17</b>
2.1 Análise de Sentimento	17
2.2 Análise de Sentimento em Tweets	18
2.3 Técnicas de Processamento de Linguagem Natural para Análise de sentimento	19
2.4 Métricas de Avaliação	20
2.5 Word Embeddings	21
<b>3 TRABALHOS RELACIONADOS</b>	<b>23</b>
3.1 Detecção de Posicionamento	23
3.2 Detecção de Posicionamento em Debates	23
3.3 Detecção de Posicionamento em Tweets	25
3.4 Semeval 2016	26
3.4.1 Os Corpora	27
3.4.2 A Métrica de Avaliação	28
3.5 Abordagens Supervisionadas de Detecção de Posicionamento em Tweets	29
3.6 Abordagens Semi-supervisionadas de Detecção de Posicionamento em Tweets	29
3.7 Considerações Finais	33
<b>4 PROCESSO DE DETECÇÃO NÃO SUPERVISIONADA DE POSICIONAMENTO EM TEXTOS DE TWEETS</b>	<b>35</b>
4.1 Visão Geral	35
4.2 Detecção de Posicionamento por Regras	39
4.2.1 Pré-processamento	39
4.2.2 Detecção de Posicionamento Preliminar	41
4.2.3 Detecção de Hashtags de Posicionamento	42
4.2.4 Detecção de Posicionamento	42
4.3 Detecção de Posicionamento por Classificação Complementar	43
4.3.1 Extração de atributos	43
4.3.2 Treinamento do Classificador	44
4.3.3 Classificação Complementar	45
4.3.4 Considerações Finais	45
<b>5 EXPERIMENTOS E ANÁLISE DE RESULTADOS</b>	<b>47</b>
5.1 Descrição dos Experimentos	47
5.2 Dados e Métricas	48
5.3 Avaliação de Desempenho	49
5.3.1 Experimento #1 - Baselines Semi-supervisionados	49
5.3.2 Experimento #2 - Baselines Supervisionados	50
5.3.3 Avaliação de Resultados dos Experimentos #1 e #2	51
5.4 Avaliação das Estratégias Usadas pelo Processo	52
5.4.1 Avaliação das Estratégias de Detecção de Posicionamento Baseado em Heurísticas	53
5.4.1.1 Cobertura das Heurísticas	53
5.4.1.2 Qualidade das Heurísticas	57
5.4.1.3 Detecção de Polaridade	60
5.4.2 Avaliação da Estratégia de Detecção de Posicionamento Complementar	61
5.5 Experimento #3 - Variações do Processo	64
5.5.1 Avaliação dos Componentes do Processo	66
5.5.2 Avaliação do Uso de Estratégias Semi-supervisionadas na Etapa de Heurísticas	67
5.5.3 Avaliação de Alternativas de Construção do Modelo Preditivo	71
5.6 Considerações Finais	74

<b>6 CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>76</b>
<b>REFERÊNCIAS .....</b>	<b>78</b>
<b>APPENDICES .....</b>	<b>81</b>
<b>APÊNDICEA .....</b>	<b>82</b>
<b>A.1 Visão Geral da Versão Anterior deste Trabalho .....</b>	<b>82</b>
<b>A.2 Criação de um Modelo Preditivo .....</b>	<b>82</b>
A.2.1 Identificação de N-gramas Alvo e Chave.....	82
A.2.2 Rotulação Automática de Tweets.....	83
A.2.3 Criação de um Modelo Preditivo de Posicionamento Utilizando Aprendizado Su- pervisionado.....	84
<b>A.3 Identificação de Posicionamento em Tweets não Rotulados .....</b>	<b>85</b>

## 1 INTRODUÇÃO

A análise de sentimento envolve a identificação automática de opiniões, sentimentos, avaliações e atitudes expressas por pessoas em linguagem escrita (LIU, 2012). Atualmente, os pesquisadores da área da análise de sentimento têm concentrado esforços em explorar informações disponíveis em redes sociais. Assim, são extraídas postagens, informações de perfis e outros dados de redes tais como Twitter ou Facebook, entre outras. Esses esforços advêm da variedade e da relevância das aplicações que podem ser construídas através do uso de técnicas de análise de sentimento aplicadas aos dados extraídos destas redes.

No escopo da análise de sentimentos, a tarefa de mineração de opinião tem grande relevância, sendo esta a área pioneira e mais explorada (LIU, 2012; TSYTSARAU; PALPANAS, 2012). Mais recentemente, pesquisas abordaram a identificação de outros sentimentos (MUNEZERO et al., 2014), tais como mineração de emoções (MOHAMMAD, 2016; BECKER; MOREIRA; SANTOS, 2017) e detecção de posicionamento (MOHAMMAD; SOBHANI; KIRITCHENKO, 2016; ANAND et al., 2011; RAJADESINGAN; LIU, 2014; WEI et al., 2016; ZARRELLA; MARSH, 2016).

A mineração de opinião consiste na análise de documentos visando a detecção do alvo da opinião e do sentimento expresso em relação a ele, medido em termos de *polaridade* (ou valência), que varia entre negativa e positiva (TSYTSARAU; PALPANAS, 2012).

Já o *posicionamento* é uma atitude cognitiva ou emocional em relação a um alvo (MOHAMMAD, 2016). Do ponto de vista de análise de sentimento, a detecção de posicionamento é a tarefa de identificar automaticamente se o autor de um texto é favorável, contrário ou nem favorável e nem contrário em relação a uma dada proposição ou alvo (MOHAMMAD et al., 2016b). O alvo pode ser uma pessoa, uma organização, uma política de governo, um movimento, uma causa, um produto, etc.

A detecção de posicionamento possui diversas aplicações (MOHAMMAD et al., 2016b), em particular quando envolve controvérsia. Exemplos são a análise de popularidade ou aprovação de um candidato a cargo eletivo, uma ideia (e.g. Mudança do clima é uma preocupação real), um movimento social (e.g. Feminismo), entre outros. Quando envolve documentos extraídos de redes sociais, permite que se avalie quantitativamente o posicionamento de porções consideráveis da população que expressam sua opinião, posicionamento e julgamento através destas plataformas. Quando aplicado a uma empresa, por exemplo, é possível avaliar o posicionamento geral quanto à sua marca ou práticas corporativas. Já um candidato a cargo eletivo pode explorar o posicionamento dos eleitores em relação a um tema de sua agenda política, ou

mesmo em relação à sua candidatura.

É importante estabelecer a diferença entre mineração de opinião e detecção de posicionamento. A mineração de opinião consiste em detectar se uma opinião expressa tem sentimento positivo ou negativo direcionado a um alvo de opinião. Contudo, ao analisar conjuntos de dados rotulados para polaridade e para posicionamento, Mohammad, Sobhani and Kiritchenko (2016) verificaram que não é possível inferir automaticamente que uma opinião positiva seja favorável a um dado alvo de posicionamento, nem que uma opinião negativa seja contrária a este. Além disso, é importante observar que o alvo da opinião pode não ser o mesmo alvo do problema de posicionamento.

Com efeito, a mineração de opinião é eficaz para a identificação de posicionamento somente se o alvo da opinião for o mesmo do posicionamento. Porém, a expressão de um posicionamento pode se dar através outras formas, como ilustrado através de exemplos reais na Tabela 1.1 em relação ao alvo Donald Trump<sup>1</sup>.

Além da expressão direta de um sentimento em relação ao alvo, um posicionamento pode ser expresso através de um sentimento em relação a um alvo relacionado, como no primeiro exemplo. No segundo exemplo da tabela, o alvo do sentimento positivo são os mexicanos, os quais recebem críticas contundentes de Trump devido à imigração ilegal, e, portanto, este tweet denota um posicionamento contrário a Trump. O terceiro exemplo menciona tema que denota clara oposição ao alvo através do tópico “Macy’s”. Esta empresa varejista descontinuou a venda de produtos da marca Trump por sua agenda política no que diz respeito aos imigrantes mexicanos, consequentemente, tweets que abordam este tópico denotam clara oposição a Trump. O posicionamento pode se apresentar também através da simples presença de uma hashtag, como no quarto exemplo, onde “#trumpforpresident” denota um posicionamento favorável apesar do texto ser neutro em relação à polaridade. Já o quinto exemplo, consiste em uma assertiva sem polaridade, na qual é possível verificar a ausência dos demais elementos descritos através dos exemplos anteriores. Um anotador humano poderia inferir que o tweet é contrário a Trump em função da referência a “Hitler”, mas é um caso de difícil detecção de posicionamento para um sistema automatizado por não haver polaridade clara ou hashtags. É necessário conhecer a semântica embutida no termo “Hitler”, e interpretar a analogia estabelecida, o que é um desafio para a própria mineração de opiniões. Esses exemplos evidenciam que a detecção de posicionamento vai além da mineração de opinião, e compartilha problemas em aberto.

---

<sup>1</sup>Exemplos extraídos dos corpora disponibilizados para a tarefa 6-B do Semeval 2016: <http://alt.qcri.org/semeval2016/task6/index.php?id=data-and-tools>

Tabela 1.1: Exemplos de formas de expressar posicionamento

Forma	Descrição	Tweet	Posicionamento
Mineração de Opinião	Opinião positiva para o alvo Trump	@realDonaldTrump your directness & honesty has moved you to the top of my list 4 #President2016! #TRUMPTRUTH	Favorável
Alvo Distinto	Opinião positiva para o alvo em apoio/oposição	WE Love <b>Mexicans</b> .WE Respect your Work Ethic.Your Love of Family.Your Loyalty.Your FOOD Your Love Of God! Lets ALL Get JOBS.VOTE	Contrário
Tópico	Presença de tópico que denota posicionamento	<b>Macy's</b> to stop selling Trump merch. I guess he'll sue them next.	Contrário
Hashtag	Presença de hashtag que denota um posicionamento	@NBCNews @realDonaldTrump or the white house!! <b>#trumpforpresident</b>	Favorável
Assertiva	Assertiva sem polaridade	Donald Trump is gonna end up like Hitler	Contrário

Fonte:o autor

Sistemas para detecção de posicionamento têm sido propostos na última década. Os trabalhos pioneiros utilizaram transcrições de discursos realizados por parlamentares no congresso americano (THOMAS; PANG; LEE, 2006) e, principalmente, sistemas de debates online (SOMASUNDARAN; WIEBE, 2009; ANAND et al., 2011). Mais recentemente, pesquisadores têm dado atenção ao tema da detecção de posicionamento em tweets, dado o avanço no uso das redes sociais.

O tema da detecção de posicionamento a partir do texto de tweets ganhou destaque a partir da tarefa 6 do International Workshop on Semantic Evaluation 2016 (Semeval 2016) (MOHAMMAD et al., 2016b). O desafio proposto consistiu na detecção de posicionamento baseada somente no conteúdo textual dos corpora de tweets fornecidos pelos organizadores da competição. A competição se dividiu em duas subtarefas: a) *sistemas supervisionados* (e.g. Zarrella and Marsh (2016), Wei et al. (2016), Boltuzic et al. (2016)) e b) *sistemas não supervisionados* (e.g. Wei et al. (2016), Dias and Becker (2016c), Krejzl and Steinberger (2016)). A esta segunda tarefa foi dada a possibilidade de utilização de uma pequena entrada adicional anotada, o que habilitou o desenvolvimento de soluções semi-supervisionadas, e, como consequência, esta abordagem foi adotada por todos os competidores. Os trabalhos que exploraram este desafio, assim como os conjuntos de dados fornecidos, tornaram-se *baselines* no tema da detecção de posicionamento em texto de tweets.

Trabalhos que utilizam abordagens supervisionadas (LAI et al., 2017; ZARRELLA; MARSH, 2016; WEI et al., 2016) costumam obter melhores resultados que aqueles que abordam o problema de forma semi-supervisionada. Contudo, a qualidade dos resultados depende da existência de um grande corpus rotulado especificamente para o domínio em análise, o que requer um grande esforço manual de anotação. Portanto, o custo subjacente à construção e aplicação de um sistema supervisionado tende a ser bastante alto e, não raramente, inviável,

pois se acrescenta ao custo a necessidade de grande conhecimento do domínio, de anotadores qualificados e um processo constante de anotação.

Já os trabalhos que empregaram abordagem semi-supervisionada no Semeval 2016, realizaram a construção dos sistemas ou processos de forma menos dispendiosa, recaindo em duas estratégias principais: a) expressões semente para a rotulação automática de tweets (DIAS; BECKER, 2016b; WEI et al., 2016; EBRAHIMI; DOU; LOWD, 2016b; EBRAHIMI; DOU; LOWD, 2016a; AUGENSTEIN et al., 2016), ou b) uso de um corpus *cross-target/cross-domain* (AUGENSTEIN; VLACHOS; BONTCHEVA, 2016; ZHANG; LAN, 2016; KREJZL; STEINBERGER, 2016). Os resultados da primeira estratégia são sensíveis à boa escolha de expressões semente, o que exige um bom conhecimento do domínio, e possivelmente até mesmo a inspeção manual de uma amostra para seleção de sementes representativas. A segunda estratégia, consiste na utilização de um corpus rotulado para outro alvo do mesmo domínio (*cross-target*) ou de outro domínio (*cross-domain*). Não se tem conhecimento da existência de corpus de treino de uso geral aplicável a diversos domínios, pois cada domínio possui vocabulário próprio. Além disso, da mesma forma que as abordagens supervisionadas, esta segunda estratégia recai na necessidade de disponibilidade ou de construção de um grande corpus rotulado.

O presente trabalho propõe e implementa um processo de detecção de posicionamento em texto de tweets, que inova na abordagem quando comparado à literatura existente por ser *não supervisionado*. A única entrada necessária ao processo é o alvo do problema de posicionamento (e.g. “Donald Trump”), além dos tweets a serem classificados. Isto elimina os problemas apontados acima para as abordagens supervisionadas e semi-supervisionadas. O processo proposto apresentou resultados promissores para alvos considerados figuras públicas, como Hillary Clinton e Donald Trump. O resultado obtido para o alvo Trump (59,71) superou o *baseline* semi-supervisionado estado-da-arte (58,03) (AUGENSTEIN et al., 2016). Já para o alvo Hillary, para o qual não existe um *baseline* semi-supervisionado, o resultado obtido (63,59) é comparável aos melhores resultados de trabalhos supervisionados (ZARRELLA; MARSH, 2016; WEI et al., 2016).

Para avaliar o processo desenvolvido em sua abrangência foram utilizados todos os seis corpora disponibilizados pelos organizadores da tarefa 6 do Semeval 2016 (MOHAMMAD et al., 2016b), a saber: candidatos a cargos eletivos (Hillary Clinton e Donald Trump), movimentos (Ateísmo, Movimento Feminista), uma causa (Legalização do Aborto) e uma proposição (Mudança do Clima é uma Preocupação Real).

O processo proposto tem como principais características:

- **Ser não supervisionado:** as únicas entradas necessárias são o alvo principal e o corpus a

ser rotulado;

- **Abordagem Híbrida:** combina heurísticas de posicionamento para rotulação automática de instâncias e aprendizado supervisionado para classificação complementar;
- **Heurísticas de posicionamento:** propomos dois tipos de regras, a saber: baseadas em hashtags de posicionamento (*regras de hashtags*) e em mineração de opinião (*regras de sentimento*). A mineração de opinião é realizada através da detecção de menção ao alvo principal do problema de posicionamento combinada com a detecção de polaridade. Já as hashtags de posicionamento são verificadas com base em um dicionário de hashtags de posicionamento extraído automaticamente. As regras baseadas em hashtags têm precedência sobre as regras de sentimento;
- **Dicionário de Hashtags de Posicionamento:** as hashtags de posicionamento são identificadas a partir da inspeção automática dos tweets. Os tweets são preliminarmente rotulados através da detecção de sentimento direcionado ao alvo do problema e são inspeccionados a fim de se extrair hashtags presentes majoritariamente em tweets de um dos lados do posicionamento, permitindo inferir o posicionamento da hashtag;
- **Classificação Complementar:** etapa complementar responsável por rotular tweets em que o posicionamento não se dá por opinião direcionada ao alvo nem por hashtags de posicionamento. Classifica estes tweets utilizando aprendizado supervisionado de máquina para detectar posicionamento baseado em alvos secundários e em outras expressões de posicionamento.

As seguintes questões de pesquisa nortearam o presente trabalho:

- Q1: Técnicas de mineração podem ser utilizadas como alternativa para realizar a tarefa de detecção de posicionamento em tweets?
- Q2: As características de um domínio influenciam a forma de expressar posicionamento?
- Q3: Existem características que definem posicionamentos e que possam ser utilizadas para propagar rótulos visando a construção automática de um corpus de treinamento para abordagem supervisionada?

As principais contribuições deste trabalho são:

- desenvolvimento de uma solução alternativa às usuais abordagens supervisionadas e semi-supervisionadas, permitindo realizar a tarefa de detecção de posicionamento com menor esforço e conhecimento do domínio;
- avanço no uso de técnicas de rotulação automática, exigindo uma entrada manual mínima para a execução do processo;
- desenvolvimento de uma abordagem para a detecção de posicionamento que se utiliza tão somente do conteúdo textual de tweets;
- avaliação da abrangência e da generalidade da solução proposta, aplicando-a a diversos domínios.

Uma versão preliminar deste trabalho foi submetida para a sub-tarefa 6-B do Semeval 2016, voltada a soluções não/semi-supervisionadas para um único alvo (i.e. Donald Trump), obtendo a terceira colocação dentre nove participantes (DIAS; BECKER, 2016c). A solução proposta era semi-supervisionada, exigindo expressões semente. Posteriormente, a pesquisa avançou ao avaliar a abrangência do processo considerando mais alvos de posicionamento, e melhorar sua eficácia (DIAS; BECKER, 2016a; DIAS; BECKER, 2016b). A presente proposta é resultado do esforço em diminuir a necessidade de conhecimento do domínio, e de inspeção/anotação de informações. Como já mencionado, este trabalho avança na pesquisa por requerer uma entrada manual mínima (apenas uma string representando o alvo), diferenciando-se dos demais trabalhos analisados que lidam com o problema de detecção de posicionamento em tweets demandando uma entrada de dados mais elaborada (WEI et al., 2016; EBRAHIMI; DOU; LOWD, 2016b; KREJZL; STEINBERGER, 2016; AUGENSTEIN et al., 2016; EBRAHIMI; DOU; LOWD, 2016a).

O restante deste trabalho está estruturado como segue. No Capítulo 2 é apresentada a fundamentação teórica necessária para o entendimento do trabalho. No Capítulo 3 são apresentados os trabalhos relacionados. O Capítulo 4 detalha o processo proposto. O Capítulo 5 descreve os experimentos realizados, discutindo os resultados obtidos. A conclusão e trabalhos futuros são endereçados no Capítulo 6.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os principais conceitos e técnicas utilizadas neste trabalho de forma resumida.

### 2.1 Análise de Sentimento

Análise de sentimento, também conhecida como mineração de opiniões, é o campo de pesquisa que analisa as opiniões, sentimentos, avaliações, atitudes e emoções direcionadas a entidades, tais como produtos, serviços, organizações, indivíduos, problemas, eventos, tópicos e seus atributos (LIU, 2012).

A mineração de opinião é a tarefa mais explorada na área da análise de sentimento e consiste em detectar se uma opinião direcionada a um alvo é negativa, positiva ou neutra. Além disso, pode ser identificada a intensidade da polaridade expressa.

A análise de sentimento tem como objeto um texto escrito, e pode ocorrer em vários níveis (LIU, 2012):

- **documento:** o documento como um todo é analisado. Por exemplo, a avaliação de um produto tem a polaridade global detectada como negativa, positiva ou neutra. Este nível de granularidade é interessante quando o documento trata de uma única entidade ou em documentos mais curtos, como tweets;
- **sentença:** cada sentença de um documento é analisada individualmente. Permite separar opiniões sobre diferentes alvos;
- **aspecto:** permite buscar opiniões sobre diferentes aspectos de uma mesma entidade. Por exemplo, para um determinado notebook (i.e. entidade), podem ser analisados os aspectos peso e desempenho e o sentimento direcionado a cada aspecto pode ser diferente. O notebook poderia ter peso leve (i.e. polaridade positiva) e desempenho ruim (i.e. polaridade negativa).

As abordagens de classificação utilizadas para detectar polaridade podem ser divididas em quatro grandes grupos (TSYTSARAU; PALPANAS, 2012):

- léxicas, com o uso de dicionários de sentimentos;

- aprendizado de máquina, com o uso predominante de técnicas de classificação ou de regressão;
- estatísticas, que se valem de técnicas para avaliar a coocorrência de termos;
- semânticas, que definem a polaridade de palavras em função de sua proximidade semântica com outras de polaridade conhecida.

Neste trabalho, a análise é realizada sobre o conteúdo textual de tweets em nível de documento, pois tweets apresentam textos curtos e os corpora analisados foram rotulados para posicionamento por tweet e não por sentença. Para a detecção de polaridade dos tweets neste trabalho, é utilizada uma API de detecção de polaridade que executa esta tarefa de forma transparente para o processo, detectando tanto a polaridade quanto a intensidade desta. A API utilizada é a IBM Alchemy API<sup>1</sup>.

## 2.2 Análise de Sentimento em Tweets

Documentos extraídos da plataforma Twitter (i.e. tweets) possuem características inerentes ao uso que, em geral, se faz desta rede social. Algumas dessas características são apresentadas a seguir:

- **Linguagem informal:** uso de gírias, vogais estendidas a fim de dar ênfase a palavra, palavras encurtadas para reduzir a digitação, uso de emoticons, etc;
- **Textos curtos:** apesar de permitir textos mais extensos, em função de uma característica inicial da ferramenta que limitava as postagens a 140 caracteres, os textos de tweets tendem a ser curtos e com pouco contexto quando analisados de forma isolada;
- **Uso de hashtags:** hashtags são amplamente utilizadas com o objetivo de reforçar opiniões e posicionamentos expressos através de tweets, prática que costuma dar contexto a textos curtos;
- **Uso de menções a perfis:** perfis do Twitter são explicitamente mencionados nas postagens através do uso do caractere “@”;
- **Atualização de tema:** os debates estão em constante mudança, o que pode fazer com que determinados temas passem a ser relevantes, ou percam a relevância a qualquer momento.

---

<sup>1</sup><https://www.ibm.com/watson/alchemy-api.html>

Essas características podem dificultar ou facilitar tarefas da área de análise de sentimento, como a mineração de opiniões e a detecção de posicionamento. O uso de linguagem informal pode dificultar o uso de técnicas usuais de classificação, como abordagem de n-gramas, pois palavras de mesmo significado podem ter a grafia modificada, por exemplo. Textos curtos reduzem a quantidade de atributos disponíveis para técnicas de classificação. Igualmente a mudança de tema pode ser um fator desafiador, já que modelos preditivos previamente construídos tem de se manter atualizados. Por outro lado, o uso de hashtags e o uso de menções podem servir de apoio nas tarefas de identificação de posicionamento e de detecção de alvo, respectivamente.

Neste trabalho, é utilizada a detecção de polaridade através de APIs de análise de sentimento combinada com a identificação de referências aos alvos de posicionamento, onde o objetivo é a detecção de hashtags que indiquem fortemente o posicionamento do autor de um tweet. Isso vai permitir que se amplie a possibilidade de detecção de posicionamento partindo de técnicas de mineração de opinião.

### 2.3 Técnicas de Processamento de Linguagem Natural para Análise de sentimento

A área de pesquisa de Processamento de Linguagem Natural dispõe de uma grande variedade de técnicas para enriquecer o conjunto de ferramentas para a execução de tarefas relacionadas à análise de sentimento (MOHAMMAD; KIRITCHENKO; ZHU, 2013). Entre as técnicas exploradas neste trabalho, estão:

- **tokenização:** descoberta dos termos que compõem as sentenças através da identificação de separadores, tais como caracteres de pontuação, tabulação, quebra de linha e espaço em branco;
- **normalização:** definir e converter termos para uma representação única. Por exemplo, menções a perfis do Twitter identificadas por @perfil podem ser convertidas para o texto `a_mention`, com o objetivo de reduzir a dimensionalidade de atributos para tarefas de classificação;
- ***term frequency-inverse document frequency (TF-IDF):*** representar os termos através de atributos numéricos considerando a relevância do termo dentro do conjunto de dados em análise.

## 2.4 Métricas de Avaliação

O presente trabalho realiza a classificação de um conjunto de tweets utilizando aprendizado supervisionado de máquina como ferramenta complementar do processo. Para isso, são utilizados algoritmos de classificação tais como SVM, *Random Forest* e *Simple Logistic*, bem como um Comitê de classificadores implementado através do meta-algoritmo Vote<sup>2</sup>.

Tarefas que têm como objetivo realizar a classificação de instâncias costumam ter sua avaliação dada pelas métricas Precisão, Revocação e Medida-F. A Tabela 2.1 apresenta uma matriz de confusão que permite entender melhor as equações utilizadas para o cálculo de tais métricas. A seguir, as métricas são apresentadas de forma resumida:

- **Precisão:** identifica o percentual de instâncias classificadas corretamente dentre as instâncias para as quais foi atribuída a classe em análise, conforme as Equações 2.1 e 2.2;
- **Revocação:** identifica o percentual de instâncias classificadas corretamente dentre as instâncias, as quais deveriam ter atribuída a classe em análise, conforme as Equações 2.3 e 2.4;
- **Medida-F (F1):** representa a média harmônica entre as métricas de precisão e revocação, conforme as Equações 2.5 e 2.6;
- **Medida-F Ponderada (F1<sub>Pond</sub>):** representa a média ponderada da medida-F, conforme Equação 2.7, onde #A representa a quantidade de instâncias da classe A e #B da classe B.

Tabela 2.1: Exemplos de formas de expressar posicionamento

		Classe Prevista	
		Classe A	Classe B
Classe Real	Classe A	Verdadeiro Positivos (VP)	Falso Negativos (FN)
	Classe B	Falso Positivos (FP)	Verdadeiro Negativos (VN)

Fonte: o autor

$$Prec_A = \frac{VP}{VP + FP} \quad (2.1)$$

$$Prec_B = \frac{VN}{VN + FN} \quad (2.2)$$

<sup>2</sup>Meta-algoritmo disponível na ferramenta Weka (HALL et al., 2009)

$$Revoc_A = \frac{VP}{VP + FN} \quad (2.3)$$

$$Revoc_B = \frac{VN}{VN + FP} \quad (2.4)$$

$$F1_A = 2 \times \frac{Prec_A \times Revoc_A}{Prec_A + Revoc_A} \quad (2.5)$$

$$F1_B = 2 \times \frac{Prec_B \times Revoc_B}{Prec_B + Revoc_B} \quad (2.6)$$

$$F1_{Pond} = \frac{\#A \times F1_A + \#B \times F1_B}{\#A + \#B} \quad (2.7)$$

Neste trabalho, essas métricas serão utilizadas para avaliar o desempenho dos componentes do processo. Já para comparar este trabalho aos *baselines*, será utilizada a média aritmética da F1 das classes favorável e contrária, representada pela Equação 2.8.

$$F1_{Média} = \frac{F1_A + F1_B}{2} \quad (2.8)$$

## 2.5 Word Embeddings

*Word embeddings* consistem em abordagens para construção e uso de representações de palavras de um vocabulário em um espaço vetorial. Dessa forma, dado um vocabulário, o *word embedding* consiste em um conjunto de pares formados por palavra e vetor de números reais, onde o vetor representa cada palavra em um espaço vetorial semântico. Assim, palavras semântica ou sintaticamente similares estarão localizadas com maior proximidade neste espaço vetorial (PENNINGTON; SOCHER; MANNING, 2014). Alguns trabalhos (e.g. Zarrella and Marsh (2016), Wei et al. (2016)) obtiveram bons resultados na tarefa da detecção de posicionamento em tweets fazendo uso de *word embeddings*.

Existem diferentes estratégias para construção desse espaço vetorial. Nos experimentos deste trabalho, foi utilizado um *embedding* pré-treinado de 200 dimensões<sup>3</sup> baseado em textos de tweets. Ele foi construído utilizando a abordagem Global Vectors (GloVe), que se baseia em uma matriz de coocorrências, considerando a frequência de coocorrência entre os termos do

---

<sup>3</sup><https://nlp.stanford.edu/projects/glove/>

conjunto de dados (PENNINGTON; SOCHER; MANNING, 2014).

Essa técnica foi utilizada neste trabalho com o objetivo de identificar atributos discriminantes que pudessem melhorar a qualidade do modelo preditivo para detecção de posicionamento em tweets construído pelo processo proposto.

### 3 TRABALHOS RELACIONADOS

Neste capítulo, são apresentados os trabalhos relacionados mais relevantes, analisando suas principais características para permitir a comparação com a abordagem de detecção de posicionamento apresentada nesta dissertação.

#### 3.1 Detecção de Posicionamento

A tarefa de detecção de posicionamento adotada neste trabalho é aquela proposta por Mohammad et al. (2016b) no Semeval 2016, a saber: "a detecção de posicionamento é a tarefa de identificar automaticamente se o autor de um texto é favorável, contrário ou nem favorável e nem contrário a uma dada proposição ou alvo".

Nas próximas seções serão apresentados os conjuntos de dados utilizados neste trabalho e alguns dos principais trabalhos desenvolvidos até o momento no tema da detecção de posicionamento.

#### 3.2 Detecção de Posicionamento em Debates

Os trabalhos pioneiros na área trataram do problema de detecção de posicionamento utilizando debates, tais como documentos de sites de debate online e fóruns (MURAKAMI; RAYMOND, 2010; ANAND et al., 2011; HASAN; NG, 2013) e, em menor quantidade, analisando transcrições de discursos do congresso americano (THOMAS; PANG; LEE, 2006).

De forma geral, essas plataformas e documentos apresentam características que permitem detectar o posicionamento do autor de um documento mais facilmente, entre elas:

- **Threads de discussão:** em geral, sites de debate implementam *threads* de discussão, onde um debatedor insere seu comentário relacionado a um comentário anterior na mesma *thread*. Esta característica permite relacionar comentários aumentando seu contexto individual. Em alguns sites, o usuário deve até mesmo informar se está refutando o comentário anterior, o que possibilita a explícita identificação de desacordo entre comentários, bem como o acordo entre comentários conectados indiretamente por transitividade de refutações;
- **Texto com mais informação textual:** estes documentos apresentam uma informação textual mais rica, contendo textos mais extensos e mais bem estruturados, em geral fazendo

menos uso de abreviações e gírias. Este vocabulário mais rico e padronizado facilita a aplicação de técnicas de classificação utilizando a abordagem de n-gramas, típica de classificação de documentos;

- **Necessidade de posicionamento:** trabalhos orientados a sites de debates partem de perguntas ou assertivas e os debatedores são convidados a participar opinando por um dos lados do debate. Neste caso não há a opção de se manter neutro. Por isso estes trabalhos abordam o problema lidando com a existência de um posicionamento e, como consequência, com a classificação em 2 classes, o que torna o problema mais simples.
- **Criação de corpus rotulado:** o voto de um congressista em votações de projetos relacionadas a determinado tema permite rotular de forma automática seus discursos, possibilitando a criação de um corpus rotulado com alta precisão para treinar um classificador. Da mesma forma, a maioria dos sites de debate solicitam que seja informado o posicionamento em relação ao alvo do debate pelo usuário, quando este insere um comentário textual. Neste caso também se tem um alto nível de precisão na anotação para formação de um corpus de treino;

Murakami and Raymond (2010) partem da hipótese de que em um debate online há uma tendência de as respostas entre usuários indicarem a refutação de ideias. Dessa forma, identificam pares de usuários com opiniões opostas, ou de usuários que concordam quanto a um tema por transitividade da oposição. Por exemplo, se “b” refuta “a” e “c” refuta “b”, então há um indício de que “c” e “a” concordam naquela discussão.

Anand et al. (2011) trabalham com a plataforma *Convinceme*<sup>1</sup> na qual é informado o lado do qual cada debatedor está posicionado, e que utiliza links de refutação entre mensagens em debates para identificar a oposição entre debatedores. Esses autores afirmam que o contexto é importante para a detecção de posicionamento em mensagens de debates. Isso ocorre em função de as refutações se referirem a mensagens anteriores e essa dependência dificultar a análise da mensagem sem o contexto. Mesmo para anotadores experientes, a simples análise do texto da refutação sem contexto se torna difícil.

Sridhar, Getoor and Walker (2014) utilizam atributos locais dos objetos (i.e. postagens) e conexões entre objetos (i.e. acordo e desacordo entre postagens) em conjunto com uma lista de regras modeladas em PSL (*Probabilistic Soft Logic*) para inferir o posicionamento das postagens fazendo uso de anotação manual de parte das postagens realizadas no site de debate.

---

<sup>1</sup><http://convinceme.net>

Hasan and Ng (2013) apresentam um método semi-supervisionado para detecção de posicionamento em sites de debates online. A partir de um conjunto de frases associadas a cada lado do posicionamento, buscam-se na web documentos que contenham as palavras dessas frases, formando um corpus de treino para construção de um classificador. Contudo esta abordagem implica em um certo esforço, pois há a necessidade de se construir um certo número de assertivas relacionadas com os dois lados do posicionamento, o que também requer um maior conhecimento do domínio em análise para a construção dessas assertivas.

Esses trabalhos partem de informações e características inerentes às plataformas citadas acima que diferem em muito daquelas encontradas na plataforma Twitter, tornando necessário que se avance na busca de abordagens alternativas.

### 3.3 Detecção de Posicionamento em Tweets

Mais recentemente, pesquisadores têm dado atenção ao problema da detecção de posicionamento em tweets (RAJADESINGAN; LIU, 2014). A detecção de posicionamento em tweets difere daquela em debates principalmente pelas características a seguir:

- As discussões se misturam em postagens (tweets) sem sequência, sobre variados tópicos, com/sem posicionamento, tornando difícil a tarefa de reconhecer *threads*;
- Apesar de eventualmente haver conexão entre tweets na forma de resposta (i.e. *reply*), não há uma informação que indique se o tweet em resposta a outro se caracteriza por ser uma refutação ou um comentário em acordo com o anterior, o que pode ser obtido em alguns sites de debate;
- Não há na plataforma a informação explícita do posicionamento relacionado ao texto, como há em certos sites de debate, onde o autor deve informar seu posicionamento ao realizar o comentário;
- Há uma quantidade menor de informações e atributos textuais a serem extraídos de tweets.

Alguns trabalhos exploraram as características estruturadas da plataforma Twitter. Rajadesingan and Liu (2014) utilizam a informação de replicação de mensagens lidas (retweets) para apoiar o seu processo de detecção do posicionamento, considerando o fato de dois usuários terem retuitado o mesmo tweet como um indício de que os autores possuem o mesmo posicionamento em relação ao alvo de posicionamento. Ebrahimi, Dou and Lowd (2016b), no entanto,

utilizam informações de retweets e conexões (seguidores) entre usuários da plataforma com o mesmo propósito. A partir do Semeval 2016, a maioria dos trabalhos abordaram a detecção de posicionamento baseando-se exclusivamente no conteúdo textual dos tweets, como será discutido no restante deste capítulo.

### 3.4 Semeval 2016

O Semeval 2016 (MOHAMMAD et al., 2016b) motivou muitos trabalhos na área, através da tarefa 6. Esta propôs a detecção de posicionamento em tweets baseada somente no seu conteúdo textual, onde, para um conjunto de tweets, deveria ser verificado o posicionamento de cada autor de tweet para um dado alvo. Três classes foram consideradas, a saber: favorável, contrário ou neutro. Foram propostas duas subtarefas:

- **Detecção supervisionada (A):** para esta sub tarefa, os competidores poderiam executar a detecção de posicionamento dos tweets através de um processo supervisionado, ou seja, poderiam ser utilizadas instâncias rotuladas para, por exemplo, treinar um classificador ou propagar rótulos. Para tal, os organizadores forneceram corpora para cinco alvos distintos: Legalização do Aborto (Aborto), Ateísmo, Mudança do Clima é uma Preocupação Real (Clima), Movimento Feminista (Feminismo) e Hillary Clinton (Hillary);
- **Detecção não supervisionada (B):** esta sub tarefa requeria o desenvolvimento de um sistema não supervisionado, onde era permitida uma anotação manual mínima, aceitando sistemas semi-supervisionados. A anotação mínima aceita poderia consistir em um conjunto de expressões semente ou em um pequeno conjunto de tweets rotulados (i.e. “uma mão cheia”). A sub tarefa “B” teve como objetivo a detecção de posicionamento em um conjunto de tweets relacionados especificamente ao alvo Donald Trump (Trump), onde o corpus fornecido era grande e não rotulado.

A Tabela 3.1 lista os competidores melhor colocados em cada tarefa junto com a colocação obtida por cada sistema. Após a competição, foram publicados outros trabalhos explorando os conjuntos de dados disponibilizados pelos organizadores do Semeval para a tarefa (EBRAHIMI; DOU; LOWD, 2016b; EBRAHIMI; DOU; LOWD, 2016a; AUGENSTEIN et al., 2016; LAI et al., 2017), utilizando os vencedores da competição (WEI et al., 2016; ZARRELLA; MARSH, 2016), e o próprio trabalho dos organizadores (SOBHANI; MOHAMMAD; KIRITCHENKO, 2016), como *baselines*. Como consequência, os trabalhos publicados utili-

Tabela 3.1: Competidores do Semeval melhor colocados em cada Subtarefa

Trabalho	Colocação	
	Subtarefa A	Subtarefa B <sup>2</sup>
Dias and Becker (2016c)		3
Zarella and Marsh (2016)	1	
Wei et al. (2016)	2	1
Boltuzic et al. (2016)	3	
Zhang and Lan (2016)	5	5
Krejzl and Steinberger (2016)	9	4

Fonte: (MOHAMMAD et al., 2016b)

zando estes corpora e os resultados obtidos por estes se tornaram, respectivamente, *benchmarks* e *baselines* para trabalhos da área de detecção de posicionamento.

### 3.4.1 Os Corpora

Os organizadores do Semeval 2016 disponibilizaram seis (6) conjuntos de dados para a realização das tarefas (MOHAMMAD et al., 2016b; MOHAMMAD et al., 2016a), listados na Tabela 3.2, junto com a respectiva quantidade de tweets.

Para cada um dos seis alvos, foram preparados um corpus de treino com dados rotulados para posicionamento e um corpus de teste. Exceto para o alvo Trump, os corpora de treino para os demais cinco alvos eram anotados para posicionamento. Já os corpora de teste não eram anotados, pois foram utilizados para avaliar as soluções desenvolvidas. Ao final da competição, foram divulgados os rótulos de posicionamento para os corpora de teste, além de mais dois tipos de rótulo para os cinco corpora de treino e seis corpora de teste: rótulos de polaridade e rótulos indicando se a opinião de cada tweet é direcionada ao alvo do posicionamento, a outro alvo ou a nenhum alvo.

Tabela 3.2: Corpora Disponibilizados pelos organizadores do Semeval 2016

Corpora	Treino/Domínio	Teste	Total
Aborto	653	280	933
Ateísmo	513	220	733
Clima	395	169	564
Feminismo	664	285	949
Hillary	689	295	984
Trump	78.156	707	78.863

Fonte: (MOHAMMAD et al., 2016a)

<sup>2</sup>o segundo colocado da subtarefa não publicou o artigo descrevendo sua solução nos anais da conferência

A Tabela 3.3 apresenta a distribuição dos tweets por posicionamento e por polaridade<sup>3</sup>. Pode ser percebida uma predominância do posicionamento contrário ao alvo principal de posicionamento, com exceção do corpus Clima, onde predomina o posicionamento favorável ao alvo principal. Já quanto à característica de sentimento dos tweets, há uma predominância da polaridade negativa para todos os alvos de posicionamento. A única exceção ocorre para o alvo Ateísmo. Para este alvo, é possível perceber uma tendência dos autores de tweets em se posicionar de forma contrária ao alvo de posicionamento. Isso ocorre através de mensagens contendo sentimento positivo direcionado a alvos de opinião relacionados à religião (e.g. “god” e “jesus”).

Tabela 3.3: Distribuição dos Corpora por Posicionamento e Polaridade

Corpus	Instâncias	Posicionamento						Polaridade					
	Total	Contrário		Favorável		Neutro		Negativo		Positivo		Neutro	
	Qtd.	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%	Qtd.	%
<b>Atheism</b>	<b>733</b>	464	63,3	124	16,92	145	19,78	258	35,2	440	60,03	35	4,77
<b>Climate</b>	<b>564</b>	26	4,61	335	59,4	203	35,99	283	50,18	175	31,03	106	18,79
<b>Feminism</b>	<b>949</b>	511	53,85	268	28,24	170	17,91	730	76,92	174	18,34	45	4,74
<b>Abortion</b>	<b>933</b>	544	58,31	167	17,9	222	23,70	634	67,95	245	26,26	54	5,79
<b>Hillary</b>	<b>984</b>	565	57,42	163	16,57	256	26,06	648	65,85	297	30,18	39	3,96
<b>Trump</b>	<b>707</b>	299	42,29	148	20,93	260	36,78	481	68,03	193	27,3	33	4,67

Fonte: <http://saifmohammad.com/WebPages/StanceDataset.htm>

### 3.4.2 A Métrica de Avaliação

A métrica para avaliação dos sistemas utilizada pelos organizadores do Semeval foi a média entre as medidas F das classes de posicionamento favorável e contrária ( $F1_{Média}$ ). Apesar da sub tarefa A consistir de 5 alvos distintos, foi considerado o vencedor da competição aquele com o maior valor para a métrica calculada reunindo todas as instâncias de teste em um corpus único. Para a sub tarefa B, a métrica foi calculada em cima dos posicionamentos detectados para as 707 instâncias de teste do alvo Donald Trump.

Os trabalhos que abordam a tarefa de detecção de posicionamento em tweets utilizam esses corpora, essa métrica e os trabalhos dos competidores do Semeval como *baselines* para avaliação de seus sistemas quanto ao desempenho.

<sup>3</sup>Informações extraídas do dashboard publicado pelos organizadores do Semeval 2016 em <http://www.saifmohammad.com/WebPages/StanceDataset.htm>

### 3.5 Abordagens Supervisionadas de Detecção de Posicionamento em Tweets

A maioria dos trabalhos exploram o problema da detecção de posicionamento em textos de tweets através de técnicas de aprendizado supervisionado de máquina (LAI et al., 2017; ZARRELLA; MARSH, 2016; SOBHANI; MOHAMMAD; KIRITCHENKO, 2016; IGARASHI et al., 2016; VIJAYARAGHAVAN et al., 2016; PATRA; DAS; BANDYOPADHYAY, 2016; BOLTUZIC et al., 2016; LIU et al., 2016; ELFARDY; DIAB, 2016; MISRA et al., 2016; BØHLER et al., 2016), tendo como base um corpus extenso rotulado para o desenvolvimento de um classificador. Nesses trabalhos são utilizadas diversas técnicas para pré-processamento e engenharia de atributos, como etapa prévia ao uso de um algoritmo de classificação para a criação de um modelo.

Em termos de engenharia de atributos, são diversas as opções utilizadas por sistemas que se propõem a realizar a tarefa de detecção de posicionamento. Abaixo são listadas algumas opções amplamente utilizadas:

- n-gramas de palavras representadas de forma binária ou TF-IDF;
- n-gramas de caracteres representadas de forma binária ou TF-IDF;
- atributos extraídos através do uso de léxicos de sentimento com base nas palavras que compõem o documento;
- propriedades gramaticais dos termos (e.g uso de prefixo para expressar *Part-of-speech*).

Trabalhos mais recentes (ZARRELLA; MARSH, 2016; WEI et al., 2016; AUGENSTEIN et al., 2016; AUGENSTEIN; VLACHOS; BONTCHEVA, 2016; ZHANG; LAN, 2016) utilizaram *word embeddings* para codificar os tweets para treino e classificação, através da aplicação de aprendizado profundo, obtendo resultados bastante positivos. Zarrella and Marsh (2016) foram vencedores da competição para a sub tarefa “A” e Wei et al. (2016) foram vencedores da sub tarefa “B”, enquanto Augenstein et al. (2016) obtiveram o melhor resultado dos trabalhos semi-supervisionados, tornando-se o principal *baseline*.

### 3.6 Abordagens Semi-supervisionadas de Detecção de Posicionamento em Tweets

Trabalhos que utilizam abordagens semi-supervisionadas para detecção de posicionamento em texto de tweets são caracterizados por uma entrada manual mínima. Esses trabalhos podem ser classificados em dois grandes grupos de acordo com o tipo de entrada:

- **Expressões semente:** esta entrada pode corresponder a palavras, expressões regulares ou hashtags semente rotuladas para posicionamento. Essas expressões rotuladas podem ser utilizadas para rotular automaticamente instâncias de um corpus pela simples verificação de sua presença nos textos dos tweets, para coletar corpora adicional no Twitter por exemplo, ou, ainda, para formar um dicionário de posicionamento identificando atributos para construção de um classificador;
- **Classificador *cross-domain/cross-target*:** abordagem definida pelo uso de um corpus anotado para outro domínio *cross-domain* (LIU, 2012), ou um corpus anotado para outro alvo do mesmo domínio *cross-target* como, por exemplo, Hillary e Trump que são figuras públicas do meio político.

Os trabalhos detalhados neste capítulo aplicaram processos semi-supervisionados ao alvo Trump permitindo que seja estabelecida uma comparação mais direta entre estas abordagens. A Tabela 3.4 apresenta um resumo das principais características desses trabalhos no que diz respeito à entrada necessária para a execução de cada processo, ao uso de fontes de dados externas, ao processamento e à abordagem de classificação, e que serão detalhadas no restante desta seção.

Nos trabalhos estudados, os classificadores *cross-domain/cross-target* desenvolvidos são aplicados ao alvo Trump e são construídos com base nos corpora da subtarefa 6-A do Semeval 2016 (AUGENSTEIN; VLACHOS; BONTCHEVA, 2016; ZHANG; LAN, 2016; KREJZL; STEINBERGER, 2016). Os temas desta subtarefa são fortemente relacionados à agenda política de Donald Trump e um dos alvos (Hillary Clinton) é uma figura pública do meio político. Isso pode indicar uma proximidade entre os tweets dos corpora da subtarefa 6-A do Semeval e os tweets do corpus Trump.

Augenstein, Vlachos and Bontcheva (2016) utilizam os corpora rotulados para os 5 alvos da subtarefa A do Semeval para treinar um autocodificador de *bag-of-words*, a fim de obter atributos para desenvolver um classificador *cross-domain* utilizando Regressão Logística. Utilizam como atributos os vetores de 100 dimensões obtidos através do autocodificador aplicado aos corpora rotulados. Um autocodificador é um algoritmo de aprendizado não supervisionado, baseado em redes neurais, que aplica *backpropagation*, definindo os valores alvo como iguais às entradas.

Zhang and Lan (2016) dividem a sua abordagem em duas etapas através da construção e sequenciamento de dois classificadores. O primeiro verifica a relevância do tweet em relação ao alvo principal do problema para detectar e descartar tweets neutros. O segundo detecta o posicionamento dos tweets relevantes, classificando-os para o posicionamento favorável ou

Tabela 3.4: Trabalhos de abordagem semi-supervisionada para detecção de posicionamento

Trabalho	Descrição	Entradas para o Processo		Fontes de Dados Externas	Processamento	Abordagem de Classificação	
		Expressões Semente	Corpora Rotulado para a Subtarefa A			Atributos	Estratégia de Classificação
Augenstein, Vlachos and Bontcheva (2016)	Treina um auto-codificador de bags-of-words para obter atributos e treinar um classificador de regressão logística	Não	Sim	Não	Não	Vetor de 100 dimensões baseado no auto-codificador aplicado aos corpora de treino	Regressão Logística
Zhang and Lam (2016)	Organiza a tarefa em 2 etapas através da construção de dois classificadores: um para analisar a relevância dos Tweets (de-sartar neutros), e um segundo classificador para detectar o posicionamento.	Não	Sim	Não	Não	Linguísticos Tradicionais, Detecção de Tópicos, Similidade, Léxico de Sentimento, Específicos para o Twitter, Relacionados a Embeddings	Regressão Logística
Krejzl and Steinberger (2016)	Utiliza o corpus anotado para o alvo Hillary Clinton para treinar um classificador de Entropia Máxima	Identifica hashtags frequentes em um dos lados do posicionamento a partir dos corpora rotulados para formar um dicionário de posicionamento baseado nestas hashtags	Apenas Hillary	Não	Não	Bag-of-words, léxicos de sentimento e dicionário de posicionamento obtido a partir das hashtags extraídas dos corpora	Entropia Máxima
Dias and Becker (2016b)	Roula tweets automaticamente a partir de um conjunto de regras baseadas em expressões semente e polaridade combinadas formando um corpus para treinar um classificador complementar	Lista de expressões semente para rotular tweets automaticamente e compor um corpus de treino	Não	Não	Aplica um conjunto de regras baseadas em alvos, palavras chave e alvos contrários, expressões semente e polaridade, rotulando os tweets de forma automática	N-gramas de 1 a 4	SVM
Wei et al. (2016)	Roula automaticamente tweets do corpus de domínio pela presença expressões semente e utiliza estes tweets rotulados juntamente com embeddings para treinar uma rede convolucional que	Lista de expressões semente para rotular tweets automaticamente e compor um corpus de treino	Não	Embeddings previamente treinados a partir de notificações do Google News	Não	Extraí vetores para as palavras a partir dos Embeddings do Google News	Redes Convolucional
Ebrahimi, Dou and Lowd (2016a)	gera rotulos iniciais para tweets com base em expressões e hashtags semente e propaga rotulos entre perfis e tweets a partir da co-ocorrência de retweets entre perfis utilizando este corpus para treinar um classificador SVM	Lista de expressões semente extraídas do corpus de domínio a partir de inspeção	Não	Coleta seguidores dos perfis que emitiram os tweets do corpus de domínio da Subtarefa B e seus tweets	Relaciona perfis do Twitter que retuitaram os mesmos tweets propagando rotulos	Bag-of-words(f-idf), Léxicos de Sentimento e API de sentimento	SVM
Ebrahimi, Dou and Lowd (2016b)	desevolvem um modelo de classificação para aplicar a tweets do corpus de domínio de Trump rotulados a partir da verificação da presença de expressões e hashtags semente, utilizando bag-of-words, utilizando léxicos de sentimento e API de detecção de sentimento para a geração de atributos	Lista de expressões semente para rotular tweets automaticamente e compor um corpus de treino	Não	Não	Não	Bag-of-words, API de sentimento, Identificação de Tópicos	Modelo Generativo Sentimento-alvo-posicionamento
Augenstein et al. (2016)	Utiliza o corpus de domínio de Trump rotulando estes tweets de forma automática com base em expressões extraídas através de inspeção, coleta tweets para treinar embeddings e utilizar em uma rede LSTM juntamente com o corpus rotulado	Lista de expressões semente extraídas do corpus de domínio a partir de inspeção	Não	Coleta tweets e utiliza para treinar embeddings	Não	Utiliza os embeddings treinados pelo processo para extrair atributos dos corpora rotulados	Rede LSTM

Fonte: o autor

contrário. O processo desenvolvido utiliza os corpora rotulados para a sub tarefa 6-A para a construção de um classificador *cross-domain* baseado em Regressão Logística.

Krejzl and Steinberger (2016) utilizam o corpus anotado para Hillary Clinton para treinar um classificador *cross-target* baseado em Entropia Máxima. Os autores argumentam que o corpus rotulado para Hillary Clinton pode ser um bom corpus de treino para o alvo Trump por ambos serem figuras públicas do meio político. Dessa forma, usam o corpus para construir um dicionário de posicionamento, constituído de hashtags de posicionamento. Essas são extraídas do corpus rotulado com base na frequência de cada hashtag em determinado lado do posicionamento neste corpus. A abordagem também faz uso de atributos baseados em bag-of-words, léxicos de sentimento e deste dicionário de posicionamento.

Todos esses trabalhos que utilizaram abordagens *cross-domain/cross-target*, só foram viabilizados pela disponibilidade de um corpus anotado para treinar um classificador, e não obtiveram resultados significativos na sub tarefa “B” da competição do Semeval 2016.

Trabalhos que fizeram uso de expressões-semente (DIAS; BECKER, 2016b; WEI et al., 2016; EBRAHIMI; DOU; LOWD, 2016a; EBRAHIMI; DOU; LOWD, 2016b; AUGENSTEIN et al., 2016) obtiveram os melhores resultados dentre os trabalhos semi-supervisionados, mas necessitam de conhecimento do domínio ou inspeção do corpus a ser classificado para anotação das expressões semente.

Em Dias and Becker (2016b) foi apresentada uma versão preliminar deste trabalho. O processo desenvolvido é semi-supervisionado e utiliza um conjunto de regras para rotular automaticamente os tweets e construir um corpus para classificação complementar daqueles não cobertos pelas regras. Esse conjunto de regras é baseado em palavras chave semente e alvos de opinião relacionados ao alvo principal do problema de posicionamento, além da polaridade detectada para cada tweet. Esta abordagem recai na necessidade de inspeção manual de uma amostra para identificação de palavras-semente e alvos de opinião relacionados, levando os autores a avançar na pesquisa em direção à abordagem não supervisionada proposta nesta dissertação. Este trabalho preliminar é detalhado no Apêndice A.

Wei et al. (2016), vencedores da sub tarefa 6-B do Semeval 2016, definem manualmente um conjunto de expressões regulares para rotular automaticamente tweets que compõem o corpus de domínio do alvo Trump com base na presença dessas expressões. Assim formam um corpus para treinar um classificador baseado em uma rede convolucional, utilizando como entrada adicional para a rede um *embedding* pré-treinado com base em notícias do Google News. A definição das expressões semente provavelmente foi realizada através de inspeção do corpus, dada a especificidade das expressões usadas, o que indica um esforço adicional para a constru-

ção do sistema.

Ebrahimi, Dou and Lowd (2016a) rotulam automaticamente o corpus de domínio de Trump com base na presença de expressões e hashtags semente, também provavelmente obtidas através de inspeção manual do corpus dada a especificidade dessas. Para aumentar o corpus, também coletam tweets de seguidores dos perfis autores dos tweets do corpus do alvo Trump. Propagam rótulos entre perfis que retuitaram os tweets coletados, e entre tweets retuitados pelos mesmos perfis com o uso de uma rede aleatória de Markov. Por fim, utilizam um algoritmos de classificação SVM.

Ebrahimi, Dou and Lowd (2016b) desenvolveram um modelo de classificação que chamam de Generativo Sentimento-alvo-posicionamento. Ele realiza a classificação de tweets rotulados a partir da detecção de presença de expressões e hashtags semente. Essas expressões foram identificadas manualmente e anotadas para posicionamento, provavelmente resultado de inspeção do corpus de domínio. Para determinar os atributos utilizados no classificador, utilizam bag-of-words, uma API de detecção de sentimento (VADER<sup>4</sup>) e léxico de sentimento. Assim como os trabalhos anteriormente citados, estes autores provavelmente recorreram à inspeção do corpus para a formação do conjunto de expressões semente, dada a especificidade das expressões usadas.

Finalmente, Augenstein et al. (2016) extraem as palavras que compõem o texto de cada alvo da sub tarefa 6-A do Semeval para coletar tweets adicionais para treinar e construir um *embedding*. Também extraem expressões semente do corpus de domínio de Trump através de inspeção manual para rotular este corpus de domínio de acordo com o lado do posicionamento identificado pelos anotadores para cada expressão. Após, utilizam o corpus rotulado e o *embedding* treinado como entradas para o desenvolvimento de um classificador baseado em uma rede neural LSTM (*Long short-term memory*). Esse trabalho foi desenvolvido após o Semeval e apresenta o melhor resultado semi-supervisionado para o alvo Trump (58,03). Por essa razão, é utilizado como *baseline* estado-da-arte neste trabalho.

### 3.7 Considerações Finais

Trabalhos que se utilizam de abordagens supervisionadas costumam obter bons resultados tanto na tarefa de detecção de polaridade, quanto na tarefa de detecção de posicionamento. Porém incorrem no alto custo, esforço e inacurácia dos processos de anotação manual de corpora. Desta forma, a tarefa de detecção de posicionamento em tweets utilizando abor-

---

<sup>4</sup><https://github.com/cjhutto/vaderSentiment>

dagens não/semi-supervisionadas surgem como uma alternativa promissora para a solução do problema.

Os trabalhos existentes na área abordam o problema de detecção de posicionamento em tweets de forma semi-supervisionada e necessitam de uma entrada para sua execução, recaindo em 2 estratégias básicas:

- o uso de um corpus rotulado para a construção de um classificador *cross-domain* ou *cross-target* para detecção de posicionamento;
- a definição manual de palavras, expressões ou hashtags semente, utilizadas para rotular automaticamente tweets e construir um classificador.

O primeiro grupo tem como problema o custo e o esforço em se construir um corpus rotulado aplicável com bons resultados a outro alvo para o qual se deseja detectar o posicionamento. O segundo recai no esforço manual em identificar as expressões semente, o que requer grande conhecimento do domínio ou, ainda, a inspeção do corpus para o qual se deseja detectar o posicionamento. Além disso, a solução é extremamente sensível à escolha destas expressões.

O trabalho apresentado nesta dissertação inova em relação às abordagens apresentadas por desenvolver um processo de detecção de posicionamento não supervisionado baseado somente no conteúdo textual de tweets, e que até o momento, não encontramos na literatura. O processo proposto não requer entrada manual de expressões semente ou de corpus previamente rotulado, além de apresentar resultados superiores para a métrica do Semeval 2016 para o alvo Trump (59,71) quando comparado ao *baseline* estado-da-arte (AUGENSTEIN et al., 2016) (58,03).

## 4 PROCESSO DE DETECÇÃO NÃO SUPERVISIONADA DE POSICIONAMENTO EM TEXTOS DE TWEETS

Neste capítulo, será apresentado o processo proposto para detecção de posicionamento em tweets baseado somente em seu conteúdo textual. Inicialmente, é dada uma visão geral do processo e, a seguir, são detalhadas cada etapa e tarefa do processo.

### 4.1 Visão Geral

O processo proposto neste trabalho realiza a detecção de posicionamento baseada exclusivamente no conteúdo dos tweets através de uma abordagem *não supervisionada*. Este processo é não supervisionado por não requerer nenhuma entrada adicional, além do próprio alvo para o qual se pretende detectar posicionamento (e.g. “Donald Trump”) e dos tweets que devem ser classificados. Essa é a característica mais importante do trabalho, já que o diferencia dos demais trabalhos estudados e apresentados no Capítulo 3.

Através da abordagem proposta, são eliminados problemas significativos não solucionados pelos trabalhos de abordagem supervisionada e semi-supervisionada, principalmente no que diz respeito às entradas necessárias. Como visto, os sistemas existentes recaem em dois tipos de entrada que oneram a construção e a execução dos sistemas de detecção de posicionamento, como discutido na Seção 3.6.

Outra característica importante é o fato de o processo proposto basear-se somente no conteúdo textual dos tweets, desconsiderando as características estruturadas da plataforma, como *retweets* ou conexões entre perfis, entre outras informações disponíveis. Isso torna o processo mais simples e menos oneroso quanto a requisitos de processamento.

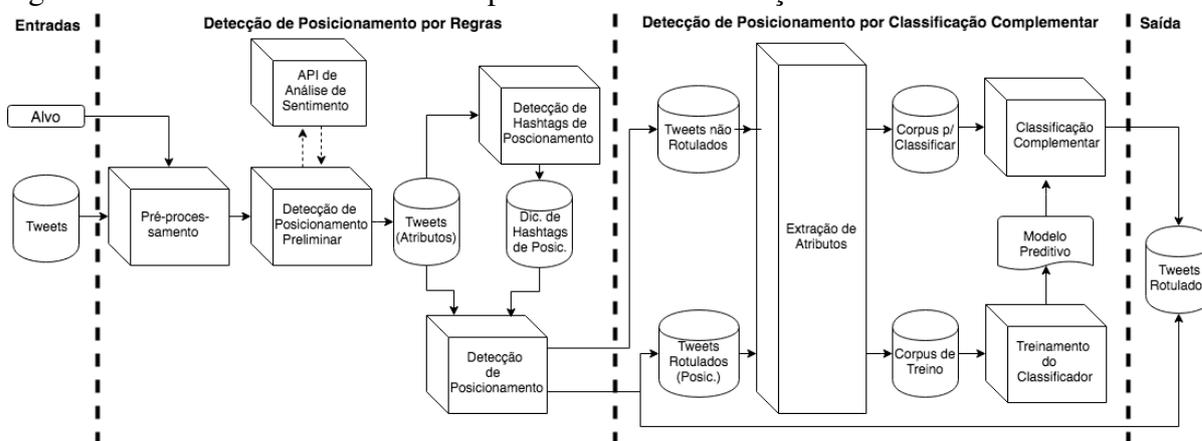
A Figura 4.1 apresenta o fluxo completo do processo, que implementa uma abordagem híbrida composta de duas etapas resumidas a seguir:

- **Detecção de posicionamento por regras:** o objetivo desta etapa é classificar tweets usando heurísticas de posicionamento, representadas por regras baseadas no sentimento em relação ao alvo principal e no uso de hashtags de posicionamento. Essas hashtags são identificadas automaticamente a partir de um subconjunto dos tweets que expressam opiniões sobre o alvo principal, compondo um dicionário de hashtags. Uma vez identificadas as hashtags, as regras são aplicadas em determinada ordem em todo conjunto de tweets. Os tweets que atenderem a alguma das heurísticas são rotulados para posicionamento, e

os demais, classificados na próxima etapa;

- **Deteção de posicionamento por classificação complementar:** atua de forma complementar à etapa anterior, classificando os tweets que não foram cobertos pelas regras de deteção de posicionamento. Esta etapa recebe como entrada os tweets rotulados na etapa anterior para treinar um classificador e aplicá-lo aos tweets que não foram rotulados, resultando na classificação da totalidade dos tweets de entrada, seja na primeira etapa ou seja na segunda.

Figura 4.1: Fluxo do Processo Não Supervisionado de Deteção de Posicionamento em Tweets.



Fonte: o autor.

As regras definidas para este trabalho estão listadas na Tabela 4.1 e se dividem em dois tipos: a) regras baseadas em hashtags de posicionamento e b) regras baseadas em sentimentos. Tweets contendo hashtags podem adicionalmente expressar opinião ou não e, conseqüentemente, as regras não são mutuamente exclusivas. Neste trabalho, as regras baseadas em hashtags têm precedência sobre as regras baseadas em opinião, pois se supõe que a deteção pela presença de hashtags de posicionamento tem melhor desempenho. Por um lado, hashtags são uma forma própria dessa mídia para expressar ideias, sentimentos e atitudes, e que dão contexto ao curto conteúdo textual dos tweets. Por outro, é mais simples e preciso detectar a presença de uma hashtag do que a polaridade e o alvo de uma opinião expressa em um tweet.

A proposta essencial deste trabalho é não depender de entradas adicionais. Com esse objetivo, propõe-se a extração automática de hashtags de posicionamento a partir de tweets onde o posicionamento é expresso na forma de opinião em relação ao alvo principal. Como foi analisado no Capítulo 1, a opinião em relação ao alvo principal é a forma mais básica de expressão de posicionamento. Essa característica é explorada por parte das regras que foram definidas para rotulação automática.

Tabela 4.1: Lista de regras por ordem de aplicação

Tipo	#	Regra	Descrição	Posicionamento
Hashtags	1	HASHTAG-FAVOR	Quantidade maior de hashtags de posicionamento favorável que contrário	Favorável
	2	HASHTAG-CONTRA	Quantidade maior de hashtags de posicionamento contrário que favorável	Contrário
Sentimento	3	ALVO-POSITIVO	Menção ao alvo e polaridade positiva	Favorável
	4	ALVO-NEGATIVO	Menção ao alvo e polaridade negativa	Contrário
	5	NEUTRO	Ausência de menção ao alvo e de hashtags, polaridade neutra	Neutro

A identificação das hashtags parte de um subconjunto de tweets previamente rotulados para posicionamento. Assim, essa rotulação preliminar é baseada apenas na polaridade e na menção ao alvo. Com base em critérios de frequência mínima, as hashtags de posicionamento favorável/contrário são extraídas desses tweets com polaridade positiva/negativa, respectivamente, assumindo que as mesmas são expressões de claro posicionamento.

Uma vez construído esse dicionário, a detecção de posicionamento por regras é realizada de forma definitiva sobre o conjunto total de tweets, considerando a ordem definida na Tabela 4.1. Devido à precedência das regras, os rótulos preliminares podem ser substituídos por novos à medida que os tweets forem filtrados por alguma das regras.

A etapa de classificação complementar tem como objetivo classificar tweets para os quais o posicionamento não se dá através de opinião direcionada ao alvo, nem pela presença de hashtags de posicionamento detectadas na etapa anterior. Nesse caso, assume-se que o posicionamento é caracterizado pela presença de termos que podem ser encontrados no tweets filtrados pelas regras e que constituem um corpus de treino para um classificador de posicionamento. Este classificador tem como atributos os termos utilizados nos tweets através de uma abordagem de n-gramas.

Note-se que com esta classificação complementar outras formas de expressão de posicionamento não cobertas pelas regras podem ser tratadas pelo processo de forma indireta, desde que representativas no corpus de treino. Tweets filtrados pelas regras podem conter termos frequentes e/ou discriminativos que denotem um posicionamento, incluindo alvos secundários ou expressões de posicionamento outras que hashtags. Exemplos de termos discriminativos poderiam ser o tópico “Macy’s” ou o alvo secundário “mexicanos”, mencionados nos exemplos da Tabela 1.1.

Experimentos demonstraram que a etapa de detecção de posicionamento por regras tem melhor desempenho que a classificação complementar (DIAS; BECKER, 2016c; DIAS; BECKER, 2016a; DIAS; BECKER, 2016b). Porém, a segunda etapa é fundamental, já que a pri-

meira não cobre a totalidade dos tweets, e portanto trata de outras formas em que o posicionamento é expresso.

Ao final do processo, o resultado será a totalidade dos tweets rotulados para posicionamento em relação ao alvo principal do problema.

Resumidamente, o processo proposto neste trabalho implementa uma abordagem composta pelas seguintes características chave:

- **Não supervisionado:** as únicas entradas necessárias são o alvo principal e o corpus a ser rotulado;
- **Abordagem Híbrida:** abordagem que combina heurísticas de posicionamento para rotulação automática e aprendizado supervisionado para classificação complementar;
- **Heurísticas:** consiste de regras baseadas em hashtags e em regras de sentimento para rotular os tweets de forma automática, bem como para gerar automaticamente um corpus de treino;
- **Dicionário de Hashtags de Posicionamento:** as hashtags de posicionamento são identificadas a partir da análise automática dos tweets com base em um posicionamento preliminar e frequência mínima;
- **Classificação Complementar:** etapa complementar responsável por rotular tweets em que o posicionamento não se dá por opinião direcionada ao alvo principal, nem por hashtags de posicionamento.

Em relação a versões prévias deste trabalho (DIAS; BECKER, 2016c; DIAS; BECKER, 2016a; DIAS; BECKER, 2016b) apresentadas no Apêndice A, as principais diferenças são:

- as versões preliminares deste trabalho eram semi-supervisionadas, exigindo a entrada manual de expressões semente (palavras ou hashtags) de posicionamento, e de alvos relacionados ao principal. Esses foram substituídos pela identificação automática de hashtags de posicionamento, que aumentou a precisão obtida pelas regras de sentimento (3 a 5 na Tabela 4.1);
- as regras baseadas na presença de expressões de posicionamento foram substituídas por regras baseadas na frequência de hashtags de posicionamento (regras 1 e 2), também mais precisas, e a classificação complementar passou a cobrir os posicionamentos envolvendo alvos secundários e expressões de posicionamento;

- foram eliminadas regras que assumiam sentimentos em relação a alvos contrários (e.g. sentimento positivo em relação a Hillary significaria posicionamento contrário em relação ao alvo principal Trump), pois essas regras foram efetivas somente em domínios específicos (e.g. Ateísmo) e, através de sua eliminação, o processo tornou-se não supervisionado.

No restante deste capítulo, cada etapa e respectivas tarefas propostas serão descritas com mais detalhes.

## 4.2 Detecção de Posicionamento por Regras

Como já mencionado, as entradas desta etapa são o alvo principal do problema de posicionamento (e.g. “Donald Trump”) e um conjunto de tweets para identificação de posicionamento. O objetivo final desta etapa é rotular o corpus de entrada utilizando regras que envolvem hashtags de posicionamento e opiniões direcionadas ao alvo de posicionamento.

A etapa é composta por uma série de tarefas: a) Pré-processamento, b) Detecção de Posicionamento Preliminar, c) Detecção de Hashtags de Posicionamento e d) Detecção de Posicionamento. Cada tweet do corpus é representado por uma tupla, e à medida que as tarefas são executadas, novos atributos são incluídos.

O resultado desta etapa são dois (2) conjuntos de tweets, que serão utilizados como entrada para a etapa de detecção de posicionamento por classificação complementar:

- Um conjunto de tweets não rotulados pelas regras, cujo rótulo deve ser determinado de forma complementar pela segunda etapa;
- Um conjunto de tweets rotulados pelas regras, que compõem o resultado final do processo e, adicionalmente, são utilizados como entrada (i.e. corpus de treino) para a próxima etapa.

No restante desta seção, cada tarefa é descrita em termos de seu objetivo, entrada, processamento e saída.

### 4.2.1 Pré-processamento

O objetivo desta tarefa é preparar o corpus para as demais tarefas do fluxo, através da limpeza do conteúdo textual dos tweets e da geração de atributos básicos.

**Entradas:**

1. String representando Alvo
2. Corpus: <Tweet>

**Processamento:**

- Limpeza dos textos de cada tweet
  1. substituição dos caracteres alfabéticos maiúsculos pelos respectivos caracteres minúsculos. Isto evita que haja redundância de termos ou atributos gerados para a tarefa de classificação. Adicionalmente, é garantido que palavras com mesmos termos sejam classificadas de acordo pelos algoritmos de classificação;
  2. substituição de menções a perfis do Twitter (e.g. @fulanodetal) pelo *token* “a\_mention”, com exceção daquelas que contêm *tokens* extraídos da *string* que representa o alvo principal do problema (e.g. @realDonaldTrump). O objetivo deste passo é manter as referências ao alvo principal do problema de posicionamento, mas ignorar referências a perfis pessoais dos autores de tweets e daqueles para os quais eles emitem seus comentários. Como consequência, ocorre uma redução do número de atributos irrelevantes para treinar o classificador na segunda etapa do processo;
  3. exclusão de aspas, vírgulas e ponto-e-vírgulas, pois são irrelevantes para o posicionamento.
  4. exclusão da hashtag “#semst”, pois ela consta em todos os tweets dos corpora disponibilizados pelos organizadores do Semeval 2016.
- Atributos básicos são adicionados às tuplas que representam o corpus:
  1. **menção\_ao\_alvo:** atributo binário indicando a presença ou ausência de *tokens* que compõem o string que identifica o alvo no conteúdo textual de cada tweet (e.g., “donald” e “trump”). Este atributo será combinado com a polaridade em tarefas posteriores para detectar o posicionamento;
  2. **tem\_hashtag:** atributo binário indicando a presença ou a ausência de hashtags. Este atributo é utilizado pela regra 5 (i.e., tweet de posicionamento neutro), para a qual uma das condições é a ausência de hashtags no corpo do tweet.

**Saídas:**

1. Alvo
2. Corpus: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags>

#### 4.2.2 Detecção de Posicionamento Preliminar

O objetivo desta tarefa é adicionar uma informação de posicionamento preliminar para cada tweet do corpus que for filtrado pelas regras de sentimento (somente as regras 3 e 4 da Tabela 4.1). Além disso, um atributo indicando o rótulo de polaridade é adicionado ao corpus.

##### Entradas:

1. Corpus: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags>

##### Processamento:

Para cada tweet:

1. o texto do atributo *Tweet\_limpo* é polarizado. Embora qualquer técnica de detecção de polaridade possa ser utilizada, neste trabalho foi utilizada a funcionalidade de detecção de sentimento em textos da IBM Alchemy, utilizando sua API online<sup>1</sup>. O texto original de cada tweet é submetido à API. Esta API retorna um escore de polaridade contendo um valor entre -1 e 1 para cada tweet analisado. O escore é analisado resultado em um rótulo negativo para um escore menor ou igual a -0,2, positivo para escore maior ou igual a 0,2 e neutro para escores entre -0,2 e 0,2. Este atributo polaridade com o respectivo valor é adicionado à respectiva tupla representando o tweet;
2. para todos os tweets que têm valor “1” no atributo *menção\_ao\_alvo*, são aplicadas as regras 3 e 4 para detectar o posicionamento preliminar do tweet, a fim de gerar rótulo de posicionamento favorável ou contrário. Se o rótulo de polaridade for positivo, será atribuído rótulo de posicionamento favorável; se a polaridade for negativa, o rótulo de posicionamento será contrário; nos demais casos o rótulo tem valor indefinido. Este atributo posicionamento com o respectivo rótulo de posicionamento preliminar é adicionado à respectiva tupla representando o tweet.

##### Saídas:

1. Corpus: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>

---

<sup>1</sup><https://www.ibm.com/watson/alchemy-api.html>

### 4.2.3 Detecção de Hashtags de Posicionamento

O objetivo desta tarefa é construir um dicionário de hashtags de posicionamento extraídas do corpus com base na frequência em que ocorrem em tweets com posicionamento preliminar favorável ou contrário.

#### Entradas:

1. Corpus<Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>

#### Processamento:

1. os tweets cujo atributo *posicionamento* for favorável ou contrário são percorridos realizando a contagem de ocorrências de cada uma de suas hashtags em tweets favoráveis e contrários, resultando em uma tabela contendo três informações para cada item: a) hashtag, b) quantidade de tweets favoráveis e c) quantidade de tweets contrários;
2. são descartadas as hashtags que: a) forem encontradas em apenas 1 tweet, ou b) as quantidades de ocorrências em tweets favoráveis e contrários forem iguais;
3. é criado um dicionário com uma entrada para cada hashtag restante. O respectivo posicionamento é *favorável* se a quantidade de ocorrências em tweets favoráveis superior à quantidade em tweets contrários, ou *contrário* se quantidade de ocorrências em tweets contrários superior à quantidade em tweets favoráveis.

#### Saídas:

1. Dicionário de hashtags de posicionamento: <hashtag, posicionamento>.

### 4.2.4 Detecção de Posicionamento

Esta tarefa conclui esta etapa e tem como objetivo rotular o maior número possível de tweets do corpus com base nas heurísticas propostas.

#### Entradas:

1. Corpus: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>
2. Dicionário de hashtags de posicionamento: <hashtag, posicionamento>

**Processamento:**

Para cada tweet do Corpus:

1. são contadas as hashtags de posicionamento favorável e de posicionamento contrário, gerando 2 informações contendo estas quantidades;
2. são aplicadas as regras de acordo com a ordem estabelecida na Tabela 4.1
3. caso o tweet não tenha sido filtrado pelas regras, irá compor o conjunto de tweets *não rotulados*;
4. caso o tweet tenha sido filtrado pelas regras, irá compor o conjunto de tweets *rotulados*.

**Saídas:**

1. Tweets não rotulados: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>
2. Tweets rotulados: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>

**4.3 Detecção de Posicionamento por Classificação Complementar**

O objetivo desta etapa é rotular de forma complementar o corpus submetido ao processo, detectando posicionamentos expressos que utilizam meios alternativos de expressão de posicionamento àqueles cobertos pelas heurísticas. Ao final do processamento, adiciona os tweets classificados ao corpus que compõe a saída do processo resultando na totalidade dos tweets rotulados para posicionamento. Como será visto com mais detalhes no Capítulo 5, esta etapa apresenta desempenho inferior quando tem os resultados comparados aos da etapa inicial, porém, é fundamental pelo fato de as regras não filtrarem 100% dos tweets.

**4.3.1 Extração de atributos**

Tarefa inicial da etapa, tem como objetivo extrair as *features* dos dois datasets de tweets que recebe como entrada, visando treinamento de um classificador (corpus rotulado) e previsão de posicionamento com base no modelo construído (corpus não rotulado). Como *features* escolheu-se n-gramas de tamanho 1 a 4. Foram realizados experimentos incorporando outros

atributos (e.g., *polaridade*, *menção\_ao\_alvo*,...), porém esses apresentaram resultados ligeiramente inferiores aos obtidos pela abordagem proposta. Detalhes sobre esses experimentos são discutidos no Capítulo 5.

**Entradas:**

1. Tweets não rotulados: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>
2. Tweets rotulados: <Tweet, Tweet\_limpo, menção\_ao\_alvo, tem\_hashtags, polaridade, posicionamento>

**Processamento:**

1. o texto limpo (*Tweet\_limpo*) de cada tweet dos datasets de entrada (rotulados e não rotulados) é processado a fim de extrair n-gramas (n=1 a 4), usando um *tokenizador*. Para cada atributo, são geradas duas representações: binária e TF-IDF, já que os experimentos revelaram melhores resultados utilizando os 2 modos de representação dos n-gramas combinados. Nos experimentos realizados, foi utilizada para esta tarefa a função *string-towordvector* da API Java da ferramenta Weka na versão 3.9 (HALL et al., 2009). Os parâmetros utilizados foram os padrões da ferramenta, exceto pelo limite de *tokens* (i.e., sem limite).

**Saídas:**

1. Corpus para classificar: <Tweet, n-gramas\_binários, n-gramas\_tf-idf>
2. Corpus de treino: <Tweet, n-gramas\_binários, n-gramas\_tf-idf, posicionamento>

### 4.3.2 Treinamento do Classificador

O objetivo desta tarefa é gerar um modelo preditivo, utilizando um algoritmo de classificação sobre o corpus de treino. Embora a abordagem seja agnóstica quanto ao algoritmo de classificação, os melhores resultados foram obtidos com um classificador SVM. Experimentos apresentados no Capítulo 5 descrevem os resultados de vários outros algoritmos e estratégias testadas, incluindo *Simple Logistic*, *Random Forest* e comitê de classificadores.

**Entradas:**

1. Corpus de treino: <Tweet, n-gramas\_binários, n-gramas\_tf-idf, posicionamento>

**Processamento:**

1. o corpus de treino gerado na tarefa anterior é utilizado para treinar um classificador de três classes: favorável, contrário e neutro. Neste trabalho, foi utilizado o algoritmo SMO<sup>2</sup> em sua configuração padrão, disponível na API Java da ferramenta Weka na versão 3.9 (HALL et al., 2009).

**Saídas:**

1. Modelo Preditivo

**4.3.3 Classificação Complementar**

A tarefa tem como objetivo rotular os tweets que restaram da etapa anterior sem rótulo de posicionamento com base no modelo preditivo gerado.

**Entradas:**

1. Modelo Preditivo;
2. Corpus para classificar: <Tweet, n-gramas\_binários, n-gramas\_tf-idf>

**Processamento:**

Para cada tweet não rotulado:

1. o tweet é submetido ao classificador para se obter um posicionamento de acordo com as classes que foram utilizadas para a construção deste modelo preditivo: favorável, contrário ou neutro. Foi utilizada a API Java da ferramenta Weka na versão 3.9 (HALL et al., 2009).

**Saídas:**

1. Tweets rotulados: <Tweet, posicionamento>

**4.3.4 Considerações Finais**

Neste capítulo, foi apresentada uma abordagem não supervisionada para detecção de posicionamento em textos de tweets. O processo utiliza uma abordagem híbrida composta por

---

<sup>2</sup>Implementação do algoritmo SVM

um conjunto de heurísticas e uma etapa de classificação complementar. O grande diferencial entre esta e as abordagens apresentadas no Capítulo 3 é o fato de o processo desenvolvido não requerer entrada de corpus anotado para posicionamento, nem de expressões semente, tornando este processo não supervisionado.

São utilizadas duas premissas na etapa de detecção por regras: a) opiniões direcionadas ao alvo principal do problema permitem inferir o posicionamento do autor, e b) hashtags podem denotar claro posicionamento e podem ser extraídas dos tweets rotulados a partir da premissa “a”.

O uso de uma etapa complementar de classificação permite rotular a totalidade dos tweets submetidos ao processo e cobrir formas de expressão de posicionamento não cobertas pelas regras, como tweets sem opinião ou com opinião direcionada a alvos relacionados ao alvo principal do problema.

## 5 EXPERIMENTOS E ANÁLISE DE RESULTADOS

Este capítulo apresenta a descrição e a análise dos experimentos e resultados obtidos.

### 5.1 Descrição dos Experimentos

A fim de responder as questões de pesquisa apresentadas no Capítulo 1, foram realizados 3 experimentos, resumidos abaixo:

1. **Experimento #1 - *Baselines* Semi-supervisionados:** o objetivo é comparar o desempenho do processo proposto com os resultados obtidos por sistemas semi-supervisionados discutidos na Seção 3.6. Para a comparação será utilizado o conjunto de dados disponibilizado para a subtarefa “B” do Semeval 2016, cujo alvo é Donald Trump;
2. **Experimento #2 - *Baselines* Supervisionados:** o objetivo é verificar a generalidade do processo proposto. Como não existem *baselines* semi-supervisionados para outros alvos, exceto versões preliminares de nosso próprio trabalho, o processo é comparado a sistemas supervisionados utilizando os alvos disponibilizados para a subtarefa “A” do Semeval 2016. Esses envolvem um alvo no mesmo domínio do Experimento #1 (i.e., Hillary Clinton), além de alvos em outros domínios (i.e. Ateísmo, Clima, Feminismo e Aborto);
3. **Experimento #3 - *Variações do processo*:** o objetivo é comparar o processo proposto com variações experimentadas para as etapas de Detecção de Posicionamento por Regras e Classificação Complementar. Mais especificamente, são explorados o uso de palavras semente, uso de perfis semente do Twitter, variações nas regras e na construção do modelo preditivo. Esses experimentos envolvem os 6 alvos propostos para o Semeval 2016.

O restante deste capítulo está organizado como segue. A Seção 5.2 descreve o uso dos corpora e as métricas para avaliação dos experimentos. A Seção 5.3 apresenta e discute os resultados dos experimentos #1 e #2. As estratégias utilizadas pelo processo proposto são avaliadas na Seção 5.4. O experimento #3 é discutido na Seção 5.5 através do estudo de variações do processo proposto. Com base nestas avaliações, respondemos na Seção 5.6 às questões de pesquisa elencadas neste trabalho.

## 5.2 Dados e Métricas

Os dados utilizados para a execução dos experimentos foram aqueles disponibilizados pelos organizadores do Semeval 2016 para execução da tarefa 6, já detalhados na Seção 3.4.1. Cada execução do processo, teve como entrada uma string caracterizando o alvo, tal como definido na tarefa do Semeval 2016, e os respectivos corpora fornecidos. Contudo, foram utilizados nos experimentos somente os dados *anotados*, como segue:

- os corpora de treino e teste *anotados* do Semeval foram submetidos à etapa de Detecção de Posicionamento por Regras, visando compor automaticamente um corpus de treino para a etapa de Classificação Complementar. No caso específico do alvo Trump, isso quer dizer que foi utilizado somente o corpus de teste fornecido para a tarefa “B”, enquanto para os demais alvos foram utilizados indistintamente os corpora de treino e teste fornecidos para a subtarefa “A”;
- exceto se explicitamente indicado, todas as métricas foram calculadas somente utilizando os corpora de *teste*, visando permitir comparação com os baselines publicados.

Para medir o desempenho do sistema proposto e compará-lo com os resultados dos trabalhos utilizados como *baselines*, foi utilizada a métrica do Semeval ( $F1_{Média}$ ), que faz a média do desempenho obtido para o posicionamento favorável e para o contrário, como já apresentado na Seção 3.4.2. Essa métrica foi calculada individualmente para cada corpus de teste submetido ao processo. Os organizadores do Semeval 2016 consideraram todos os corpora de teste (i.e. Ateísmo, Clima, Feminismo, Hillary e Aborto) como um único corpus para avaliação dos competidores da subtarefa “A”, calculando para este corpus consolidado a mesma métrica  $F1_{Média}$ . Para poder avaliar melhor as características da solução proposta em relação as especificidades de cada corpus, nossa avaliação utiliza a  $F1_{Média}$  sobre o conjunto de todos os 5 corpora, bem como através da avaliação desta métrica para cada um dos 5 corpora de forma individual.

Já para medir a eficiência das diferentes tarefas do processo, tais como a detecção de posicionamento por regras e a detecção de polaridade, foram utilizadas as métricas de precisão, revocação e medida-F, amplamente utilizadas para análise de desempenho de sistemas envolvendo classificação e detalhadas na Seção 2.4.

Tabela 5.1: Experimento comparado aos baselines para a subtarefa B

Trabalho	F1 <sub>Média</sub>	Estratégia	Referência
1-RH-RS-SVM	<b>59,71</b>	não supervisionada	Este Trabalho
Bicond	58,03	expressões semente	(AUGENSTEIN et al., 2016)
pkudblab	56,28	expressões semente	(WEI et al., 2016)
WIC	50,58	expressões semente	(DIAS; BECKER, 2016b)
UWB	42,02	cross-target	(KREJZL; STEINBERGER, 2016)

### 5.3 Avaliação de Desempenho

Nesta seção, são apresentados os experimentos #1 e #2, que comparam os resultados obtidos através do processo proposto nesta dissertação (Capítulo 4) com os respectivos *baselines*. A execução do processo proposto será identificada por “1-RH-RS-SVM”, onde “RH” indica o uso de regras baseadas em hashtags, “RS” indica o uso de regras baseadas em sentimento e “SVM” indica o uso de um classificador SVM na etapa de classificação complementar. Inicialmente, são apresentados os resultados obtidos e comparados aos *baselines* supervisionados e semi-supervisionados e, em seguida, a aplicabilidade da proposta é avaliada de acordo com os resultados obtidos para os corpora.

#### 5.3.1 Experimento #1 - Baselines Semi-supervisionados

O experimento #1 consiste na execução do processo para o alvo Trump e na comparação do resultado global obtido de acordo com a métrica utilizada pelo Semeval 2016 (F1<sub>Média</sub>) com os *baselines*. As entradas do processo para este experimento são o corpus disponibilizado para o alvo Trump e próprio alvo (i.e. “donald trump”). A Tabela 5.1 apresenta o resultado comparado a quatro trabalhos semi-supervisionados apresentados no Capítulo 3.

O *baseline Bicond* é considerado o estado-da-arte, por ser o melhor resultado conhecido para o alvo Trump. O mesmo foi publicado após a competição do Semeval 2016 superando o vencedor da subtarefa “B” na competição, a saber *pkudblab*. O trabalho *UWB* também está entre os competidores do Semeval e foi o que obteve o resultado mais significativo (42,02) dentre os de abordagem *cross-domain/target*. O trabalho *WIC* consiste em uma versão preliminar do processo proposto, publicada após o Semeval 2016. A mesma é semi-supervisionada, baseada em palavras chave e alvos semente e um conjunto distinto de heurísticas.

Este trabalho obteve um resultado superior em quase 2 pontos percentuais ao estado-da-arte (58,03). É importante considerar que este trabalho é não supervisionado, o que torna esse resultado ainda mais relevante. Em relação ao nosso trabalho prévio, o presente trabalho avança

Tabela 5.2: Experimento comparado aos Baselines para os corpora da sub tarefa A

TRABALHO	HILLARY	CLIMA	FEMINISMO	ABORTO	ATEÍSMO	TAREFA A	REFERÊNCIA
1-RH-RS-SVM	<b>63,59</b>	47,44	43,78	38,34	0,00	59,01	Este Trabalho
Disc-STS	74,48	41,18	57,90	67,94	65,52	71,03	(EBRAHIMI; DOU; LOWD, 2016a)
MITRE	57,67	41,63	62,09	57,28	61,47	67,82	(ZARRELLA; MARSH, 2016)
WIC	62,87	53,67	47,69	48,29	54,23	60,17	(DIAS; BECKER, 2016b)

consistentemente nos resultados para o alvo Trump com um incremento de nove pontos, além de avançar na independência de inspeção de corpus.

### 5.3.2 Experimento #2 - Baselines Supervisionados

O experimento #2 consiste na execução do processo proposto no Capítulo 4 para os alvos que foram utilizados para avaliação dos sistemas supervisionados no Semeval 2016. O processo foi executado individualmente para cada alvo, tendo como entrada todos os respectivos tweets fornecidos (i.e. treino e teste) e o string do alvo em análise (i.e. “atheism”, “climate change is a real concern”, “feminist movement”, “hillary clinton” e “legalization of abortion”). Como já destacado, o cálculo da métrica de avaliação do Semeval 2016 foi realizado utilizando somente o corpus de teste, permitindo a comparação com os *baselines* adotados.

A Tabela 5.2 apresenta os resultados para cada alvo e para o corpus consolidado considerando todos os corpora de teste. O *baseline Disc-STS* foi publicado após o Semeval apresentando o melhor resultado conhecido para a sub tarefa “A”. *MITRE* foi o vencedor do Semeval 2016 nesta tarefa. O *baseline WIC*, como já discutido na seção anterior, é uma versão preliminar semi-supervisionada do processo proposto nesta dissertação.

O resultado obtido para o alvo *Hillary* (63,59) é superior ao vencedor supervisionado do Semeval (*MITRE*), e comparável ao trabalho preliminar semi-supervisionado (*WIC*), mas fica aquém do estado da arte (*Disc-STS*).

Para o alvo *Clima*, obteve-se um resultado (47,44) superior aos *baselines* supervisionados (*Disc-STS* e *MITRE*), embora um pouco inferior ao *baseline* semi-supervisionado (*WIC*). Esse desempenho ocorre em função das características deste corpus: o menor número de tweets no corpus de teste (169), maior número de instâncias para a classe favorável (72,78% ou 123 tweets), e conjunto de dados desbalanceado, já que as instâncias da classe contrária representam apenas 6,51% do total (11 tweets).

A  $F1_{Média}$  para os alvos *Feminismo* (43,78) e *Aborto* (38,34) é inaceitável e fica abaixo de todos *baselines*. Uma análise desses resultados revela que o problema está na representatividade do alvo principal fornecido como entrada, i.e. “Feminist Movement” e “Legalization

of Abortion”. Na Seção 5.5.2, será demonstrado que resultados substancialmente melhores são obtidos quando do uso adicional de expressões ou perfis semente.

Quanto ao alvo *Ateísmo*, o problema também está relacionado ao alvo, mas de uma maneira um pouco diferente. Neste corpus, a grande maioria das opiniões são dirigidas a alvos de tema religioso como “jesus” e “god”, como formas de expressar posicionamento em oposição ao Ateísmo. Nesse caso, o processo também se revelou inadequado, e inferior às variações do processo proposto que envolvem expressões ou perfis semente (Seções 5.5.2 e 5.5.3).

Em resumo, os resultados obtidos para o experimento #2 relevam um bom desempenho do processo não supervisionado somente para alvos onde grande parte do posicionamento se dá através da expressão de opinião direcionada ao alvo principal do problema de posicionamento (Hillary e Trump). Já para os demais, as opiniões se dão através de alvos secundários relacionados (Aborto, Feminismo), pelo uso de hashtags (Clima), ou alvos por oposição (Ateísmo), os quais não são facilmente detectáveis. Esse e outros elementos que procuram explicar os resultados obtidos serão explorados ao longo da Seção 5.4.

### 5.3.3 Avaliação de Resultados dos Experimentos #1 e #2

A Tabela 5.3 apresenta os resultados obtidos segundo três métricas: medida-F (F1) por classe (contrária, favorável e neutra), a sua média ponderada para as três classes ( $F1_{\text{Pond}}$ ), além da métrica utilizada para comparação com os *baselines* ( $F1_{\text{Média}}$ ).

Os alvos Hillary e Trump tiveram os melhores resultados para a  $F1_{\text{Média}}$ . Isso ocorre em função de esses serem os únicos alvos com a F1 balanceada entre as classes de posicionamento (i.e. favorável e contrário). Os demais alvos têm F1 para as classes de posicionamento bastante desbalanceadas em favor da classe prevalente no respectivo corpus.

É interessante observar através da Tabela 5.3, a diferença de enfoque entre a métrica utilizada pelo Semeval ( $F1_{\text{Média}}$ ) e a métrica mais usual para avaliação de sistemas de detecção de classes ( $F1_{\text{Pond}}$ ). Enquanto a  $F1_{\text{Pond}}$  é calculada com base no peso em que cada classe tem no corpus, a  $F1_{\text{Média}}$  privilegia o equilíbrio entre o desempenho obtido para as classes de posicionamento favorável e contrário.

Ao observar a  $F1_{\text{Pond}}$ , é possível verificar um desempenho variando de 53,39 (Trump) a 72,19 (Clima). Avaliando os sistemas através desta métrica, observa-se um ótimo desempenho para o alvo Clima (72,19) e resultados significativos para os demais alvos, com exceção do Ateísmo: Trump (53,39), Feminismo (60,33), Aborto (61,21) e Hillary (63,70). Contudo este

Tabela 5.3: Medida-F por Alvo X Classe

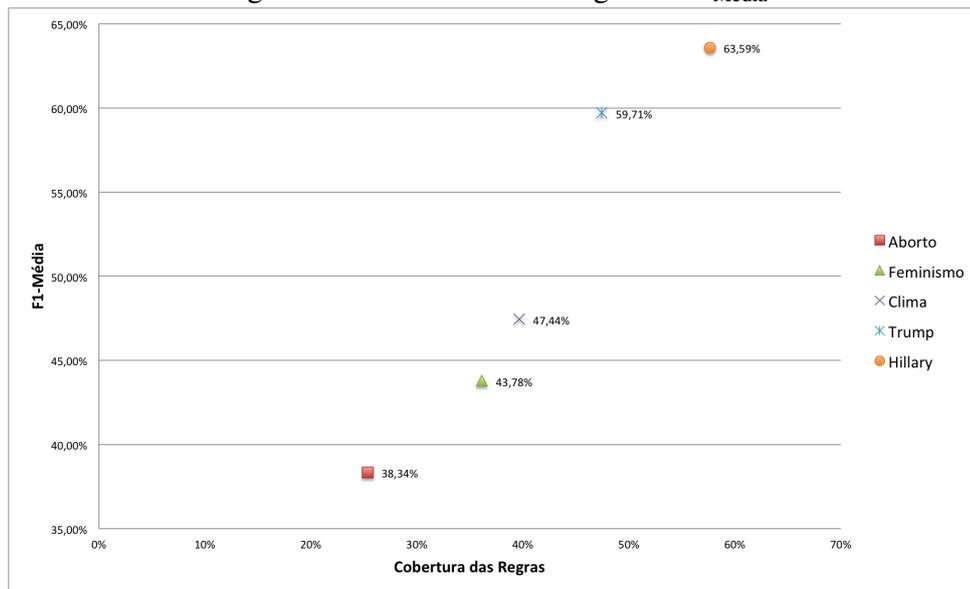
ALVO	CLASSE	F1	F1 <sub>Pond</sub>	F1 <sub>Média</sub>
Trump	CONTRA	<b>59,49</b>	53,39	59,71
	FAVOR	<b>59,92</b>		
	NEUTRO	27,27		
Hillary	CONTRA	<b>73,12</b>	63,70	63,59
	FAVOR	<b>54,55</b>		
	NEUTRO	23,30		
Clima	CONTRA	14,29	72,19	47,44
	FAVOR	<b>80,60</b>		
	NEUTRO	28,57		
Feminismo	CONTRA	<b>76,61</b>	60,33	43,78
	FAVOR	10,96		
	NEUTRO	22,95		
Aborto	CONTRA	<b>72,99</b>	61,21	38,34
	FAVOR	3,70		
	NEUTRO	16,67		
Ateísmo	CONTRA	0,00	2,06	0,00
	FAVOR	0,00		
	NEUTRO	22,67		

critério não permitiria a comparação com outros sistemas que tenham explorado o problema de detecção de posicionamento em textos de tweets.

A partir dos resultados obtidos ( $F1_{Média}$ ), constata-se que o processo desenvolvido pode ser aplicado com êxito a alvos/domínios que se caracterizam pela tendência dos usuários em expressar opinião direcionada ao alvo primário e em usar hashtags para reforçar o posicionamento. Já para alvos/domínios que apresentam diferentes características quanto à forma de expressar posicionamento (e.g. Clima e Ateísmo), torna-se necessário que sejam buscadas alternativas de detecção de posicionamento, tais como o uso de expressões ou perfis semente (Seção 5.5.2 e 5.5.3).

#### 5.4 Avaliação das Estratégias Usadas pelo Processo

Nesta seção, são analisadas as etapas do processo com base no desempenho obtido nos experimentos #1 e #2. Para a etapa de detecção de posicionamento baseado em heurísticas, são analisadas a cobertura e a precisão das regras bem como o desempenho da tarefa de detecção de polaridade. Para a etapa de classificação complementar, é analisado o desempenho do respectivo modelo preditivo construído.

Figura 5.1: Cobertura das Regras x  $F1_{Média}$ 

Fonte: o autor.

#### 5.4.1 Avaliação das Estratégias de Detecção de Posicionamento Baseado em Heurísticas

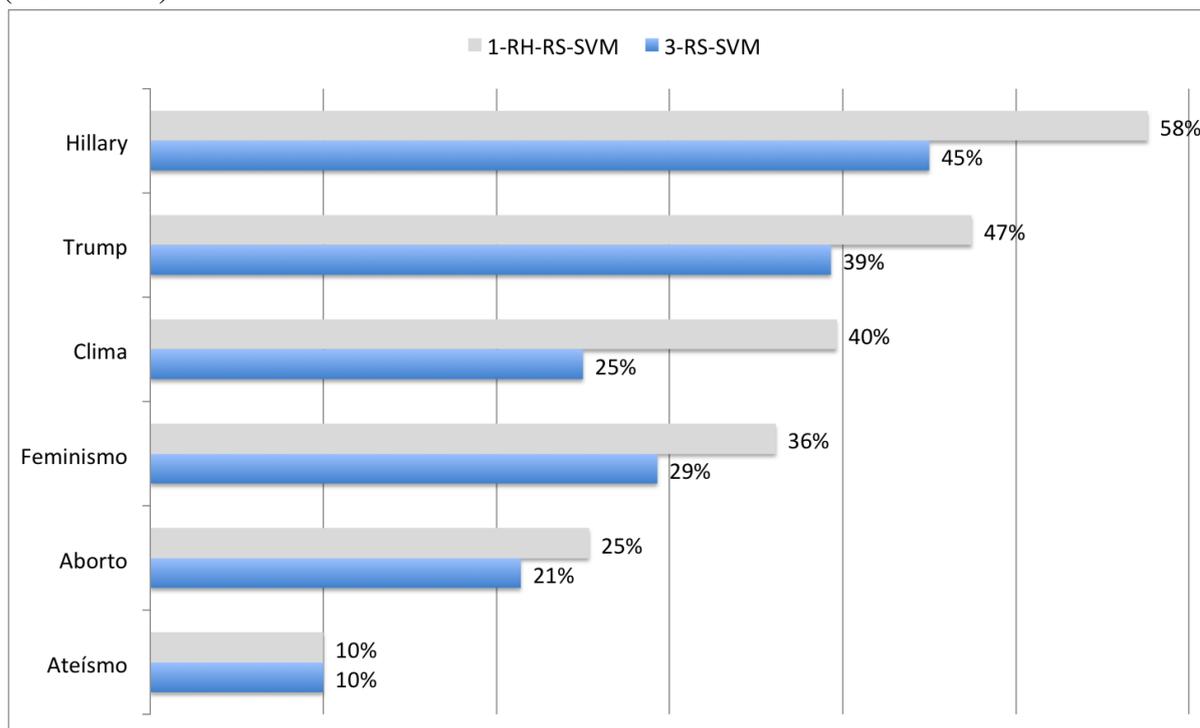
Nesta seção, avalia-se a contribuição das heurísticas propostas para a tarefa de detecção de posicionamento. Esta análise é feita com base na representatividade de cada uma das regras nos corpora estudados (Seção 5.4.1.1), precisão em relação ao posicionamento a ser detectado (Seção 5.4.1.2) e precisão da polaridade utilizada para as regras de sentimento (Seção 5.4.1.3).

##### 5.4.1.1 Cobertura das Heurísticas

Resultados preliminares desta pesquisa (DIAS; BECKER, 2016b; DIAS; BECKER, 2016c; DIAS; BECKER, 2016a) mostraram que quanto maior for a cobertura das regras, melhor será o desempenho do processo. A Figura 5.1 mostra que os resultados obtidos pelos experimentos #1 e #2 confirmam esta tendência. Isso ocorre porque o desempenho das regras é superior ao desempenho da etapa de classificação complementar, como será detalhado na Seção 5.4.2. O melhor resultado foi obtido para o alvo Hillary (63,59), que também apresentou maior cobertura percentual das regras (58%). O alvo Trump apresenta a segunda maior cobertura geral (47%) e, conseqüentemente, o segundo melhor resultado (59,71).

É importante notar que as hashtags de posicionamento, ainda que extraídas com base na expressão direta de sentimento ao alvo principal, têm impacto direto no aumento da cobertura. Esse fato é evidenciado pela Figura 5.2, que compara a cobertura das 5 regras propostas na Tabela 4.1, com o uso exclusivo de regras sentimentos (i.e. apenas regras 3, 4 e 5), referenciada

Figura 5.2: Cobertura: Processo completo (1-RH-RS-SVM) x Apenas regras de sentimento (3-RS-SVM)



Fonte: o autor.

na figura como 3-RS-SVM<sup>1</sup>. Nota-se que com o uso das hashtags de posicionamento, o aumento da cobertura das regras é significativo para os alvos Clima (15 pontos percentuais - pp), Hillary (13 pp), Trump (8 pp) e Feminismo (7 pp). Apenas para o alvo Aborto o ganho foi pouco significativo (4 pp), em função da menor cobertura obtida para as regras 3 e 4. Já o alvo Ateísmo não apresentou ganho, pela característica já mencionada de os tweets apresentarem menções a alvos de opinião em oposição ao Ateísmo.

Esta comparação permite concluir que a mineração de opinião não é suficiente para a detecção de posicionamento, mas funciona como um elemento muito importante pois auxilia a encontrar outro elemento fundamental de posicionamento, i.e. hashtags de posicionamento. Essas regras agem de forma complementar, já que os tweets que não seriam rotulados pela aplicação das técnicas de mineração de opinião (i.e. menção ao alvo combinada com polaridade) acabam sendo rotulados em função de menções às hashtags detectadas. A Tabela 5.4 apresenta os dicionários de hashtags de posicionamento gerados para cada alvo, de forma a avaliar quantitativa e qualitativamente as hashtags detectadas.

Para verificar a representatividade de cada uma das regras propostas na Tabela 4.1, a Figura 5.3 detalha a respectiva cobertura quando aplicada ao corpus de teste de cada alvo.

<sup>1</sup>Varição 3 do processo, usando apenas regras de sentimento (RS) combinadas com classificador SVM

Tabela 5.4: Dicionários de hashtags de posicionamento por Alvo

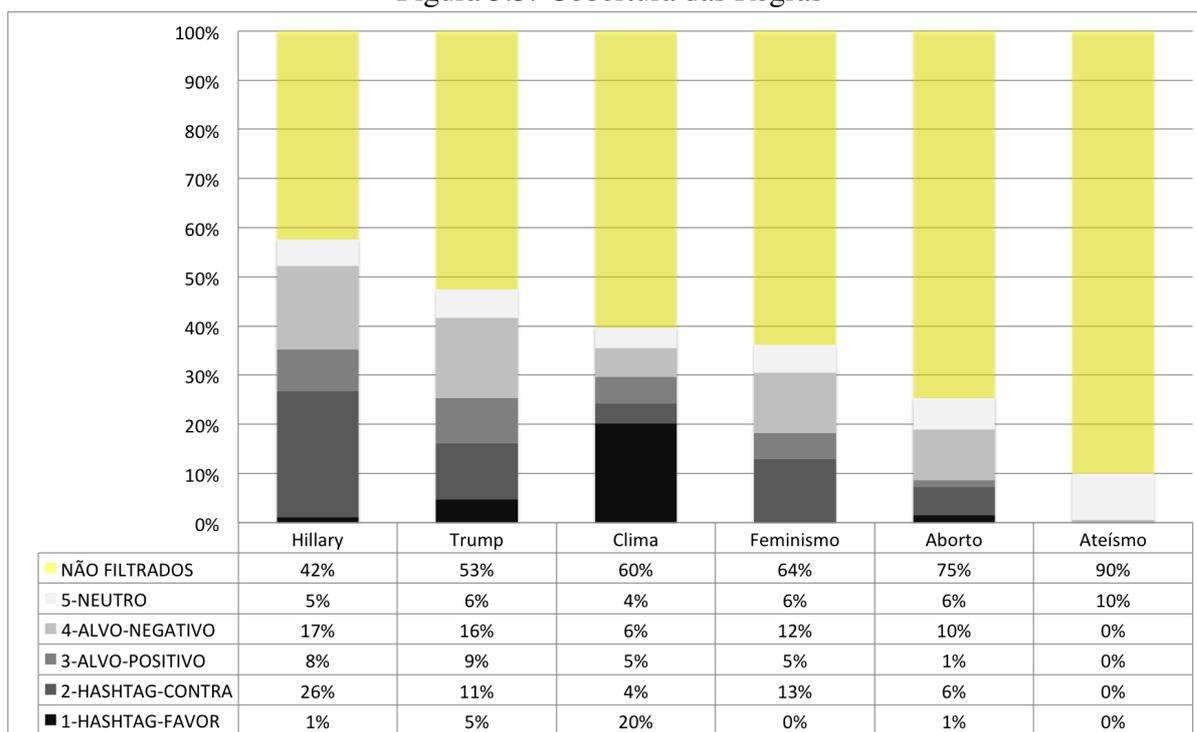
Alvo	HASHTAGS	
	FAVOR	CONTRA
Ateísmo	-	-
Clima	#csota, #actonclimate, #ff, #loveourplanet #mychangeforclimate, #sustainabledevelopment #green, #climatehealth, #cfcc15, #tip #mission, #greatman, #ghg, #abpoli, #cop21	#earthquake #floods #auspol
Feminismo	-	#feminists, #killallmen, #howtospotafeminist #antifeminist, #antifeminism, #heforshe, #wtf #sexism, #mra, #spankafeminis, #sjw #notyourshield, #gamergate, #womensrights, #dontmancriminate, #yesallwomen
Hillary	#itrusthillary	#bernie2016, #hillaryclinton, #liar, #tcot #tlot, #hillaryforprison, #benghazi, #pjnet #obama, #democrats, #whyimnotvotingforhillary #uniteblue, #monicalewinsky, #wakeupamerica #liberals, #potus, #billclinton, #hillaryonenn
Aborto	#prolifeyouth #abortionrights	#deathpenalty, #alllivesmatter, #tcot, #scotus, #abortionismurder, #feminism, #letstalkabortion, #ccot
Trump	#greta, #trumpforpresident, #trumpisright #onethousandtweets, #makeamericagreatagain	#dumptrump, #tcot, #nbc, #cnn, #usa #gop, #latino, #ac360, #yourefired #election2016

Em geral, pode ser verificado que a cobertura das regras que detectam posicionamento contrário (2-HASHTAG-CONTRA e 4-ALVO-NEGATIVO) é superior à cobertura das regras que detectam posicionamento favorável (1-HASHTAG-FAVOR e 3-ALVO-POSITIVO). Isso ocorre para quase todos os alvos, pois o posicionamento contrário é predominante nos corpora, como já destacado na Seção 3.4.1. As exceções são os alvos Ateísmo, que se caracteriza por não apresentar menções ao alvo primário, e Clima, para o qual predomina o posicionamento favorável. Essa maior cobertura das regras de posicionamento contrário reflete a tendência dos usuários do Twitter em expressar seu posicionamento através de críticas.

A cobertura da regra 4-ALVO-NEGATIVO é a mais significativa para todos os alvos, com exceção dos alvos Hillary e Clima. O alvo Hillary tem uma cobertura superior para a regra 2-HASHTAG-CONTRA, indicando uso relevante de hashtags para expressar posicionamento contrário ao alvo, como pode ser observado na Tabela 5.4. Já o alvo Clima tem a regra 1-HASHTAG-FAVOR como a mais significativa. Isso ocorre em função de: a) apresentar um posicionamento com forte prevalência da classe favorável (59,40%) em contraste com a abrangência do posicionamento contrário (4,61%) (Tabela 3.2), b) não ter um alvo de opinião bem definido (“Climate Change is a real Concern”) e c) possuir um número representativo de hashtags detectadas do lado favorável do posicionamento (Tabela 5.4).

A regra oposta 3-ALVO-POSITIVO tem representatividade significativa para todos os alvos, variando de 5% a 9%. A exceção ocorre para o alvo Aborto, para o qual a cobertura

Figura 5.3: Cobertura das Regras



Fonte: o autor.

desta regra é de apenas 1%, impactando, conseqüentemente, na detecção de hashtags de posicionamento e na cobertura da regra 1-HASHTAG-FAVOR (1%). Na Seção 5.5.2, demonstra-se que os usuários do Twitter tendem a usar hashtags de posicionamento para este alvo, mas estas puderam ser detectadas em maior quantidade através do uso de tweets extraídos de perfis semente.

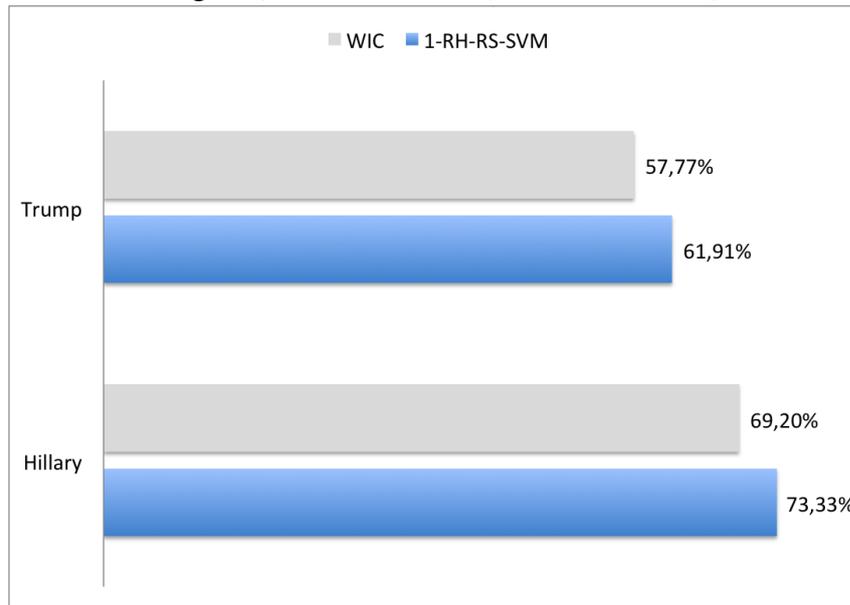
As regras 3 e 4 não possuem cobertura para o alvo Ateísmo. Conseqüentemente, não são detectadas hashtags de posicionamento e a cobertura é nula para as regras 1 e 2. Isso explica o resultado da métrica  $F1_{Média}$  (0,00) para este alvo.

A regra 2-HASHTAG-CONTRA apresenta uma cobertura significativa para todos os alvos variando de 4% a 26%. Já a regra 1-HASHTAG-FAVOR possui representatividade apenas para os alvos Trump (5%) e Clima (20%). As hashtags identificadas nos experimentos #1 e #2 são listadas na Tabela 5.4.

A cobertura da regra 5-NEUTRO, que varia de 4% a 6% apenas, não está de acordo com a representatividade da classe de posicionamento neutro. Esta tem seu menor percentual para o alvo Ateísmo (12,73%) e a maior abrangência para o alvo Trump (36,78%).

Como demonstrado, o uso de regras de sentimento, tanto para detecção de posicionamento quanto para extração de hashtags, combinado com o uso de regras baseadas em hashtags, podem compor um sistema capaz de detectar posicionamento em textos de tweets. Os resulta-

Figura 5.4: Precisão das Regras (Média Ponderada): Este Trabalho (1-RH-RS-SVM) X WIC



Fonte: o autor.

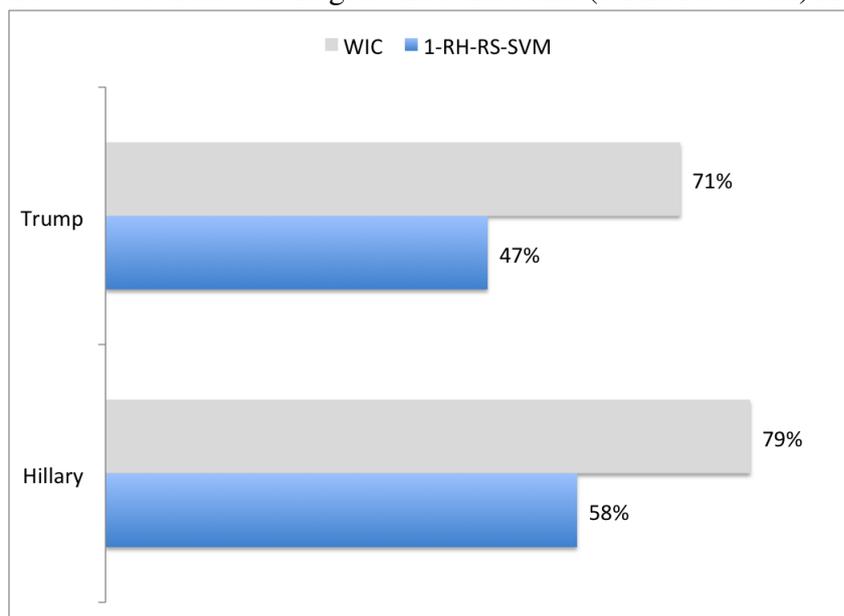
dos são significativos principalmente em alvos/domínios onde a expressão do posicionamento se dá através de sentimento dirigido ao alvo principal do problema, já que esta é a base de detecção automática de hashtags.

#### 5.4.1.2 Qualidade das Heurísticas

Apesar de demonstrado que quanto maior a representatividade das regras, melhor será o resultado do processo, outro componente importante é a qualidade das regras. A Figura 5.4 mostra a média ponderada da precisão das regras deste trabalho comparada à de sua versão preliminar semi-supervisionada (WIC). O presente trabalho apresenta **precisão superior para os alvos Hillary (73,33% contra 69,20%) e Trump (61,91% contra 57,77%), resultando em um desempenho superior para os alvos Hillary (63,59 contra 62,87) e Trump (59,71 contra 50,58)**, como pode ser verificado nas Tabelas 5.1 e 5.2. Porém, o trabalho WIC apresentou uma cobertura (Figura 5.5) bastante superior para os alvos Hillary (79% contra 58%) e Trump (71% contra 47%).

O equilíbrio da qualidade das regras de posicionamento favorável e contrário também é um ponto importante para o desempenho da solução. Na Figura 5.6 é possível verificar a precisão das regras de forma combinada por posicionamento. Nota-se que os alvos de melhor desempenho global (Hillary e Trump) têm a precisão das regras de posicionamento mais balanceado que os alvos com pior desempenho. É possível observar, ainda, que o alvo Clima

Figura 5.5: Cobertura das Regras: Este Trabalho (1-RH-RS-SVM) X WIC



Fonte: o autor.

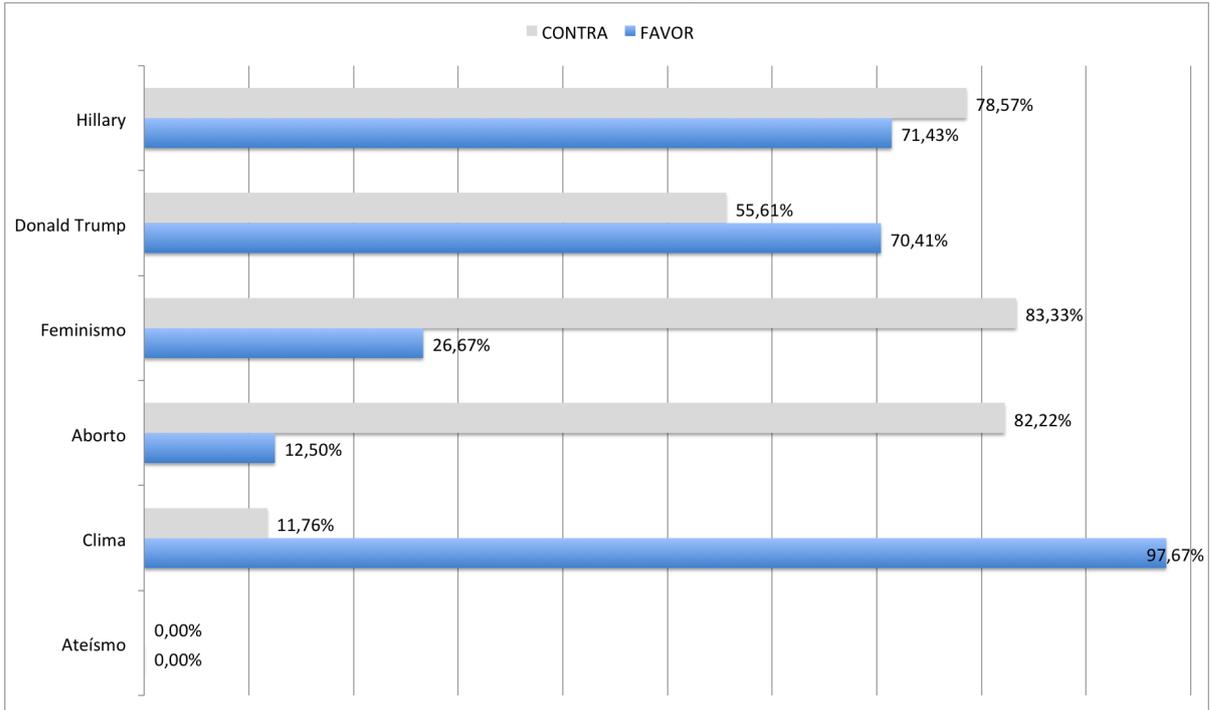
apresenta o melhor resultado para uma das classes (92,67%), mas este resultado não está equilibrado com o da classe contrária (11,76%), resultando em um escore inferior para a  $F1_{Média}$ .

A Figura 5.7 apresenta a precisão das regras de 1 a 5 medidas para cada corpus de teste, permitindo que se analise a qualidade de cada regra definida para o processo de acordo com as características dos alvos.

Ao analisar as regras de forma individual, pode-se perceber que a precedência das regras baseadas em hashtags sobre as regras de sentimento também melhora o desempenho geral do processo. Como evidenciado na Figura 5.7, em geral, as regras baseadas em hashtags (1 e 2) apresentam desempenho próximo ou superior ao das respectivas regras baseadas em sentimento (3 e 4). A exceção ocorre para os alvos Trump e Clima. Para o alvo Trump, a regra 4 apresenta um desempenho superior (67%) ao da regra 2 (40%). Já para o alvo Clima, o desempenho é inaceitável para ambas as regras: 2-HASHTAG-CONTRA (0%) e 4-ALVO-NEGATIVO (20%).

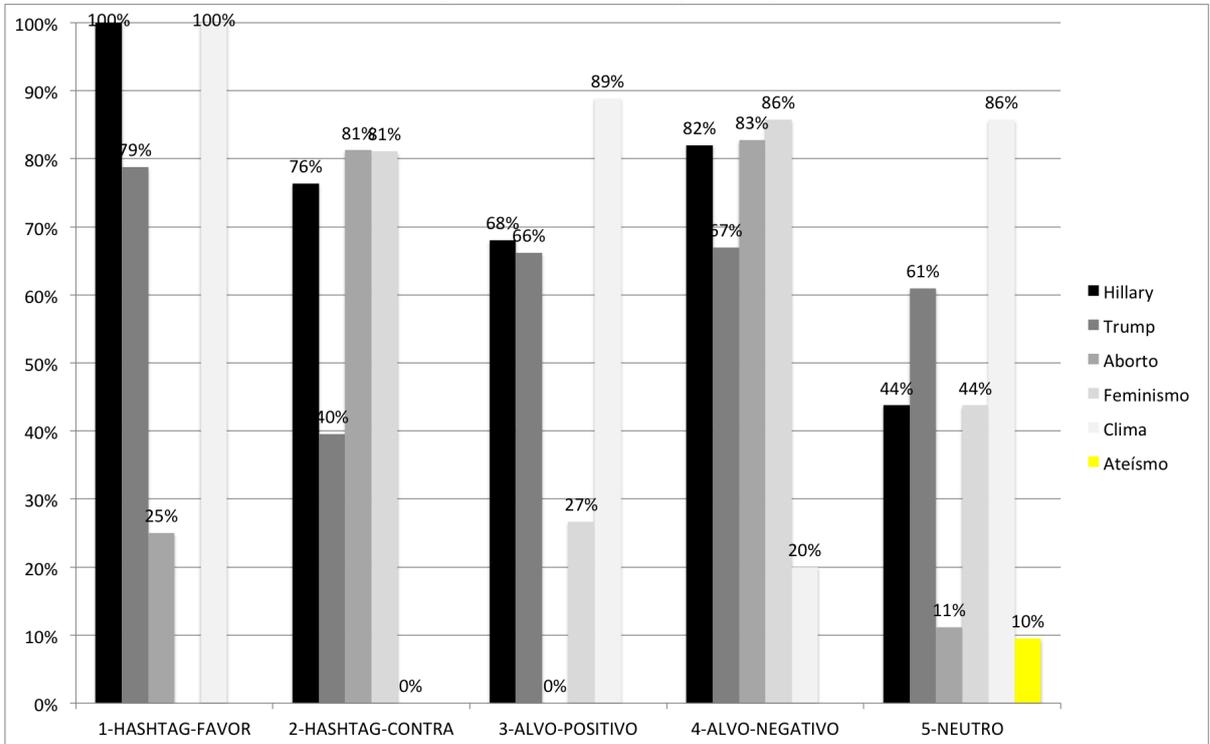
A regra 2-HASHTAG-CONTRA teve um desempenho próximo ao da regra 4-ALVO-NEGATIVO para os alvos Feminismo, Hillary e Aborto, variando entre 76% e 81%. O desempenho dessa regra para o alvo Clima é de 0% de precisão, mas este alvo é o que apresenta a menor representatividade para a classe de posicionamento não prevalente (6,51%). O desempenho dessa regra é inaceitável para o alvo Trump (40%), mas esse resultado não acompanha o bom desempenho das demais regras para este alvo. Isso resulta em uma  $F1_{Média}$  para o alvo Trump (59,71) inferior à obtida para o alvo Hillary (63,59).

Figura 5.6: Precisão das Regras por Posicionamento



Fonte: o autor.

Figura 5.7: Precisão por Regra



Fonte: o autor.

A regra 1-HASHTAG-FAVOR apresentou um desempenho superior para aqueles alvos em que obteve uma cobertura significativa, i.e. Trump e Clima. Foi obtida uma precisão de 79% para uma cobertura de 5% para o alvo Trump. Já para o alvo Clima, foi obtida uma precisão de 100% para uma cobertura de 20%. Apesar da baixa representatividade para o alvo Hillary (1%), foi obtida uma precisão de 100%.

A regra 4-ALVO-NEGATIVO teve um bom desempenho para os alvos Feminismo (86%), Hillary (82%) e Aborto (83%), Trump (67%), e inaceitável para o alvo Clima (20%). Esses resultados evidenciam o melhor desempenho da detecção de posicionamento na classe prevalente.

A regra 3-ALVO-POSITIVO apresentou desempenho aceitável para os alvos Hillary (68%) e Trump (66%), e significativo para o alvo Clima (89%), o que é importante, pois a classe favorável é prevalente para este alvo. A precisão desta regra para os alvos Feminismo (27%) e Aborto (0%) é inaceitável.

A regra 5-NEUTRO apresentou resultados em geral inaceitáveis, com exceção dos alvos Clima (86%) e Trump (61%). É interessante observar que esses dois alvos têm a classe neutra como segunda mais prevalente, com percentuais relevantes (36,78% para o alvo Trump e 35,99% para o alvo Clima). A detecção da classe neutra tem importância menor na abordagem utilizada para comparação, pois não compõe a métrica de avaliação, mas pode impactar de forma indireta no resultado global.

Foi demonstrada a melhor qualidade das regras baseadas em hashtags quando comparadas às respectivas regras de sentimento. Isto evidencia que está correta a hipótese de que a expressão de posicionamento é mais clara através do uso de hashtags do que através de sentimento dirigido ao alvo do problema, mesmo que estas hashtags sejam extraídas a partir da opinião emitida.

Demonstrou-se a relevância da qualidade das regras para que o sistema obtenha melhores resultados. Contudo, é importante frisar que esta qualidade deve estar acompanhada de equilíbrio. De acordo com os critérios utilizados, não é suficiente que um sistema tenha um desempenho superior para a maioria dos tweets. É importante que o desempenho seja superior em ambas as classes de posicionamento.

#### *5.4.1.3 Detecção de Polaridade*

A tarefa de detecção de polaridade tem relevância no processo desenvolvido, pois essa informação serve de base para as regras de sentimento, e por consequência, para as regras de hashtags de posicionamento. Como os corpora de treino e teste do Semeval 2016 foram anotados para polaridade, é possível analisar o desempenho desta tarefa comparando os rótulos

obtidos através do uso da API adotada e os dados anotados fornecidos junto com os corpora.

Conforme descrito no Capítulo 4, a proposta deste trabalho utiliza a API IBM Alchemy<sup>2</sup>, mas esse componente pode ser facilmente substituído por outra API ou método de detecção de polaridade com base em conteúdo textual de documentos.

A Figura 5.8 apresenta a precisão da detecção de polaridade por alvo e pelas classes de sentimento (positivo e negativo). Para a polaridade, repete-se a característica já percebida no caso do posicionamento: ocorre um melhor desempenho da detecção de polaridade na classe prevalente. Portanto, observa-se um desempenho superior para a classe negativa em geral. A exceção é o alvo Ateísmo, para o qual o posicionamento contrário é expresso através de opiniões favoráveis a temas religiosos.

A melhor precisão obtida para detecção da classe negativa impacta diretamente no resultado obtido pelas regras de sentimento (3 e 4), pois a regra 4-ALVO-NEGATIVO apresenta melhor desempenho que a regra 3-ALVO-POSITIVO para a maioria dos alvos de posicionamento. Outras estratégias podem ser experimentadas com o objetivo de melhorar o resultado global do processo através do incremento da qualidade da detecção de polaridade.

#### 5.4.2 Avaliação da Estratégia de Detecção de Posicionamento Complementar

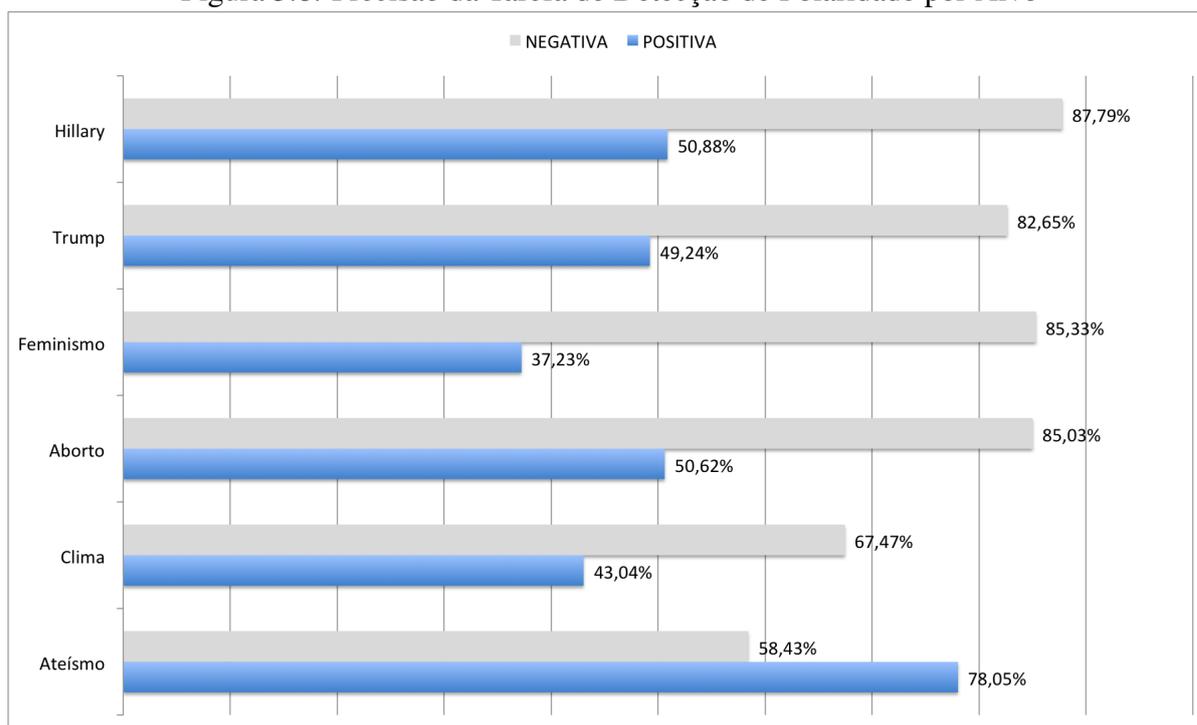
Como já mencionado, a etapa de classificação complementar desempenha um papel importante no processo proposto devido à cobertura das regras. Como pode ser observado através da Figura 5.3, o máximo de cobertura atingido pelas regras foi de 58% para o alvo Hillary, seguido de 47% para o alvo Trump. Os demais alvos apresentam coberturas ainda menores, variando entre 10% e 40%.

A Tabela 5.5 apresenta os resultados para precisão, revocação e medida-F por alvo e por classe. Adicionalmente, é apresentada a média ponderada para cada métrica. Os resultados foram calculados para cada etapa do processo: a) Detecção de Posicionamento por regras (HEURÍSTICAS), b) Detecção de Posicionamento por Classificação Complementar (CLASSIF. COMPLEMENTAR) e c) resultado combinado (PROCESSO COMPLETO). A coluna “QTD” apresenta a quantidade de tweets classificada em cada etapa do processo.

Através da Tabela 5.5, é possível confirmar os resultados superiores da primeira etapa comparados aos da segunda. Esse mesmo comportamento já havia sido observado para versões anteriores do presente trabalho (DIAS; BECKER, 2016b; DIAS; BECKER, 2016a; DIAS;

<sup>2</sup><https://www.ibm.com/watson/alchemy-api.html>

Figura 5.8: Precisão da Tarefa de Detecção de Polaridade por Alvo



Fonte: o autor.

BECKER, 2016c). Isso evidencia a importância da cobertura e da qualidade das regras, como já ressaltado na Seção 5.4.1.

A medida-F ponderada ( $F1_{Pond}$ ) da etapa de classificação complementar é superior para os alvos Clima (66,84) e Aborto (61,08). Já para os alvos Hillary (42,44), Trump (43,40) e Feminismo (54,83) o desempenho é inferior. Para o alvo Ateísmo o desempenho é inaceitável (2,11), em função dos resultados obtidos na primeira fase.

O modelo preditivo para o alvo Hillary apresenta o segundo pior desempenho (42,44). Isto se deve a um modelo de classificação desbalanceado, que classifica a grande maioria dos tweets como contrários. Porém, isto não chega a impactar no resultado final, pois a maioria dos tweets de posicionamento favorável (37) foram rotulados na primeira etapa, restando apenas 8 tweets favoráveis para classificar.

Um modelo que classifica de forma equilibrada apresenta melhor desempenho de acordo com os critérios do Semeval 2016. O modelo preditivo gerado para o alvo Trump tem o resultado F1 mais equilibrado, justificado pelas respectivas medidas-F, 57,08 para a classe contrária, e 40,00 para a classe favorável. Consequentemente, o resultado final é mais equilibrado, resultando em uma  $F1_{Média}$  superior (59,71) a dos demais alvos.

Os demais modelos preditivos, ainda que tenham desempenho ( $F1_{Pond}$ ) superior ao do alvo Trump, tendem a rotular os tweets de acordo com a classe prevalente. Isto ocorre princi-

Tabela 5.5: Resultados por Estapa e Alvo

ALVO	CLASSE	HEURÍSTICAS				CLASSIF. COMPLEMENTAR				PROCESSO COMPLETO				
		QTD	PREC	REV	F1	QTD	PREC	REV	F1	QTD	PREC	REV	F1	F1 <sub>Média</sub>
Hillary	CONTRA	115	78,57	86,09	82,16	57	45,22	91,23	60,47	172	62,66	87,79	73,12	63,59
	FAVOR	37	71,43	54,05	61,54	8	0,00	0,00	0,00	45	68,97	44,44	54,05	
	NEUTRO	18	43,75	38,89	41,18	60	55,56	8,33	14,49	78	48,00	15,38	23,30	
	MÉD. POND.	170	73,33	74,12	73,33	125	36,47	62,60	42,44	295	62,48	70,69	63,70	
Trump	CONTRA	151	55,61	72,19	62,82	148	41,27	92,57	57,08	299	46,59	82,27	59,49	59,71
	FAVOR	119	70,41	57,98	63,59	29	72,73	27,59	40,00	148	70,64	52,03	59,92	
	NEUTRO	65	60,98	38,46	47,17	195	68,97	10,26	17,86	260	64,29	17,31	27,27	
	MÉD. POND.	335	61,91	60,60	60,06	372	57,82	53,51	43,40	707	58,57	58,92	53,39	
Feminismo	CONTRA	76	83,33	78,95	81,08	107	59,12	100,00	74,31	183	66,01	91,26	76,61	43,78
	FAVOR	19	26,67	21,05	23,53	39	0,00	0,00	0,00	58	26,67	6,90	10,96	
	NEUTRO	8	43,75	87,50	58,33	36	0,00	0,00	0,00	44	41,18	15,91	22,95	
	MÉD. POND.	103	69,81	68,93	68,70	182	43,62	73,79	54,83	285	56,82	69,84	60,33	
Aborto	CONTRA	58	82,22	63,79	71,84	131	62,23	89,31	73,35	189	66,09	81,48	72,99	38,34
	FAVOR	8	12,50	12,50	12,50	38	0,00	0,00	0,00	46	12,50	2,17	3,70	
	NEUTRO	5	11,11	40,00	17,39	40	23,81	12,50	16,39	45	17,95	15,56	16,67	
	MÉD. POND.	71	69,36	56,34	61,32	209	52,52	73,84	61,08	280	56,67	67,90	61,21	
Clima	CONTRA	3	11,76	66,67	20,00	8	0,00	0,00	0,00	11	11,76	18,18	14,29	47,44
	FAVOR	57	97,67	73,68	84,00	66	64,71	100,00	78,57	123	74,48	87,80	80,60	
	NEUTRO	7	85,71	85,71	85,71	28	0,00	0,00	0,00	35	85,71	17,14	28,57	
	MÉD. POND.	67	92,58	74,63	81,31	102	55,05	85,07	66,84	169	72,85	77,30	72,19	
Ateísmo	CONTRA	12	0,00	0,00	0,00	148	0,00	0,00	0,00	160	0,00	0,00	0,00	0,00
	FAVOR	8	0,00	0,00	0,00	24	0,00	0,00	0,00	32	0,00	0,00	0,00	
	NEUTRO	2	9,52	100,00	17,39	26	13,13	100,00	23,21	28	12,79	100,00	22,67	
	MÉD. POND.	22	0,87	9,09	1,58	198	1,19	9,09	2,11	220	1,16	9,09	2,06	

Tabela 5.6: Instâncias de Treino do Modelo Preditivo

ALVO	FAVORÁVEL		CONTRÁRIO		NEUTRO		TOTAL
	QTD	PREC.(%)	QTD	PREC.(%)	QTD	PREC.(%)	
Hillary	28	71	132	78	79	24	239
Trump	98	70	196	56	41	61	335
Feminismo	33	39	268	79	62	31	363
Aborto	45	9	161	70	63	27	269
Clima	96	70	44	11	31	58	171
Ateísmo	0	0	4	50	62	15	66

palmente com os alvos Clima e Feminismo, para os quais todas as instâncias receberam o rótulo da respectiva classe prevalente. Contudo, também ocorre para os alvos Hillary e Aborto, para os quais alguns poucos tweets são classificados como neutros.

A classe neutra em todos os casos apresenta resultados (F1) inaceitáveis quando avaliada tanto para o processo completo, quanto para cada uma das etapas dos processos. Acredita-se que isso se deve principalmente à dificuldade de determinar uma regra com bom desempenho para a detecção de tweets nesta classe.

Apesar de a Tabela 5.5 apresentar as métricas calculadas somente para os tweets dos corpora de teste, é importante lembrar que todos tweets anotados dos corpora de treino/teste do Semeval 2016 foram rotulados durante a primeira etapa visando compor o corpus de treino utilizado para treinar o classificador nesta segunda etapa. A Tabela 5.6 apresenta a distribuição destes corpora de treino e a qualidade dos rótulos gerados medida através da métrica precisão.

Tabela 5.7: Experimentos com Variações do Processo

VARIACÃO	ABORDAGEM SUPERVISÃO-NADA	ETAPA 1 - HEURÍSTICAS						ETAPA 2 - CLASSIFICAÇÃO COMPLEMENTAR			
		ALVOS SEMENTE (AS)	PERFIS SEMENTE PARA DETECTAR HASHTAGS (PSH)	DETECÇÃO POR POLARIDADE (POL)	REGRAS DE HASHTAGS DE POSICIONAMENTO (RH)	REGRAS DE SENTIMENTO (RS)	REGRAS DE SENTIMENTO PARA ALVO CONTRÁRIO (RC)	PERFIS SEMENTE PARA TREINO (PST)	ATRIBUTOS ADICIONAIS (AA)	EMBEDDINGS (EMB)	ALGORITMO
1-RH-RS-SVM	NAO				X						SVM
2-POL	NAO			X							SVM
3-RS-SVM	NAO					X					SVM
4-RH-SVM	NAO				X	NEUTRO					SVM
5-AS-RH-RS-RC-SVM	SEMI	X			X	X	X				SVM
6-PSH-RH-RS-SVM	SEMI		X		X	X					SVM
7-RH-RS-PST-SVM	SEMI				X	X		X			SVM
8-RH-RS-AA-SVM	NAO				X	X			X		SVM
9-RH-RS-EMB-SVM	NAO				X	X				X	SVM
10-RH-RS-SL	NAO				X	X					SL
11-RH-RS-RF	NAO				X	X					RF
12-RH-RS-VOTE	NAO				X	X					VOTE

fonte: o autor

Os corpora de treino apresentados na Tabela 5.6 possuem poucas instâncias e não estão balanceados. Na Seção 5.5.3, são apresentados experimentos no sentido de aumentar o tamanho dos corpora e melhorar a qualidade do modelo preditivo.

Em conclusão, a etapa de classificação complementar é necessária para rotular todos os tweets e é o componente do processo que requer mais atenção a fim de melhorar a sua qualidade contribuindo para o resultado final do sistema. Foram experimentadas alternativas para melhoria do modelo preditivo com certo êxito de acordo com o alvo e domínio, porém impactando na entrada necessária para execução do processo. Essas alternativas serão apresentadas na Seção 5.5.3.

### 5.5 Experimento #3 - Variações do Processo

Nesta seção serão avaliadas variações no processo principal divididas em três grupos: a) componentes do processo (Seção 5.5.1), b) uso de alvos e perfis semente na etapa de heurísticas (Seção 5.5.2) e c) alternativas de construção do modelo preditivo (Seção 5.5.3).

A Tabela 5.7 identifica e diferencia cada variação indicando os componentes utilizados por cada uma delas agrupados por etapa do processo e se a variação resulta em uma solução não supervisionada ou semi-supervisionada. As variações do processo proposto são identificadas por composições de siglas que identificam cada componente utilizado. A Tabela 5.8 descreve de forma resumida o significado de cada sigla.

A Tabela 5.9 apresenta o resultado obtido através de cada variação para cada alvo utilizando a métrica de avaliação do Semeval 2016 ( $F1_{Média}$ ). A primeira linha (1-RH-RS-SVM) representa os resultados obtidos através do processo proposto, que é usado como *baseline*.

<sup>3</sup><https://nlp.stanford.edu/projects/glove/>

<sup>4</sup>implementação do SVM disponível na API Java do Weka

<sup>5</sup>disponível na API Java do Weka

Tabela 5.8: Componentes Utilizados nas Variações do Processo

SIGLA	NOME	DESCRIÇÃO
RH	Regras de Hashtags p/ de Posicionamento	<ul style="list-style-type: none"> <li>- objetivo: rotular tweets automaticamente com base em hashtags de posicionamento</li> <li>- implementado através das regras 1-HASHTAG-FAVOR e 2-HASHTAG-CONTRA</li> <li>- comparam o número de menções de hashtags favoráveis e contrárias ao alvo principal</li> </ul>
RS	Regras de Sentimento	<ul style="list-style-type: none"> <li>- objetivo: rotular tweets automaticamente com base em menção ao alvo combinada com a polaridade do tweet</li> <li>- regras 3-ALVO-POSITIVO, 4-ALVO-NEGATIVO e 5-NEUTRO</li> <li>- verificam menção ao alvo principal e, caso existam, outros alvos semente favoráveis ao alvo principal</li> <li>- 5-NEUTRO considera a ausência de hashtags</li> </ul>
POL	Detecção por Polaridade	<ul style="list-style-type: none"> <li>- objetivo: rotular tweets automaticamente com base na polaridade</li> <li>- tweets positivos são rotulados como favoráveis, negativos como contrários e neutros como neutros</li> <li>- dispensa a detecção de hashtags de posicionamento, o uso de regras e a etapa de classificação complementar</li> </ul>
AS	Alvos Semente	<ul style="list-style-type: none"> <li>- objetivo: permitir a identificação de menções favoráveis/contrárias ao alvo principal através de alvos indiretamente relacionados</li> <li>- alvos adicionais são inseridos como entrada do processo</li> <li>- são relacionados ao alvo principal com anotação contrário/favorável</li> </ul>
RC	Regras de Sentimento p/Alvo Contrário	<ul style="list-style-type: none"> <li>- objetivo: rotular tweets automaticamente com base em menção a alvos semente contrários ao alvo principal e a polaridade do tweet</li> <li>- regras 6-ALVO-CONTRA-POSITIVO e 7-ALVO-CONTRA-NEGATIVO</li> <li>- uso combinado com alvos semente (AS)</li> </ul>
PSH	Perfis Semente p/ Detectar Hashtags	<ul style="list-style-type: none"> <li>- objetivo: incrementam o número de hashtags de posicionamento detectadas</li> <li>- são perfis claramente favoráveis/contrários ao alvo principal e recebem este rótulo</li> <li>- tweets destes perfis são coletados recebendo o mesmo rótulo do perfil semente</li> <li>- os tweets coletados são inseridos como entrada da tarefa de detecção de hashtags de posicionamento</li> </ul>
PST	Perfis Semente para Treino	<ul style="list-style-type: none"> <li>- objetivo: incrementar o volume do corpus de treino da etapa de classificação complementar</li> <li>- são perfis claramente favoráveis/contrários ao alvo principal e recebem este rótulo</li> <li>- tweets destes perfis são coletados recebendo o mesmo rótulo do perfil semente</li> <li>- os tweets coletados são inseridos como entrada da tarefa de extração de atributos para compor o corpus de treino</li> </ul>
AA	Atributos Adicionais	<ul style="list-style-type: none"> <li>- objetivo: melhorar a qualidade do modelo preditivo através de atributos discriminantes</li> <li>- são extraídos dos corpora durante o processo e adicionados aos corpora de treino e teste usados para construção e aplicação do classificador complementar</li> <li>- polaridade (negativa, positiva ou neutra)</li> <li>- escore de polaridade contendo valores entre -1 e 1</li> <li>- quantidade de hashtags</li> <li>- atributo binário indicando menção ao alvo principal</li> </ul>
EMB	Embeddings	<ul style="list-style-type: none"> <li>- objetivo: melhorar a qualidade do modelo preditivo através de atributos discriminantes</li> <li>- embedding pré-treinado GloVe<sup>3</sup> 200-d é inserido como entrada na tarefa de extração de atributos</li> <li>- gera 200 atributos representando o vetor médio das palavras que compõem cada tweet</li> </ul>
SVM	SVM	<ul style="list-style-type: none"> <li>- objetivo: algoritmo usado para construção do modelo preditivo da etapa de classificação complementar</li> <li>- algoritmo SMO do Weka<sup>4</sup></li> </ul>
SL	Simple Logistic	<ul style="list-style-type: none"> <li>- objetivo: melhorar o resultado da etapa de classificação complementar</li> <li>- algoritmo de classificação Simple Logistic substitui o SVM</li> </ul>
RF	Random Forest	<ul style="list-style-type: none"> <li>- objetivo: melhorar o resultado da etapa de classificação complementar</li> <li>- algoritmo de classificação Random Forest substitui o SVM</li> </ul>
VOTE	Comitê	<ul style="list-style-type: none"> <li>- objetivo: melhorar o resultado da etapa de classificação complementar</li> <li>- meta-algoritmo VOTE<sup>5</sup></li> <li>- foi escolhida uma combinação de algoritmos para compor o comitê: SMO, Simple Logistic e Random Forest</li> <li>- comitê calcula a média das probabilidades dos algoritmos por classe para realizar a classificação</li> </ul>

Fonte: o autor

Tabela 5.9: Resultados dos Experimentos

VARIAÇÃO	ATEÍSMO	CLIMA	FEMINISMO	HILLARY	ABORTO	TASK A	TRUMP
<b>1-RH-RS-SVM</b>	<b>0,00</b>	<b>47,44</b>	<b>43,78</b>	<b>63,59</b>	<b>38,34</b>	<b>59,01</b>	<b>59,71</b>
2-POL	23,33	28,62	47,61	50,06	34,99	41,11	49,19
3-RS-SVM	0,00	36,25	42,04	58,42	32,11	48,13	55,78
4-RH-SVM	0,00	41,88	39,47	41,60	34,53	57,12	43,78
5-AS-RH-RS-RC-SVM	<b>54,02</b>	47,20	<b>51,79</b>	63,59	43,44	<b>61,73</b>	<b>59,92</b>
6-PSH-RH-RS-SVM	10,20	<b>51,39</b>	43,14	63,66	<b>50,20</b>	56,00	55,53
7-RH-RS-PST-SVM	47,12	40,18	42,50	<b>64,51</b>	47,67	51,34	47,53
8-RH-RS-AA-SVM	0,00	47,44	43,35	63,39	37,5	58,49	59,53
9-RH-RS-EMB-SVM	0,00	47,44	43,73	62,25	38,19	58,73	58,96
10-RH-RS-SL	0,00	45,98	42,77	60,94	20,03	52,54	56,49
11-RH-RS-RF	0,00	47,44	43,78	63,76	39,78	59,57	57,69
12-RH-RS-VOTE	0,00	47,22	43,78	63,59	38,34	58,87	58,88

fonte: o autor

### 5.5.1 Avaliação dos Componentes do Processo

O objetivo desta seção é demonstrar a importância dos componentes inseridos no processo para a obtenção dos resultados atingidos para os alvos que representam figuras públicas (i.e. Hillary e Trump). Tanto as regras de sentimento quanto as regras baseadas em hashtags contribuem para um melhor resultado principalmente quando aplicadas a alvos que identifiquem figuras públicas. Para confirmar isto, foram realizadas 3 variações do processo proposto:

- **2-POL:** consiste em detectar o posicionamento exclusivamente com base na polaridade dos tweets que contêm menções ao alvo principal. Tweets com polaridade positiva são rotulados como favoráveis, com polaridade negativa são rotulados como contrários e os tweets de polaridade neutra são rotulados como neutros. Esta abordagem dispensa a etapa de classificação complementar, bem como toda a parte do processo relacionada à identificação de hashtags, já que a totalidade dos tweets foram polarizados;
- **3-RS-SVM:** consiste em detectar o posicionamento na primeira etapa do processo através da aplicação das regras de sentimento 3, 4 e 5, eliminando toda a parte do processo relacionada à identificação e aplicação das regras baseadas em hashtags;
- **4-RH-SVM:** elimina-se as regras 3 e 4 da tarefa de detecção de posicionamento da etapa. Essas duas regras de sentimento são utilizadas somente na tarefa de detecção de posicionamento preliminar para permitir a detecção de hashtags de posicionamento e construção do dicionário. O posicionamento é definido usando somente as regras 1, 2 e 5.

A simples detecção da polaridade não é suficiente para um processo de detecção de posicionamento em relação a um dado alvo, confirmando que há várias formas de expressar posicionamento, conforme já demonstrado no Capítulo 1. Os resultados (Tabela 5.9) da variação

*2-POL* são bastante inferiores para quase todos os alvos quando comparados ao processo proposto (*1-RH-RS-SVM*). Os únicos alvos para os quais os resultados são superiores são Ateísmo e Feminismo. Isso ocorre pela representatividade fraca dos alvos utilizados como entrada (i.e. “Atheism” e “Feminist Movement”).

Os resultados inferiores da variação *3-RS-SVM*, já descritos na Seção 5.4.1.1, confirmam que a mineração de opinião pode ser utilizada como técnica para apoiar a detecção de posicionamento, mas não é suficiente.

Finalmente, a cobertura obtida utilizando somente as regras baseadas em hashtags (variação *4-RH-SVM*) é insuficiente tanto para rotular um volume de tweets que gere bons resultados, quanto para formar um corpus de treino para o classificador complementar. Os resultados obtidos por esta variação, quando comparados aos do processo proposto (*1-RH-RS-SVM*), mostram que as regras de hashtags cobrem apenas uma parcela menor dos tweets em um processo de detecção de posicionamento, como pode ser verificado na Figura 5.2.

Dessa forma, através dos resultados superiores obtidos para os alvos relacionados a figuras públicas, pôde ser verificada a importância da combinação dos componentes utilizados pela etapa de heurísticas do processo proposto.

### 5.5.2 Avaliação do Uso de Estratégias Semi-supervisionadas na Etapa de Heurísticas

As duas estratégias exploradas nesta seção envolvem a entrada de expressões semente ou de perfis semente, configurando os sistemas propostos como semi-supervisionados.

A variação *5-AS-RH-RS-RC-SVM*, baseada na entrada de alvos semente, busca inspiração em versões anteriores deste trabalho (DIAS; BECKER, 2016b; DIAS; BECKER, 2016a; DIAS; BECKER, 2016c). Esses trabalhos utilizaram a estratégia de inspeção e seleção de n-gramas relacionados ao alvo para compor o que chamamos no presente trabalho de alvos semente. Já a variação *6-PSH-RH-RS-SVM*, busca inspiração em Rajadesingan and Liu (2014), através da indicação de perfis semente notoriamente apoiadores ou opositores ao alvo principal para que sejam coletados seus tweets e componham o corpus necessário para detecção de hashtags de posicionamento.

A primeira alternativa (*5-AS-RH-RS-RC-SVM*) a ser apresentada diferencia-se do processo proposto (*1-RH-RS-SVM*) pelo uso de alvos semente e de regras em oposição ao alvo principal de posicionamento (Tabela 5.10). Dessa forma, se baseia no fato observado de que tweets podem expressar posicionamento favorável ou contrário a um alvo de forma indireta, através de alvos de opinião claramente apoiadores ou opositores do alvo principal. Por exem-

plo, ao emitir um sentimento positivo direcionado ao alvo contrário “god”, está sendo emitido um posicionamento contrário ao alvo principal em um debate sobre Ateísmo. Ou ao expressar uma opinião positiva dirigida ao alvo “republican” em um debate sobre as eleições presidenciais de 2017, pode-se inferir um posicionamento favorável ao alvo principal Trump. Esta estratégia demonstrou-se eficiente, principalmente para o alvo Ateísmo, pois se expressa o posicionamento muito mais com base em temas religiosos do que propriamente em ateísmo.

A segunda alternativa (*6-PSH-RH-RS-SVM*) insere o componente de perfis semente no processo proposto (*1-RH-RS-SVM*). Esses perfis semente (Tabela 5.11) são notoriamente apoiadores ou opositores do alvo principal e tem seu posicionamento propagado para seus tweets, que são coletados pelo processo. São extraídas hashtags destes tweets a fim de enriquecer o dicionário de hashtags de posicionamento, visando aumentar a cobertura das regras baseadas em hashtags. Esta estratégia resultou em um bom desempenho em domínios onde não há um alvo representativo (e.g. Clima e Aborto).

A seguir, as variações exploradas nesta seção são detalhadas:

- **5-AS-RH-RS-RC-SVM:** esta variação do experimento consiste em inserir alvos semente adicionais na entrada do processo. Alvos semente recebem rótulos indicando seu posicionamento em relação ao alvo principal: a) favoráveis ou b) contrários. Os primeiros são alvos que são tratados como se a opinião expressa estivesse sendo direcionada ao alvo principal, porém de forma indireta. Já o alvos contrários são opositores do alvo principal. Os alvos que forem anotados como favoráveis serão utilizados para a detecção de menção ao alvo para aplicação das regras de sentimento (3-ALVO-POSITIVO, 4-ALVO-NEGATIVO e 5-NEUTRO). Já os alvos anotados como contrários serão utilizados para aplicação das regras de sentimento contrário (6-ALVO-CONTRA-POSITIVO e 7-ALVO-CONTRA-NEGATIVO) (Tabela 5.10). As regras 6 e 7 atuam de forma similar às regras de sentimento (i.e. regras 3 e 4), porém consistem em verificar menções a alvos contrários. Neste experimento, as regras 3 e 4 são mutuamente exclusivas em relação às regras 6 e 7. Ainda, o tweet deve conter menções somente a alvos favoráveis para aplicação das regras 3 e 4, ou somente a alvos contrários para aplicação das regras 6 e 7. Os alvos usados estão listados na Tabela 5.12 e foram definidos após a inspeção dos corpora, através da verificação das palavras com maior número de menções para cada alvo;
- **6-PSH-RH-RS-SVM:** consiste em inserir perfis do Twitter como entradas adicionais para a tarefa. Para esta variação, foram escolhidos os perfis da Tabela 5.11, obtidos através de pesquisa manual em mecanismos de busca (e.g. google.com). Os perfis semente devem

Tabela 5.10: Regras de Sentimento Contrário

#	Regra	Descrição	Posicionamento
6	ALVO-CONTRA-POSITIVO	Menção ao alvo contrário, ausência de menção ao alvo favorável e polaridade positiva	Contrário
7	ALVO-CONTRA-NEGATIVO	Menção ao alvo contrário, ausência de menção ao alvo favorável e polaridade negativa	Favorável

fonte: o autor

Tabela 5.11: Perfis Semente utilizados para os experimentos 6-PSH-RH-RS-SVM e 7-RH-RS-PST-SVM

Alvo	Perfis Semente			
	Favorável		Contrário	
	Perfil	Tweets(#)	Perfil	Tweets(#)
Ateísmo	@americanatheist	3.200	@Godly_Life	3.231
Clima	@ClimateReality	3.236	@realDonaldTrump	3.239
Feminismo	@weneedfeminism	3.201	@notofeminism	983
Hillary	@HillaryClinton	3.206	@SenSanders	3.203
			@realDonaldTrump	3.239
Aborto	@prochoiceny	3.229	@prolifecampaign	2.796
Trump	@realDonaldTrump	3.239	@HillaryClinton	3.206

fonte: o autor

ser anotados como favoráveis ou contrários. Então, o processo realizará a coleta de seus tweets mais recentes, propagando o rótulo do perfil semente para estes tweets. Os tweets coletados são inseridos no fluxo do processo como tweets rotulados para uso exclusivo na tarefa de detecção de hashtags de posicionamento.

Além de manter o mesmo nível de desempenho do processo proposto para os alvos Hillary e Trump, a variação 5-AS-RH-RS-RC-SVM foi a que apresentou o melhor resultado para os alvos Ateísmo (54,02) e Feminismo (51,79). O desempenho muito superior aos das demais variações para o alvo Ateísmo se explica pela definição dos alvos semente contrários listados na Tabela 5.12, que permitem a detecção de posicionamento através da menção a estes alvos (e.g. “jesus” e “god”), como já mencionado na Seção 3.4.1. Enquanto o alvo Ateísmo se beneficia da aplicação das regras 6 e 7, o alvo Feminismo é favorecido pelo aumento da cobertura das regras 3 e 4 através de alvos semente favoráveis (e.g. “female”, “woman” e “girls”). O Alvo Aborto apresenta um incremento no resultado de 38,34 para 43,44, o que é um ganho relevante e ocorre de forma similar ao Ateísmo em função dos alvos semente contrários (e.g. “baby” e “child”).

Tabela 5.12: Alvos Semente utilizados para o experimento 5-AS-RH-RS-RC-SVM

Alvo	Alvos Semente	
	Favorável	Contrário
Ateísmo	atheism, atheist	bible, book, christ, church, faith, god, heaven, holy, islam, jesus, lord, mary, prayer, relig, rosary, spirit
Clima	climate, change, concern, real, green, cop21, obama	gop
Feminismo	feminist, movement, feminists, feminism, female, women, woman, girls	family
Hillary	hillary, clinton, democrat	
Aborto	legalization, abortion, abort	catholic, fetus, kids, unborn, god, child, baby, children, babies
Trump	donald, trump, republican	

fonte: o autor

O ponto negativo desta solução é que ela torna o processo semi-supervisionado e requer o esforço de inspeção manual das palavras mais frequentes dos corpora, além de tornar necessário que o anotador tenha algum conhecimento do domínio a que pertence o alvo do problema. Desta forma, a solução é muito sensível a esta entrada, dependendo da qualidade dos alvos semente selecionados e da correta anotação destes.

Mudanças nos debates podem impactar na qualidade das sementes previamente escolhidas. Isto dificulta a aplicação de solução baseada em alvos semente, pois os alvos relacionados de que tratam os tweets podem mudar com o passar do tempo. Por exemplo, durante as prévias para as eleições presidenciais, Bernie Sanders poderia ser um bom alvo semente contrário ao alvo principal Hillary Clinton, por ser seu oponente como pré-candidato à presidência pelos Democratas. Porém, terminadas as prévias, outros alvos podem ter se tornado mais relevantes, como o alvo semente contrário Donald Trump. Esse comportamento tornaria necessária uma revisão periódica dos alvos semente para promover uma maior aderência entre os alvos informados e os alvos efetivamente mencionados nos tweets. Essa necessidade constante de revisão dos alvos semente poderia tornar a solução proibitiva.

A variação 6-PSH-RH-RS-SVM apresenta um resultado bastante superior aos demais para o alvo Aborto (50,20) e o melhor resultado para o alvo Clima (51,39). Os resultados mantêm-se próximos dos obtidos pelo processo proposto (1-RH-RS-SVM) para os alvos Hillary (63,66) e Feminismo (43,14), ocorre uma pequena redução para o alvo Trump (55,53) e o resultado ainda é insuficiente para o alvo Ateísmo (10,20).

A variação 6-PSH-RH-RS-SVM pode ser considerada semi-supervisionada da mesma forma que a variação 5, já que requer uma entrada adicional para execução do sistema. Porém, esta variação do processo é menos sensível à seleção de sementes que a variação 5-AS-RH-RS-RC-SVM, pois são informados apenas perfis semente (Twitter) e não requer a análise de uma grande quantidade de termos. Ainda assim, é uma solução que depende de intervenção manual, resultando em um sistema sensível à qualidade dos perfis semente selecionados e ao conhecimento do domínio por parte do anotador.

Da mesma forma que ocorre para a variação 5-AS-RH-RS-RC-SVM, mudanças de alvos relacionados e de temas podem ocorrer com o passar do tempo. A mesma mudança mencionada anteriormente para o alvo semente Bernie Sanders pode ser aplicada ao perfil semente Bernie Sanders. Este perfil semente pode ter sido um bom perfil semente contrário ao alvo Hillary durante as prévias dos Democratas, mas talvez tenha sido um bom perfil semente favorável durante a campanha presidencial efetiva, por ser do mesmo partido de Hillary.

Resumindo, as soluções envolvendo sementes (alvos ou perfis) recaem basicamente nos problemas abaixo:

- **Esforço Manual de Anotação:** pode ser dispendioso ter que inspecionar termos ou buscar perfis do Twitter para uso como sementes;
- **Conhecimento do Domínio:** a qualidade das sementes selecionadas depende do conhecimento que o anotador possui do domínio em análise;
- **Atualização das Sementes:** a mudança de temas em discussão relacionados ao alvo do problema de posicionamento pode mudar com o passar do tempo, tornando necessária uma atualização das sementes.

Apesar dos problemas mencionados acima e de essas estratégias configurarem um processo semi-supervisionado, elas podem ser utilizadas com resultados relativamente positivos dependendo do conhecimento que se tem do domínio em que será aplicado o processo e do conhecimento que se tem das características do mesmo.

### 5.5.3 Avaliação de Alternativas de Construção do Modelo Preditivo

Nesta seção serão analisadas alternativas avaliadas para melhoria do modelo preditivo utilizando técnicas variadas, tais como enriquecimento do corpus de treino, engenharia de atri-

butos através atributos adicionais ou *embeddings* e uso de comitês de algoritmos de classificação variados. As variações mais relevantes são as que seguem:

- **7-RH-RS-PST-SVM:** perfis semente (Twitter) favoráveis e contrários são selecionados por um anotador. Os tweets destes perfis são coletados automaticamente e são rotulados de acordo com os rótulos dados aos perfis pelo anotador. Após isso, estes tweets são acrescentados ao conjunto de tweets rotulados para compor o corpus de treino do modelo preditivo. Foram utilizados os mesmos perfis semente e tweets coletados utilizados no experimento *6-PSH-RH-RS-SVM* descrito na seção anterior, e que constam na Tabela 5.11;
- **8-RH-RS-AA-SVM:** foram adicionados atributos ao corpus de treino e às instâncias submetidas à tarefa de classificação complementar. Os atributos adicionais são: a polaridade detectada para o tweet através da API Alchemy (negativa, positiva ou neutra), atributo de representatividade binária identificando a presença ou ausência do alvo no tweet, o escore de polaridade detectado para o tweet através da API Alchemy e um atributo numérico informando a quantidade de hashtags presentes no tweet;
- **9-RH-RS-EMB-SVM:** esta variação se baseia em uma técnica recentemente utilizada para identificar palavras relacionadas através do treinamento não supervisionado de *embeddings* com base em um conjunto de documentos textuais, no caso são tweets coletados para construção destes *embeddings* pré-treinados (i.e. GloVe) (PENNINGTON; SOCHER; MANNING, 2014). Para esta variação, foi calculado o vetor médio de cada tweet composto pela média das 200 dimensões dos vetores obtidos para cada palavra que compõe o tweet<sup>6</sup>. Estes vetores são obtidos do *embedding* pré-treinado GloVe. Cada uma das 200 dimensões do vetor médio calculado se converte em um atributo;
- **10-RH-RS-SL:** substituição do algoritmo de classificação SMO (SVM) pelo algoritmo *Simple Logistic*;
- **11-RH-RS-RF:** substituição do algoritmo de classificação SMO (SVM) pelo algoritmo *Random Forest*.
- **12-RH-RS-VOTE:** uso de um comitê de algoritmos de classificação por média das probabilidades calculadas pelos algoritmos de classificação que compõem o comitê. O meta-classificador Vote está disponível na API Weka e foram inseridos como parâmetros três

---

<sup>6</sup><https://softwaremill.com/deep-learning-for-nlp/>

algoritmos de classificação para compor o comitê: Simple Logistic, Random Forest e SMO (SVM);

Através dos resultados apresentados na Tabela 5.9, nota-se que as diferentes estratégias tentadas pouco influenciaram os resultados, em alguns casos até impactando negativamente.

A variação *7-RH-RS-PST-SVM* apresenta o melhor resultado para o alvo Hillary (64,51), quase um ponto superior ao do processo proposto e o segundo melhor resultado para os alvos Ateísmo (47,12) e Aborto (47,67). Mas os resultados são inferiores para os demais alvo, apresentando uma perda relevante para o alvo Trump. Isso pode ter ocorrido pelos fatores levantados na seção anterior: a) qualidade da seleção das sementes e b) desatualização das sementes considerando os diferentes momentos de coleta, que podem estar em descompasso com os temas discutidos nos tweets utilizados para avaliar o sistema.

As variações *8-RH-RS-AA-SVM* e *9-RH-RS-EMB-SVM* utilizam a engenharia de atributos e tem por objetivo melhorar a qualidade do modelo preditivo através da inclusão de atributos adicionais (AA e EMB). As duas variações apresentaram desempenho ligeiramente inferior ao processo proposto para todos os alvos utilizados para os experimentos. Nenhuma destas estratégias para melhorar o poder discriminatório dos atributos apresentou bons resultados.

As variações *10-RH-RS-SL* e *11-RH-RS-RF* consistem na substituição do algoritmo SVM por outros dois algoritmos. Foram realizados testes com outras alternativas de algoritmos, porém esses dois apresentaram os melhores resultados dentre os testados. O algoritmo *Random Forest* apresentou resultados ligeiramente superiores aos do processo proposto para os alvos Hillary (63,76) e Aborto (39,78) e inferior ao resultado obtido pelo processo proposto para o alvo Trump em 2 pontos. Já o algoritmo *Simple Logistic* apresenta resultado bastante inferior para o alvo Aborto (20,03) e um pouco inferior para o alvo Trump (57,69).

Finalmente, a variação *12-RH-RS-VOTE*, inspirada em Silva, Hruschka and Hruschka (2014), é baseada na ideia de que a combinação de algoritmos de classificação através de um comitê pode melhorar o resultado da tarefa de classificação. O resultado obtido para o alvo Trump é quase 1 ponto inferior ao do processo proposto. Para os demais alvos, o resultado é equivalente ou ligeiramente inferior.

Das variações apresentadas nesta seção, a *7-RH-RS-PST-SVM* é a única que transforma o sistema em semi-supervisionado em função da entrada adicional de perfis semente. As demais alternativas avaliadas afetam componentes pontuais da etapa de classificação complementar e tiveram pequeno efeito no desempenho do sistema, seja no sentido de melhorá-lo ou de torná-lo inferior ao do processo proposto.

Como já explorado neste capítulo, a etapa de classificação complementar apresenta de-

sempenho inferior ao da etapa de heurísticas e, como demonstrado acima, os esforços realizados no sentido de melhorar o desempenho desta etapa não surtiram o efeito desejado.

## 5.6 Considerações Finais

Considerando os resultados obtidos, as análises desses resultados, bem como as alternativas testadas, as questões de pesquisa destacadas no Capítulo 1 são respondidas a seguir:

*Q1: Técnicas de mineração podem ser utilizadas como alternativa para realizar a tarefa de detecção de posicionamento em tweets?*

Pode-se concluir que técnicas de mineração de opiniões podem ser utilizadas como ferramenta de apoio à detecção de posicionamento, mas que não são suficientes. Isso se evidencia pelos resultados obtidos para alvos relacionados a figuras públicas (e.g., Hillary e Trump) e pelas análises das etapas do processo proposto baseadas em cobertura (Seção 5.4.1.1) e qualidade (Seção 5.4.1.2) das regras e avaliação da classificação complementar (Seção 5.4.2).

*Q2: As características de um domínio influenciam a forma de expressar posicionamento?*

As características de um domínio podem influenciar a forma de expressar posicionamento, conforme demonstrado através dos experimentos realizados, principalmente com os alvos Ateísmo, Feminismo e Aborto, onde se revela uma fraca representatividade dos alvos estabelecidos (i.e., “Atheism”, “Feminist Movement” e “Legalization of Abortion”) para a detecção de posicionamento através do processo proposto. Mas, através das variações propostas, tais como uso de alvos e perfis semente, pode-se cobrir os diferentes cenários encontrados para cada domínio.

*Q3: Existem características que definem posicionamentos e que possam ser utilizadas para propagar rótulos visando a construção automática de um corpus de treinamento para abordagem supervisionada?*

Características específicas na forma de expressar posicionamento podem ser exploradas para se rotular tweets de forma automática, obtendo-se resultados bastante superiores. Exemplos dessas características são: a) emissão de opinião dirigida a alvos em oposição ao alvo principal do problema (Ateísmo e Aborto), b) uso de hashtags de posicionamento (Hillary e Trump) e c) falta de um alvo claro para o posicionamento (Clima).

A estratégia de rotulação baseada em heurística apresenta resultados promissores para alvos que representem figuras públicas quando aplicada através de regras de sentimento, e superiores a esses quando aplicadas através dos dicionários de hashtags construídos pelo processo.

Para os demais alvos, foram obtidos resultados superiores quando foram utilizados alvos semente e/ou perfis semente, tornando-se uma alternativa para tais domínios.

Já os resultados obtidos através da estratégia de classificação complementar ficam aquém dos da primeira etapa e, apesar de exploradas diversas técnicas a fim de aprimorá-los, esta etapa desponta como um componente a ser estudado em trabalhos futuros.

Os resultados obtidos para alvos relacionados a figuras públicas são bastante promissores quando comparados a sistemas semi-supervisionados, porém ficam aquém de sistemas supervisionados. Isto pode ser verificado através dos resultados obtidos para os alvos Hillary e Trump.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho propôs um processo não supervisionado de detecção de posicionamento em textos de tweets através de uma abordagem híbrida que se divide em 2 grandes etapas: a primeira composta por um conjunto de heurísticas para rotulação automática de tweets e a segunda pela construção e aplicação de um modelo preditivo para classificação complementar. Foram apresentados experimentos com o objetivo de comparar o presente trabalho ao estado-da-arte e avaliar os componentes e estratégias utilizadas neste processo. Finalmente, foram avaliadas estratégias alternativas para o processo proposto no intuito de compará-las a este quando aplicadas aos distintos alvos utilizados nos experimentos.

O processo proposto tem desempenho numericamente superior ao estado-da-arte para o alvo Trump e é comparável a trabalhos supervisionados para o alvo Hillary. Dessa forma, conclui-se que o processo pode ser aplicado com êxito a alvos que representem figuras públicas.

O processo desenvolvido é inovador quando comparado aos sistemas que abordam o problema da detecção de posicionamento por ser não supervisionado, já que não foi encontrado nenhum trabalho com esta característica na literatura revisada. Os trabalhos encontrados utilizam corpora rotulados, ou necessitam de expressões semente provavelmente resultado de inspeção manual de uma amostra do corpus. Com isso, o sistema desenvolvido tem o menor esforço para realizar a tarefa ao ser comparado aos trabalhos estudados.

Ao analisar a abrangência do processo desenvolvido, verificou-se que ele é adequado à detecção de posicionamento para alvos onde uma parte considerável dos tweets avaliados expressam posicionamento através da emissão de opinião com sentimento direcionado ao próprio alvo do problema de posicionamento, e através de hashtags que denotem posicionamento claro. Dessa forma, o trabalho obteve resultados promissores para alvos considerados figuras públicas como Hillary Clinton e Donald Trump.

O trabalho de pesquisa desenvolvido até o momento resultou em quatro publicações:

- Dias, M. e Becker, K. (2016). Inf-ufrgs-opinion-mining at semeval-2016 task 6:Automatic generation of a training corpus for unsupervised identification of stance in tweets. Em **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)**, pgs 378–383, San Diego, California. Association for Computational Linguistics.
- Dias, M. e Becker, K. (2016). Detecção semi-supervisionada de posicionamento em tweets baseada em regras de sentimento. Em **Anais do Simpósio Brasileiro de Banco de Dados 2016**, Salvador, BA., 2016.

- Dias, M. e Becker, K. (2016). Detecção semi-supervisionada de posicionamento em tweets. Em **Anais do SBBD/WTDBD - XV Workshop de Teses e Dissertações em Banco de Dados**. Salvador, BA., 2016.
- Dias, M. e Becker, K. (2016). An Heuristics-based, Weakly-Supervised Approach for Classification of Stance in Tweets. Em **Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence**. Omaha, Nebraska, USA, 2016.

O presente trabalho apresenta as limitações a seguir:

- o desempenho do modelo preditivo é muito inferior ao das heurísticas, resultando em um desempenho final inferior;
- a regra de detecção de neutros apresenta baixa precisão, impactando da detecção de posicionamento;
- devem ser consideradas as características do domínio antes da aplicação do processo, pois, como já discutido, o processo pode ser aplicado com êxito a alvos que representem figuras públicas.

O processo proposto pode evoluir com os objetivos de se tornar mais abrangente e de melhorar o desempenho mesmo para figuras públicas. Para isso, trabalhos futuros podem explorar: a) a pesquisa de alternativas que permitam que o desempenho obtido para figuras públicas possa ser reproduzido para outros domínios, tais como formas de detecção de alvos relacionados indiretamente ao alvo principal, b) o uso de técnicas de aprendizado profundo com o objetivo de melhorar o desempenho da etapa de classificação complementar e c) o uso de técnicas e APIs para detecção de polaridade a fim de melhorar o desempenho das regras baseadas em sentimento.

## REFERÊNCIAS

- ANAND, P. et al. Cats rule and dogs drool!: Classifying stance in online debate. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis**. [S.l.], 2011. p. 1–9.
- AUGENSTEIN, I. et al. Stance detection with bidirectional conditional encoding. **Empirical Methods in Natural Language Processing**, n. 2010, p. 876–885, 2016.
- AUGENSTEIN, I.; VLACHOS, A.; BONTCHEVA, K. Usfd at semeval-2016 task 6: Any-target stance detection on twitter with autoencoders. **Proceedings of SemEval**, p. 389–393, 2016.
- BECKER, K.; MOREIRA, V. P.; SANTOS, A. G. L. dos. Multilingual emotion classification using supervised learning: Comparative experiments. **Inf. Process. Manage.**, v. 53, n. 3, p. 684–704, 2017. Available from Internet: <<https://doi.org/10.1016/j.ipm.2016.12.008>>.
- BØHLER, H. et al. Idi@ ntnu at semeval-2016 task 6: Detecting stance in tweets using shallow features and glove vectors for word representation. **Proceedings of SemEval**, p. 445–450, 2016.
- BOLTUZIC, F. et al. Takelab at semeval-2016 task 6: Stance classification in tweets using a genetic algorithm based ensemble. **Proceedings of SemEval**, p. 464–468, 2016.
- DIAS, M.; BECKER, K. Detecção semi-supervisionada de posicionamento em tweets baseada em regras de sentimento. In: **Anais do Simpósio Brasileiro de Banco de Dados 2016**. Salvador, BA: [s.n.], 2016.
- DIAS, M.; BECKER, K. An heuristics-based, weakly-supervised approach for classification of stance in tweets. In: **Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence**. [S.l.: s.n.], 2016.
- DIAS, M.; BECKER, K. Inf-ufrgs-opinion-mining at semeval-2016 task 6: Automatic generation of a training corpus for unsupervised identification of stance in tweets. In: **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)**. San Diego, California: Association for Computational Linguistics, 2016. p. 390–395.
- EBRAHIMI, J.; DOU, D.; LOWD, D. A joint sentiment-target-stance model for stance classification in tweets. In: **COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan**. [S.l.: s.n.], 2016. p. 2656–2665.
- EBRAHIMI, J.; DOU, D.; LOWD, D. Weakly supervised tweet stance classification by relational bootstrapping. In: **Proceedings of EMNLP**. [S.l.: s.n.], 2016.
- ELFARDY, H.; DIAB, M. Cu-gwu perspective at semeval-2016 task 6: Ideological stance detection in informal text. **Proceedings of SemEval**, p. 434–439, 2016.
- HALL, M. et al. The weka data mining software: an update. **ACM SIGKDD explorations newsletter**, v. 11, n. 1, p. 10–18, 2009.

- HASAN, K. S.; NG, V. Stance classification of ideological debates: Data, models, features, and constraints. In: **IJCNLP**. [S.l.: s.n.], 2013. p. 1348–1356.
- IGARASHI, Y. et al. Tohoku at semeval-2016 task 6: Feature-based model versus convolutional neural network for stance detection. **Proceedings of SemEval**, p. 401–407, 2016.
- KREJZL, P.; STEINBERGER, J. Uwb at semeval-2016 task 6: Stance detection. In: **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)**. San Diego, California: Association for Computational Linguistics, 2016. p. 420–424.
- LAI, M. et al. Friends and enemies of clinton and trump: Using context for detecting stance in political tweets. **arXiv preprint arXiv:1702.08021**, 2017.
- LIU, B. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.
- LIU, C. et al. Iucl at semeval-2016 task 6: An ensemble model for stance detection in twitter. **Proceedings of SemEval**, p. 394–400, 2016.
- MISRA, A. et al. Nlds-ucsc at semeval-2016 task 6: A semi-supervised approach to detecting stance in tweets. **Proceedings of SemEval**, p. 420–427, 2016.
- MOHAMMAD, S. M. Challenges in sentiment analysis. **A Practical Guide to Sentiment Analysis**, D. Das, E. Cambria, and S. Bandyopadhyay, Eds. Springer, 2016.
- MOHAMMAD, S. M. et al. A dataset for detecting stance in tweets. In: **Proceedings of 10th edition of the the Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia**. [S.l.: s.n.], 2016.
- MOHAMMAD, S. M. et al. Semeval-2016 task 6: Detecting stance in tweets. **Proceedings of SemEval**, v. 16, 2016.
- MOHAMMAD, S. M.; KIRITCHENKO, S.; ZHU, X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. **arXiv preprint arXiv:1308.6242**, 2013.
- MOHAMMAD, S. M.; SOBHANI, P.; KIRITCHENKO, S. Stance and sentiment in tweets. **arXiv preprint arXiv:1605.01655**, 2016.
- MUNEZERO, M. D. et al. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. **IEEE Transactions on Affective Computing**, v. 5, n. 2, p. 101–111, April 2014. ISSN 1949-3045.
- MURAKAMI, A.; RAYMOND, R. Support or oppose?: classifying positions in online debates from reply activities and opinion expressions. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 23rd International Conference on Computational Linguistics: Posters**. [S.l.], 2010. p. 869–875.
- PATRA, B. G.; DAS, D.; BANDYOPADHYAY, S. Ju nlp at semeval-2016 task 6: Detecting stance in tweets using support vector machines. **Proceedings of SemEval**, p. 440–444, 2016.
- PENNINGTON, J.; SOCHER, R.; MANNING, C. D. Glove: Global vectors for word representation. In: **EMNLP**. [S.l.: s.n.], 2014. v. 14, p. 1532–1543.

RAJADESINGAN, A.; LIU, H. Identifying users with opposing opinions in twitter debates. In: SPRINGER. **International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction**. [S.l.], 2014. p. 153–160.

SILVA, N. F. D.; HRUSCHKA, E. R.; HRUSCHKA, E. R. Tweet sentiment analysis with classifier ensembles. **Decision Support Systems**, Elsevier, v. 66, p. 170–179, 2014.

SOBHANI, P.; MOHAMMAD, S. M.; KIRITCHENKO, S. Detecting stance in tweets and analyzing its interaction with sentiment. **c2016 The\* SEM 2016 Organizing Committee. All papers c2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0>**, p. 159, 2016.

SOMASUNDARAN, S.; WIEBE, J. Recognizing stances in online debates. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1**. [S.l.], 2009. p. 226–234.

SRIDHAR, D.; GETOOR, L.; WALKER, M. Collective stance classification of posts in online debate forums. **ACL 2014**, v. 109, 2014.

THOMAS, M.; PANG, B.; LEE, L. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 2006 conference on empirical methods in natural language processing**. [S.l.], 2006. p. 327–335.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, v. 24, n. 3, p. 478–514, 2012.

VIJAYARAGHAVAN, P. et al. Deepstance at semeval-2016 task 6: Detecting stance in tweets using character and word-level cnns. **arXiv preprint arXiv:1606.05694**, 2016.

WEI, W. et al. pkudblab at semeval-2016 task 6 : A specific convolutional neural network system for effective stance detection. In: **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)**. San Diego, California: Association for Computational Linguistics, 2016. p. 396–400.

ZARRELLA, G.; MARSH, A. Mitre at semeval-2016 task 6: Transfer learning for stance detection. In: **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)**. San Diego, California: Association for Computational Linguistics, 2016. p. 470–475.

ZHANG, Z.; LAN, M. Ecnu at semeval-2016 task 6: Relevant or not? supportive or not? a two-step learning system for automatic detecting stance in tweets. **Proceedings of SemEval**, p. 451–457, 2016.

# Appendices

## ApêndiceA

Neste anexo, são apresentadas a abordagem proposta em versão anterior do presente trabalho, junto com as principais diferenças em relação à abordagem proposta nesta dissertação.

### A.1 Visão Geral da Versão Anterior deste Trabalho

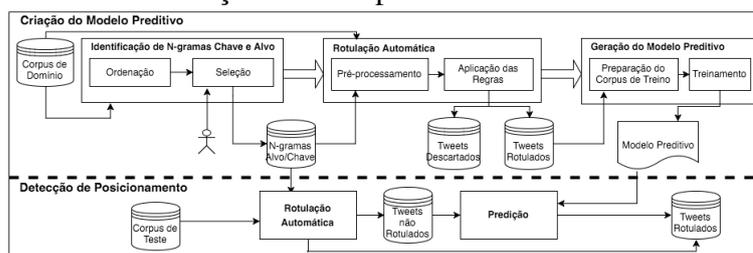
A Figura A.1 destaca os principais pontos da abordagem semi-supervisionada para detecção de posicionamento baseado no conteúdo de tweets proposta em trabalhos anteriores (DIAS; BECKER, 2016b; DIAS; BECKER, 2016a; DIAS; BECKER, 2016c). A abordagem é semi-supervisionada porque pressupõe como entrada um pequeno conjunto de n-gramas que representem os alvos de opinião direta ou indiretamente relacionados, ou expressões explícitas de posicionamento (e.g. hashtags ou tópicos tipicamente mencionados por um dos lados no domínio). Ela compreende dois aspectos: a) criação de um modelo preditivo a partir da criação automática de um corpus de treinamento usando heurísticas, e b) a detecção de posicionamento. No restante da seção, são discutidas as principais atividades necessárias para identificação de posicionamento em conteúdos de tweets.

### A.2 Criação de um Modelo Preditivo

#### A.2.1 Identificação de N-gramas Alvo e Chave

O processo inicia com a identificação de um conjunto de n-gramas (uni-gramas, bi-gramas e tri-gramas) que sejam representativos para expressar posicionamentos no domínio, os quais dividimos em dois grandes grupos: *alvo* e *chave*.

Figura A.1: Processo de Detecção Semi-supervisionada de Posicionamento em Tweets



fonte: (DIAS; BECKER, 2016a)

Os n-gramas alvo são aqueles utilizados em tweets para representar direta ou indiretamente alvos a favor da questão, ou o lado opositor. Por exemplo, para o problema de posicionamento em relação a políticos (e.g. Hillary Clinton), alvos do lado favorável são variações do nome do político (e.g. Hillary, Clinton), seus aliados (e.g. Bill, Obama), ou seu partido (e.g. democratas), enquanto para o lado contrário poderíamos citar seus concorrentes (e.g. republicanos, Trump, Bush). Além de entidades, há também assuntos caracterizando alvos, tais como elementos do programa de governo. Por exemplo, opiniões sobre “imigrantes” ou “mexicanos” caracterizam posicionamento sobre Trump, enquanto que “saúde pública” é utilizado para se posicionar em relação à Hillary.

Já os n-gramas chave são normalmente hashtags ou expressões que, quando mencionadas, denotam uma forte tendência quanto ao posicionamento do tweet. Exemplos de n-gramas de posicionamento favorável/contrário à Hillary são “readyforhillary” e “stophillary2016”, respectivamente.

N-gramas alvo e chave devem ser separados em 2 subgrupos distintos: *Favoráveis* e *Contrários*, considerando o alvo principal. Por exemplo, ”hillary“ é um n-grama alvo favorável ao alvo principal Hillary, enquanto que ”trump“ é um n-grama alvo contrário.

Propõe-se que os n-gramas sejam extraídos de um conjunto de tweets no domínio (identificado na Figura A.1 como *Corpus de Dominio*), em um processo semi-automático. Um procedimento extrai deste corpus os n-gramas mais frequentes, para um limiar dado, e os ordena por frequência. Esta lista é apresentada a um anotador para inspeção manual, de forma que possa analisá-los, selecionando os mais relevantes.

## A.2.2 Rotulação Automática de Tweets

A rotulação automática de tweets permite compor um corpus de treinamento para aprendizagem supervisionada, evitando o custo associado a esta atividade quando executada de forma manual. Para este fim, propusemos um conjunto de regras, apresentadas na Tabela A.1, as quais representam nossa premissa básica de que posicionamentos podem ser expressos na forma de sentimentos positivos ou negativos em relação a alvos direta ou indiretamente relacionados. As regras 1 e 2 inferem o posicionamento do tweet com base na presença de n-gramas chave, enquanto que as regras de 3 a 6 rotulam o tweet com base na presença de n-gramas de alvo combinado com a polaridade detectada para o texto do Tweet. A regra 7 assume que não há posicionamento para tweets onde não foi detectado sentimento. Dado o tamanho de um tweet, assume-se que o alvo de um sentimento é um n-grama alvo.

Regra	Descrição	Posicionamento
1 - CHAVE-FAVOR	Presença de n-grama(s) chave favorável(eis) e ausência de n-grama chave contrário	FAVORÁVEL
2 - CHAVE-CONTRA	Presença de n-grama(s) chave contrário(s) e ausência de n-grama chave favorável	CONTRÁRIO
3 - ALVO-FAVOR-POSITIVO	Presença de n-grama(s) alvo favorável(eis), ausência de n-grama contrário e polaridade positiva	FAVORÁVEL
4 - ALVO-FAVOR-NEGATIVO	Presença de n-grama(s) alvo favorável(eis), ausência de n-grama contrário e polaridade negativa	CONTRÁRIO
5 - ALVO-CONTRA-POSITIVO	Presença de n-grama(s) alvo contrário(s), ausência de n-grama favorável e polaridade positiva	CONTRÁRIO
6 - ALVO-CONTRA-NEGATIVO	Presença de n-grama(s) alvo contrário(s), ausência de n-grama favorável e polaridade negativa	FAVORÁVEL
7 - NEUTRO	Tweet de polaridade neutra	SEM POSIC.
Outros casos		DESCARTAR

Tabela A.1: Regras para Rotulação Automática

Esta etapa é dividida em dois passos: *pré-processamento*, que gera as informações necessárias para aplicação das regras, e a *aplicação das regras* propriamente dita. Como resultado, tem-se dois conjuntos de tweets: um rotulado com o posicionamento assumido (FAVORÁVEL, CONTRÁRIO ou NENHUM), e um de tweets descartados, i.e. que não foram filtrados por nenhuma regra. Dado o conjunto de n-gramas chave e alvo, o passo de pré-processamento tem por objetivo gerar as seguintes informações para cada tweet do Corpus de Domínio: a) Presença de pelo menos um n-grama chave favorável ao alvo principal; b) Presença de pelo menos um n-grama chave contrário ao alvo principal; c) Presença de pelo menos um n-grama alvo favorável ao alvo principal; d) Presença de pelo menos um n-grama alvo contrário ao alvo principal e e) Polaridade do Tweet.

### A.2.3 Criação de um Modelo Preditivo de Posicionamento Utilizando Aprendizado Supervisionado

Considerando o texto dos tweets do Corpus do Domínio, e os respectivos rótulos atribuídos automaticamente na etapa anterior, o objetivo desta etapa é compor um corpus de treinamento, e usá-lo para treinar um classificador visando a geração de um modelo preditivo. Como já salientado, apenas os tweets que foram filtrados pelas regras 1 a 7 da Tabela A.1 são considerados para o corpus de treinamento. Em nossos experimentos, esta etapa foi realizada utilizando um algoritmo de classificação (SVM).

Em termos de processamento textual dos tweets, a criação do corpus de treinamento implica as seguintes ações: a) substituição de menções a perfis de twitter pelo uni-grama "a\_mention", desde que não estejam relacionados aos n-gramas alvo identificados na primeira etapa do processo (e.g. "realdonaldtrump"), b) remoção de stopwords, c) extração de todos uni-gramas, e d) criação do corpus usando peso binário para os atributos (i.e. presença ou ausência do uni-grama no texto).

### **A.3 Identificação de Posicionamento em Tweets não Rotulados**

Nossa abordagem identifica o posicionamento contido no texto de tweets combinando a rotulação automática, no exato processo descrito na Seção A.2.2, com o modelo preditivo gerado através do processo descrito na Seção A.2.3. Dado um corpus contendo um conjunto de tweets não rotulados, representado na Figura A.1 como *Corpus de Teste*, primeiro tenta-se rotulá-los com o apoio das regras da Tabela A.1. Somente os tweets que não são filtrados por nenhuma das regras (*Tweets Não Rotulados*) são submetidos ao modelo preditivo.