UM MODELO DE CLASSIFICAÇÃO DE POLARIDADE EM CINCO NÍVEIS

FELIPE ALVES FERREIRA

UM MODELO DE CLASSIFICAÇÃO DE POLARIDADE EM CINCO NÍVEIS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal do Amazonas como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientadores: Dr. David Braga Fernandes de Oliveira e
Dr. Marco Antonio Pinheiro de Cristo

Manaus

Fevereiro de 2017

Alves Ferreira, Felipe

F383u Um modelo de classificação de polaridade em cinco níveis / Felipe Alves Ferreira. — Manaus, 2017
xx, 48 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal do Amazonas
Orientador: Dr. David Braga Fernandes de Oliveira

1. Análise de sentimento. 2. Polaridade. 3. Opinião. 4. Classificação. I. Título.

FOLHA DE APROVAÇÃO

"Um modelo de Classificação de Polaridade em Cinco Níveis"

FELIPE ALVES FERREIRA

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof. Marco Antonio Pinheiro de Cristo - PRESIDENTE

Prof. João Marcos Bastos Cavalcanti - MEMBRO EXTERNO

Prof. Moisés Gomes de Carvalho - MEMBRO EXTERNO

Prof. David Braga Fernandes de Oliveira - MEMBRO EXTERNO

Manaus, 22 de Fevereiro de 2017

A todos que me acompanharam nessa trajetória.

Agradecimentos

Inicialmente, agradeço a Deus pelo dom da vida, pela família em que me permitiu nascer, e por ter me contemplado com a capacidade de me maravilhar com a natureza e suas criaturas. Sem Ele nenhum agradecimento faria sentido.

A minha mãe, Maria do Socorro Alves Ferreira por sempre cuidar de mim e por sempre me incentivar a estudar, torcendo pelas minhas vitórias.

Agradeço também a minha esposa, Mariana Ferreira por sempre estar ao meu lado em todos os momentos.

Ao meu filho Rafael Ferreira por ser meu grande parceiro nos momentos de descontração, transformando os momentos difíceis em alegria.

Aos meus orientadores: Dr. David Fernandes e Dr. Marco Cristo por acreditarem em mim, pelos direcionamentos e por serem profissionais exemplares.

Ao INDT e a Globo.com pelo apoio e auxílio financeiro.

A minha família, a qual amo muito, pelo carinho, paciência e incentivo.

Resumo

Análise de sentimento é a área de estudo que observa as opiniões das pessoas, sentimentos, avaliações, atitudes e emoções em torno de entidades como produtos, serviços, organizações e eventos. No mundo real, empresas e organizações frequentemente estão interessadas em saber opiniões públicas sobre seus produtos e serviços. Os consumidores também estão interessados em saber a opinião de quem adquiriu um produto antes de comprá-lo. Outras pessoas estão interessadas em saber a opinião dos outros sobre determinados candidatos de um pleito eleitoral antes de tomar uma decisão sobre quem irá votar. O objetivo deste trabalho é desenvolver um método de aprendizagem supervisionada capaz de classificar tweets em cinco níveis de polaridade (aprovação completa, opinião positiva pontual, opinião neutra, opinião negativa pontual e rejeição completa) usando para isso tweets do contexto político como estudo de caso. Para tal, investigamos se técnicas que fazem detecção de polaridade de três níveis são capazes de oferecer boas evidências para o treino e testes de classificadores no contexto dos cinco níveis de polaridade nos tweets. Partindo dessa ideia, propomos estratégias de aprendizagem com os classificadores Árvores de Decisão, Naive Bayes e SVM usando como *features* o modelo de *bag-of-words*, as evidências a partir de resultados de métodos que classificam polaridade em três níveis e um modelo de extração de meta-informação. Os resultados mostraram que existe um ganho nas acurácias dos classificadores ao combinar os diferentes modelos de *features*.

Palavras-chave: Análise de Sentimento, Classificação, Polaridade, Opinião, Naive Bayes, SVM, Árvores de Decisão.

Abstract

Sentiment analysis is the area of study that observes people's opinions, feelings, assessments, attitudes and emotions around entities such as products, services, organizations and events. In the real world, companies and organizations are often interested in knowing public opinions about their products and services. Consumers are also interested in knowing the opinion of those who bought a product before buying it. Other people are interested in knowing the opinions of others about certain candidates in an election process before making a decision about who will vote. The objective of this work is to develop a supervised learning method capable of classifying tweets in five levels of polarity (complete approval, punctual positive opinion, neutral opinion, punctual negative opinion and complete rejection) using tweets of the political context as a case study. In order to do this, we investigated whether three-level polarity detection techniques are capable of providing good evidence for training and classifier testing in the context of the five polarity levels in tweets. Based on this idea, we propose learning strategies with the Decision Trees, Naive Bayes and SVM classifiers using as features models: bag-of-words, evidence from results of methods that classify polarity in three levels and a meta-level extraction method. The results showed that there is a gain in the accuracy of the classifiers when combining the different models of features.

Keywords: Sentiment Analysis, Classification, Polarity, Opinion, Naive Bayes, SVM, Decision Trees.

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1

Introdução

Análise de sentimento é a área de estudo que observa as opiniões das pessoas, sentimentos, avaliações, atitudes e emoções em torno de entidades como produtos, serviços, organizações e eventos. Opiniões são o centro de quase todas as atividades humanas. Sempre que precisamos tomar uma decisão, queremos consultar as opiniões de outros. Na atualidade, empresas e organizações frequentemente estão interessadas em saber opiniões públicas sobre seus produtos e serviços. Os consumidores também estão interessados em saber a opinião de quem adquiriu um produto antes de comprá-lo. Outras pessoas estão interessadas em saber a opinião dos outros sobre determinados candidatos de um pleito eleitoral antes de tomar uma decisão sobre quem irá votar.

A importância da área de análise de sentimentos cresceu bastante com o surgimento das mídias sociais como fóruns, blogs e microblogs, pois tais sistemas que permitem a interação social a partir do compartilhamento de informações sobre os mais variados assuntos, também são uma vasta fonte de opiniões e emoções das pessoas sobre os mais diversos temas de interesse. Essas fontes tem sido objetos de diversos estudos que buscam, por exemplo, compreender melhor a relação entre o estado emocional e as palavras que as pessoas usam nas mídias socias[Beasley & Mason, 2015], detectar sentimentos sobre eventos cotidianos como festivais ou lançamento de filmes[Schinas et al., 2013] e até encontrar termos extremos que ajudam a identificar automaticamente o sentimento sobre temas conflitantes de uma sociedade[Makrehchi, 2014].

Em geral, muitos trabalhos da literatura sobre análise de sentimento visam investigar as opiniões expressas em determinadas mensagens com dois níveis (positiva/negativa) ou três níveis de polaridade (positiva, neutra e negativa). Na literatura encontramos diversos trabalhos com o intuito de identificar a polaridade (ou opinião) de uma mensagem de texto em relação ao seu objeto principal. Por exemplo [Hu & Liu, 2004] propõe um método para classificar a polaridade sobre as características dos

produtos de sites de vendas na internet. Ele cria um léxico (dicionário de sentimento) usando *WordNet* [Miller et al., 1990; Fellbaum, 1998] a partir dos termos extraídos dos comentários para então fazer a predição de polaridade em três níveis. O método avalia somente as frases onde os consumidores expressam suas opiniões sobre as características dos produtos e então sumariza o resultado para calcular o escore de sentimento de cada característica.

O trabalho realizado por [Pla & Hurtado, 2014] propõe um modelo para identificar a tendência política dos usuários no Twitter. A partir dos tweets dos usuários, ele extraiu e classificou as entidades relacionadas ao contexto político como partidos e candidatos, em três níveis de polaridade. Usando tais polaridades, o autor estabeleceu uma métrica para inferir se um usuário tem preferência por partidos de esquerda, de direita ou centrais.

Baseado em uma busca de tópicos, [Mukherjee et al., 2012] desenvolveu um sistema para classificar os tweets relacionados ao tópico de consulta em três níveis de polaridade. Ele recupera e categoriza os tweets pertencentes a uma entidade pesquisada com base no seu conteúdo de sentimento para atribuir uma pontuação de sentimento global como sendo positivo, negativo ou neutro.

No entanto, em dados contextos é necessário classificar as sentenças em mais de três classes de opiniões. Por exemplo, no contexto envolvendo qualidade de serviços de telefonia móvel, um tweet pode expressar uma rejeição completa (ex: *Eu odeio aquela operadora e jamais seria um cliente dela*); pode expressar uma opinião negativa mas pontual (ex: *Entrei em contato com a operadora e me deparei com atendentes despreparadas e que não sabiam me dar nenhuma informação correta*); pode expressar uma opinião neutra(ex: *Não conheço os serviços daquela operadora pois nunca fui cliente dela*); pode expressar uma opinião positiva mas pontual (ex: *Enfim a operadora me deu um desconto por ser um bom cliente*); e pode expressar uma aprovação completa (ex: *Sempre gostei dessa operadora, ela sempre me tratou muito bem*)

Em geral, técnicas que avaliam polaridades com mais de três níveis permitem uma compreensão mais sofisticada do sentimento do que técnicas que avaliam apenas três níveis pois é possível ajustar o valor do sentimento de um conceito relativo às modificações que podem cercá-lo. As técnicas de três níveis apenas oferecem uma visão macro sobre a polaridade de um texto.

O objetivo deste trabalho é desenvolver um método de aprendizagem supervisionada capaz de classificar os tweets em cinco níveis de polaridade usando para isso tweets do contexto político como estudo de caso. Queremos separar uma opinião *negativa pontual* de uma *rejeição completa*, separar uma opinião *positiva pontual* de uma *aprovação completa* e de separar opinião *neutra*.

Por exemplo, no contexto político, podemos separar uma *rejeição completa* (ex: eu odeio a Dilma) de uma opinião *negativa pontual* (ex: a Dilma não devia ter colocado o Eduardo Braga no ministério de minas e energia); separar uma opinião *positiva pontual* (ex: a Dilma conseguiu equilibrar as contas do governo) de uma *aprovação completa* (ex: a Dilma é a melhor presidente de todos os tempos) e separar uma opinião *neutra* (ex: não sei o que pensar da Dilma).

Tais separações são importantes pois permitem estudos mais aprofundados onde uma classificação de três níveis não é muito útil, pois, se quiséssemos estudar a rejeição de um candidato, produto ou serviço, uma opinião negativa pontual não é um indício tão forte de rejeição.

As cinco classes de polaridade propostas se diferem de outros trabalhos da literatura onde também levam em consideração polaridade de cinco níveis. Por exemplo, submetendo o tweet de *Aprovação completa* "A Dilma é a melhor presidente de todos os tempos"para a técnica proposta por [Thelwall, 2013] que faz uso de uma escala dual numérica de cinco níveis ((1 à 5) para positivo e (-1 à -5) para negativo) para medir a intensidade da polaridade de um texto, o resultado obtido é uma intensidade positiva de valor dois (pouco positiva) e uma intensidade negativa de valor menos um (pouco negativa). De fato, as cinco classes apresentadas por este trabalho, exigem uma aprendizagem sobre o conceito das mesmas pois as polaridades *Aprovação completa*, *Positiva pontual*, *Neutra*, *Negativa pontual* e *Rejeição completa* podem representar diferentes polaridades em outras técnicas da literatura.

Em nosso trabalho, queremos investigar se técnicas que fazem detecção de polaridade de três níveis são capazes de oferecer boas evidências para o treino e testes de classificadores no contexto dos cinco níveis de polaridade nos tweets. Partindo dessa ideia, propomos estratégias de aprendizagem com os classificadores Decision Trees, Naive Bayes e SVM usando dois modelos de representação de *features*.

Como uma técnica para a extração dos diferentes modelos de *features*, propomos o uso dos resultados da ferramenta online iFeel [Gonçalves et al., 2013] que usa 19 técnicas diferentes onde cada uma delas classifica a sentença em três níveis de polaridade.

Uma outra técnica usada neste trabalho é a técnica de extração de meta-informações proposta por [Canuto et al., 2016]. Essa técnica fornece cinco tipos de meta-características a partir dos vizinhos mais próximos de cada instância usando uma busca com kNN.

Para avaliar a eficácia de tais abordagens nesse novo contexto, comparamos a acurácia dessas abordagens com um método Naive, no caso o modelo de *bag-of-words*. Como resultado de nossa avaliação, observamos que os melhores resultados de classificação para cada algoritmo foi obtido usando as combinações de *bag-of-words* e meta-

características e os piores resultados foram obtidos usando as *features* a partir da iFeel.

1.1 Objetivos

1.1.1 Geral

Através de uma abordagem de aprendizado supervisionado, desenvolver um método capaz de classificar opiniões expressas em mensagens do Twitter em cinco níveis de polaridade. As polaridades a serem identificadas serão: *Aprovação completa, positivo pontual, neutro, negativo pontual e rejeição completa.*

1.1.2 Específicos

Este trabalho possui os seguintes objetivos específicos:

1. Criar uma coleção de tweets sobre candidatos do contexto político para a análise de polaridade de cinco níveis. Para esta coleção, selecionar uma amostra de forma aleatória para então rotulá-la de acordo com os níveis de polaridade através de avaliadores humanos;

2. Implementar diferentes métodos de extração de *features* que serão usados como base de comparação;

3. Verificar como diferentes estratégias de *features* podem ser utilizadas no aprendizado de modelos de classificação para fins de comparação e escolha do modelo mais apropriado ao contexto das cinco classes de polaridade.

1.2 Organização da dissertação

Além deste capítulo introdutório, esta dissertação está dividida da seguinte maneira: O Capítulo 2 apresenta uma revisão da literatura sobre os diferentes métodos de análise de sentimentos; o Capítulo 3 descreve a nossa proposta para a classificação de polaridade de cinco níveis. Os experimentos e os resultados obtidos são mostrados no Capítulo 4 e por fim, no Capítulo 5 descrevemos nossas conclusões.

Capítulo 2

Trabalhos Relacionados

Com o crescimento das redes sociais na web, análise de sentimento e mineração de opiniões tornaram-se objetos de estudo para muitas pesquisas. Neste capítulo, examinamos as diferentes técnicas utilizadas para medir os sentimentos de textos online. Os trabalhos estão organizados em técnicas que avaliam duas polaridades, três polaridades e mais de três polaridades.

2.1 Técnicas que avaliam duas polaridades

Com o propósito de avaliar comentários sobre produtos em sites de e-commerce, [Hu & Liu, 2004] constrói um léxico para prever a polaridade sobre cada característica do produto que foi comentada pelos clientes. Fazendo uso de técnicas de mineração de dados e processamento de linguagem natural (NLP), o método faz a extração das características que foram comentadas pelos clientes, identifica as sentenças de opinião em cada review e decide se cada sentença de opinião é positiva ou negativa. Para definir a polaridade de cada sentença, primeiro um conjunto de palavras adjetivas normalmente usadas para expressar opiniões é identificado usando NLP e em seguida é determinada a orientação semântica de cada palavra (positiva ou negativa) através de uma técnica de bootstrapping utilizando WordNet [Miller et al., 1990; Fellbaum, 1998]. Após a identificação semântica de cada palavra de opinião, a classificação de uma sentença de opinião em positiva ou negativa é definida pela orientação dominante dos termos, ou seja, se em uma sentença houver mais palavras de opiniões positivas, então a sentença de opinião será positiva e vice-versa.

Utilizando uma abordagem para a classificação em dois níves de polaridade (positiva, negativa) de textos em geral com base na sua estrutura discursiva, [Heerschop et al., 2011] desenvolveu uma ferramenta chamada `Pathos`. Essa ferramenta primeiro

identifica partes do discurso (*POS tagging*) e os lemas de todas as palavras para então fazer uma desambiguação. Após essa etapa, é empregado um analisador de discurso para transformar o texto de entrada em unidades elementares do discurso [Mann & Thompson, 1988] que são usadas para atribuir um peso a cada palavra individual. O sentimento geral de um texto é então calculado como uma média ponderada de pontuações de palavras individuais obtidas de um léxico de sentimentos chamado SentiWordNet [Baccianella et al., 2010]. O SentiWordNet é um dicionário criado com base em um dicionário léxico inglês chamado WordNet [Miller, 1995] e é bastante utilizado em *opinion mining* para fazer um agrupamento de substantivos, adjetivos e outras classes gramaticais em conjuntos de sinônimos chamados synsets. SentiWordNet associa três pontuações (positivo, negativo e neutro) com o synset do dicionário WordNet para indicar o sentimento do texto.

Usando redução de dimensionalidade para inferir a polaridade de conceitos do senso comum, o SenticNet [Cambria et al., 2010] é uma ferramenta pública para minerar opiniões de textos em linguagem natural através de uma abordagem semântica ao invés de explorar apenas evidências sintáticas. O autor cria uma coleção de conceitos do senso comum com uma polaridade positiva (1) ou negativa (-1). Com esse objetivo, diferentemente de SentiWordNet, o autor descarta conceitos com polaridade neutra ou quase neutra. Além disso, enquanto SentiWordNet armazena três valores para cada synset, em SenticNet cada conceito é associado somente a um valor representando sua polaridade a fim de evitar redundâncias. Portanto, conceitos no SenticNet como "causar boa impressão", "parecer atraente", "mostrar apreço"ou "bom negócio"provavelmente terão polaridade próximas de 1, enquanto que conceitos como "ser demitido", "deixar para trás"ou "perder o controle"terão polaridades próximas de -1.

Com o intuito de avaliar o efeito de fatores externos como fenômenos naturais no sentimento das mensagens de uma rede social como o Twitter, usando técnicas de aprendizado de máquina em um corpus do Twitter correlacionado com a condição atmosférica no tempo e localização dos tweets, [Hannak et al., 2012] encontra o sentimento agregado que segue situações climáticas distintas, temporalidade e padrões sazonais. Primeiro ele constrói um conjunto de tweets positivos e negativos considerando tweets com somente um emoticon positivo ou negativo. Em seguida ele tokeniza os tweets ignorando hashtags, usernames, e URL. Ele calcula a fração relativa de vezes que cada token ocorre com um emoticon positivo ou negativo e usa isso como escore. De forma semelhante ele calcula o escore de sentimento de um tweet observando ocorrências dos tokens listados e obtendo a média ponderada no escore de sentimento individual do token para ser o escore de sentimento do tweet. Com a lista de palavras resultante,

o autor examina os padrões de sentimentos que existem. Para tal fim, o autor optou por árvores de decisão pois ela pode lidar com todos os tipos de atributos e valores ausentes, fornece uma explicação sobre o como a árvore fez uma certa predição para um determinado input além de ser considerada um dos melhores modelos tanto para classificação quanto para problemas de regressão e também facilita o treinamento e consulta em paralelo [Caruana & Niculescu-Mizil, 2006]. O autor agrupou os tweets em buckets de hora em hora por localidade, obtendo o escore do sentimento do bucket através da média dos sentimentos de todos os tweets. Para simplificar a criação das árvores,o escore do sentimento obtido através da média foi reduzido para um valor binário expressando (1) para positivo e (0) para negativo. Usando dados históricos coletados sobre o clima e correlacionando-os com os tweets, o autor modelou o problema para que dado as variáveis de entrada: localidade, mês, tempo(dia do mês, dia da semana e hora do dia) e a temperatura, para que o preditor seja capaz indicar o sentimento agregado.

O [De Smedt & Daelemans, 2012] descreve um pacote de programação para Python 2.4+ chamado "Pattern". Ele foi desenvolvido para lidar com NLP, web mining e análise de sentimentos. A análise de sentimento positivo/negativo é oferecida através da média de pontuações de adjetivos na sentença de acordo com um pacote de léxico de adjetivos.

[Mohammad et al., 2013] propõe a criação de um dicionário léxico de uma forma semelhante ao [Mohammad, 2012] mas os autores usaram a ocorrência de emoticons para classificar o tweet como positivo ou negativo. Em seguida, o escore do n-grama foi calculado com base na frequência que ele ocorre em cada classe de tweets.

2.2 Técnicas que avaliam três polaridades

Um dos métodos mais simples para inferir a polaridade de uma mensagem de redes sociais como o Twitter é a utilização de emoticons. Usando um conjunto de emoticons positivos, negativos e neutros, [Gonçalves et al., 2013] determina se a classe de polaridade é positiva, negativa ou neutra através da identificação de tais emoticons. Caso apareça mais de um, a classe de polaridade é determinada pelo primeiro emoticon que aparece na mensagem. No entanto, a taxa de mensagens de redes sociais que contém ao menos um emoticon é muito baixa comparado ao número total de mensagens que poderiam expressar alguma classe de emoção, de maneira que a técnica apenas classifica mensagens com emoticons, ou seja, o número de mensagens classificadas é consequentemente baixo. Estudos recentes mostram que essa taxa é menor que 10%

[Park et al., 2013].

Um sistema em tempo real para a análise de sentimentos do público sobre os candidatos a presidência dos Estados Unidos em 2012 foi a proposta de Wang et al. [2012]. Ele tira proveito de postagens que contém opiniões (positivas, negativas e neutras) e pontos de vista sobre os partidos e candidatos buscando explorar se o Twitter fornece evidências sobre o desenrolar das campanhas e as indicações de mudanças de opinião pública. Durante um dos principais eventos políticos da campanha, o primeiro debate em Iowa (15/12/2011) foram coletados mais de meio milhão de tweets em poucas horas.

Para lidar com um grande volume de tweets a fim de processa-los em tempo real, o autor implementou uma ferramenta capaz de capturar expressões de reações instantâneas sobre eventos no momento em que eles ocorrem, mostrando por exemplo, como uma platéia reage a declarações de um candidato durante um debate político. Além da complexidade da tarefa e da necessidade de lidar com um fluxo muito rápido de tweets, outros dois aspectos importantes foram tratados. O primeiro deles é que por conta da informalidade e da limitação de caracteres, a linguagem usada nas mensagens fogem de qualquer estrutura gramatical o que torna a tarefa de análise sentimental mais difícil. O segundo aspecto é que os tweets em geral e os tweets sobre política tendem a ser sarcásticos em sua maioria conforme estudo de González-Ibáñez et al. [2011]. Para a coleta dos tweets relevantes, foram usadas combinações de palavras-chave sobre os nomes dos candidatos e eventos. Essas combinações foram submetidas a ferramenta comercial de provisão de dados do Twitter chamada Gnip. Baseando-se em técnicas de processamento de linguagem natural (NLP), o autor usou um tokenizador disponibilizado por Krieger & Ahn [2010] para tratar URLs, emoticons comuns, números de telefones, tags html, menções do twitter e hashtags, números com frações e decimais, repetição de símbolos e caracteres unicode.

O modelo de análise de sentimentos usado no sistema foi baseado na premissa que opiniões expressadas seria altamente subjetiva e contextualizada. Por conta disso foi usado uma abordagem crowdsourcing através da Amazon Mechanical Turk(AMT) para rotular os sentimentos em tweets do domínio político. A cada Turker foi perguntado sua idade, gênero e sua preferência política e então foram convidados a rotular um conjunto de tweets em três sentimentos: positivo, negativo, neutro ou incerto, se o tweet foi sarcástico ou humorístico, o sentimento em uma escala de positivo a negativo, e a preferência política do autor do tweet em uma escala deslizante de conservador a liberal. O modelo de sentimento é baseado no rótulo do sentimento e nos rótulos de sarcasmo e humor.

O classificador estatístico escolhido foi o Naive Bayes para a modelagem de ca-

racterísticas de unigramas a partir da tokenização dos tweets preservando pontuações que podem significar sentimento como emoticons e exclamação. Para a exibição gráfica sobre os sentimentos, é feito uma agregação dos tweets e os sentimentos em períodos de tempo para cada candidato. A cada um minuto, é mostrado em tempo real o número de tweets de cada candidato e a cada cinco minutos, é exibido o número de tweets positivos, negativos, neutros. Também é mostrado um trending de palavras dos últimos cinco minutos. Esse trending é calculado usando o TF-IDF onde tweets sobre todos os candidatos em um minuto são tratados como um único documento, as palavras que são tendências são tokens do minuto atual com os maiores pesos de TF-IDF quando o corpus das duas últimas horas.

O trabalho de [Gonçalves et al., 2013] tem como propósito detectar oscilações de humor de usuários no Twitter chamado PANAS-t. O método consiste em uma versão adaptada de (PANAS) Positive Affect Negative Affect [Watson et al., 1988], método bem conhecido na psicologia com um grade conjunto de palavras, cada uma delas associada a um dos onze estados de espírito (surpresa, medo, jovialidade, segurança, serenidade, tristeza, culpa, hostilidade, timidez, fadiga e atenção). O PANAS original consistiu em duas escalas de 10 itens de humor para proporcionar um resumo de medidas de PA (Positive Affect) e NA (Negative Affect). Em suma, PA reflete a extensão em que uma pessoa se sente com entusiasmo, ativa, e alerta.

Um PA alto é um estado alto de energia, concentração total, e um envolvimento prazeroso, enquanto que a baixa PA é caracterizada por tristeza e letargia. Em contraste, NA é uma dimensão geral de angústia subjetiva e engajamento desagradável que integra uma variedade de estados de humor aversivos, incluindo raiva, desprezo, nojo, culpa, medo e nervosismo. Mais tarde, o mesmo autor criou uma versão expandida chamada PANAS-x, que não só mede as duas escalas de ordem superior originais (PA e NA) mas também os 11 sentimentos: medo, tristeza, culpa, hostilidade, timidez, fadiga, surpresa, jovialidade, auto-confiança, atenção e serenidade. [Gonçalves et al., 2013] propõe o PANAS-t, uma versão de 11 sentimentos da escala psicométrica adaptado ao contexto do Twitter.

Para verificar se tweets que expressam um tipo específico de sentimento aumentou ou diminui para um dado evento (ex: morte de uma celebridade ou desastre natural), o autor primeiro investiga quais os tipos de sentimentos aparecem nos períodos em que nenhum evento significativo acontece. Para isso, ele agregou os sentimentos durante um longo período de tempo e considerou a proporção de cada tipo de sentimento como baseline. Então, pela comparação da proporção dos tweets que contém um sentimento específico durante um dado evento contra todo o baseline foi possível saber como os sentimentos mudam em relação a presença de um dado evento no dataset. Para estes

processos, o autor assumiu que cada tweet normalizado pode ser mapeado para um único sentimento.

Quando o tweet teve qualquer adjetivo correspondente a um dos 11 sentimentos, ele foi associado ao sentimento correspondente como o sentimento principal do tweet. No caso em que nenhuma palavra de sentimento apareceu no tweet, o método não pode inferir o sentimento para o dado tweet. Esta é uma limitação da maioria das técnicas semelhantes. O valor de baseline para cada sentimento é definido como a proporção que divide o número de ocorrências de tweets de cada tipo de sentimento pelo número total de tweets normalizados. Uma vez estabelecido o baseline de sentimentos, tornou-se possível computar o aumento relativo ou diminuição nos sentimentos para uma amostra particular de tweets. Dessa forma, o PANAS-t escore de um sentimento ($P(s)$) é definido como um vetor de 11 dimensões onde a função de escore calcula um valor entre -1 e +1 para cada sentimento. Por exemplo, um evento com $P(medo) = 0$ significa que o evento não teve qualquer aumento ou diminuição para o sentimento "medo"em comparação com o dataset. Um valor positivo de 0.3 significa um aumento de 30% e etc.

Baseado em *crowdsourcing*, [Mohammad & Turney, 2013] desenvolveu um léxico de sentimento geral. Cada entrada lista a associação de um token com 8 sentimentos básicos: alegria, tristeza, raiva e etc., definidas por [Plutchik, 1980]. O léxicon proposto inclui unigramas e bigramas do Thesaurus e também palavras da WordNet.

Inferir classificações adicionais de comentários de usuários através da análise de sentimento de comentários e integrar seus outputs em um modelo de vizinho mais próximo (Nearest Neighbor(NN)) que fornece recomendações de multimídia sobre TED talks é o trabalho proposto por [Pappas & Popescu-Belis, 2013]. Ele se concentra na investigação de comentários de usuários que não são acompanhados de rótulos de classificações explícitas e sua utilidade para uma tarefa de filtragem colaborativa de uma classe como bookmarking, onde somente as ações do usuário são dadas como grau de confiança. Ele propõe um modelo de sentimento do vizinho mais próximo chamado SANN. O modelo supera significativamente, em mais de 25% em dados não vistos, diversos baselines competitivos.

Vader [Hutto & Gilbert, 2014] é um modelo baseado em regras simples para a análise de sentimento. Ele cria um conjunto de características léxicas (juntamente com suas medidas de intensidade de sentimento) que são especialmente preparadas para detectar sentimentos em contextos de microblogs. Em seguida, ele combina essas características lexicais com a consideração de cinco regras gerais que incorporam convenções gramaticais e sintáticas para expressar e enfatizar a intensidade do sentimento.

O trabalho de [Thelwall, 2013] constrói um dicionário léxico anotado por humanos

e aprimorado com o uso de Machine Learning.

Uma ferramenta gratuita para desambiguar tweets usando dicionário léxico com heurísticas para detectar palavras negativas mais alongadas e avaliação de hashtags é o trabalho de [lev, 2013]. A ferramenta se chama Umigon [1]. Além disso, Umigon oferece indicadores para *features* adicionais semânticas presentes nos tweets, como por exemplo, indicações de tempo ou marcadores de subjetividade.

Enquanto que Semantria é uma ferramenta paga que emprega análise multi-nível de sentenças. Basicamente, ela tem quatro níveis: parte do discurso, atribuição de pontuação anteriores de dicionários, aplicação de intensificadores e, finalmente, técnicas de aprendizado de máquina para entregar um peso final da frase.

Com o objetivo de proporcionar uma comparação entre diversos métodos de classificação de polaridade existentes, [Gonçalves et al., 2013] criou um serviço Web gratuito chamado iFeel. Essa ferramenta permite comparar o resultado de 19 técnicas da literatura na classificação de um texto em três níveis de polaridade (positiva, neutra e negativa) [Ribeiro et al., 2015]. As técnicas utilizadas são: afinn[Nielsen, 2011], emolex[Mohammad & Turney, 2013], emoticons [Gonçalves et al., 2013], emoticonds[Hannak et al., 2012], happinessindex[Bradley et al., 1999], opinnionfinder[Wilson et al., 2005], nrchashtagMohammad [2012], opinionlexicon[Hu & Liu, 2004], panas[Gonçalves et al., 2013], sann[Pappas & Popescu-Belis, 2013], sasa[Wang et al., 2012], senticnet[Cambria et al., 2010], sentiment140[Mohammad et al., 2013], sentistrength[Thelwall, 2013], sentiwordnet[Baccianella et al., 2010], socal[Taboada et al., 2011], stanford[Socher et al., 2013], umigonlev [2013] e vaderHutto & Gilbert [2014]

Para usar a ferramenta, basta submeter um arquivo csv com as sentenças. Ao final do processo, o iFeel entrega 19 resultados sendo um para cada técnica implementada onde cada valor representa uma polaridade -1 (negativa), 0 (neutra) ou 1 (positiva). Em nosso trabalho, submeteremos o nosso dataset à iFeel e usaremos os 19 resultados como *features* para uma nova etapa de classificação de cinco níveis de polaridade.

Uma proposta de aprendizagem automática para classificar o sentimento de mensagens curtas/reviews é o trabalho de [Canuto et al., 2016]. O autor propõe o uso de informações derivadas de características de meta-nível, isto é, características derivadas principalmente da representação original de *bag-of-words*. Baseada no KNN, dentre as meta-características, ele extrai informações derivadas da distribuição de sentimentos entre os k vizinhos mais próximos de um determinado documento de teste x, a distribuição das distâncias de x para seus vizinhos, a polaridade dos documentos vizinhos obtidas por métodos lexicais não supervisionados. O conjunto de *features* proposto

[1] Umigon: http://www.umigon.com/

pelo autor tem como objetivo a transformação do espaço original de *features* para um novo espaço menor e mais informativo. Em nosso trabalho faremos o uso dessa técnica como forma de extrair as *features* para o nosso problema.

2.3 Técnicas que avaliam mais de três polaridades

Criar uma escala numérica de nove níveis para representar o índice de felicidade em diversos textos online como letras musicais e blogs, o trabalho proposto por [Dodds & Danforth, 2010] faz uso dos termos da ANEW [Bradley et al., 1999] para construir uma escala de valência (sentimentos) entre 1 (baixa felicidade) e 9 (alta felicidade). ANEW é uma coleção de 1.034 palavras do idioma inglês associadas com as suas dimensões afetivas de sentimentos, excitação e dominância. O método inicialmente realiza um cálculo de frequência com que cada palavra da ANEW aparece no texto e, em seguida calcula a média ponderada da valência dos termos para determinar o escore geral do texto.

Construir um léxico de sentimento baseado no Twitter incluindo gírias da Internet e palavras obscenas e avaliar o seu desempenho é o propósito de [Nielsen, 2011]. O autor rotulou manualmente a AFINN, uma lista de palavras do idioma inglês onde cada termo recebeu um valor inteiro entre -5 (muito negativo) e +5 (muito positivo) representado a valência. Em seguida, ele fez um comparativo de desempenho da AFINN com outras listas de palavras afetivas como a ANEW (Affective Norms for English Words) [Bradley et al., 1999] usando um dataset de tweets rotulados pela Amazon Mechanical Turk(AMT). Como método para computar o sentimento geral de um tweet, o autor primeiro identificou as palavras e então obteve a valência de cada termo consultado os léxicos de sentimentos. A soma das valências das palavras dividido pelo número de palavras representa a força de sentimento combinado para o tweet.

Seguindo uma abordagem léxica, [Taboada et al., 2011] propõe um método chamado Semantic Orientation CALculator(SO-CAL) para extração de sentimento de um texto no idioma inglês. Ele usa dicionários de palavras anotadas com suas orientações semânticas (polaridade e força) e incorpora intensificação e negação. A técnica é aplicada à tarefa de classificação de polaridade. Para tal, ele criou um léxico com unigramas (verbos, advérbios, substantivos e adjetivos) e multi-gramas contendo verbos compostos (*phrasal verbs*) e intensificadores a partir de um corpus de comentários sobre livros, carros, computadores, hotéis, filmes, música e telefones. Cada termo foi pontuado manualmente com um valor entre -5 (extremamente negativo) e +5 (extremamente positivo) onde termos pontuados como 0 (neutros) foram descartados.

O trabalho proposto por [Mohammad, 2012] mostra a criação automática de um dicionário léxico usando hashtags. O autor coletou tweets usando as hashtags #anger, #disgust, #fear, #happy, #sadness e #surprise para criar um corpus emocional de tweets (TEC) representando respectivamente as seis emoções: raiva, nojo, medo, alegria, tristeza e surpresa. Em seguida, utilizando o corpus TEC, ele verifica a frequência de cada n-grama específico em uma emoção e calcula sua força de associação para essa emoção. Ele repete o mesmo processo para um dataset de manchetes de jornais e então cria um léxico de sentimento chamado NRC.

Enquanto que [Socher et al., 2013] propõe um modelo chamado Recursive Neural Tensor Network(RNTN) que processa todas as sentenças lidando com suas estruturas em computa as interações entre eles. Esta abordagem é interessante uma vez que RNTN leva em conta a ordem das palavras numa frase, que é ignorado na maior parte dos métodos.

Capítulo 3

Classificação de Polaridade em Cinco Níveis

Neste capítulo apresentamos a nossa proposta de aprendizagem supervisionada para a detecção das cinco classes de polaridade em postagens de microblogs. O método proposto identificará, nos tweets que expressam opiniões sobre entidades predeterminadas, cinco classes de polaridades, visando contribuir com uma classificação eficiente para o auxílio da compreensão do sentimento público além de ser genérica e passível de ser aplicada a contextos como o de produtos, marcas, serviços, turismo e etc. O método proposto atua apenas na classificação da polaridade expressa no tweet, ou seja, não está no contexto da solução contemplar técnicas que façam o reconhecimento prévio de entidades em tweets.

Para realizar a classificação de opinião, serão consideradas as cinco classes de polaridade descritas no Capítulo 1: *aprovação completa, opinião positiva pontual, neutro, opinião negativa pontual, e rejeição completa.*

Tweet	**Polaridade**
a Dilma é a melhor presidente de todos os tempos	aprovação completa
a Dilma conseguiu equilibrar as contas do governo	positiva pontual
não sei o que pensar da Dilma	neutra
a Dilma não devia ter colocado o Eduardo no ministério de minas e energia	negativa pontual
eu odeio a Dilma	rejeição completa

Tabela 3.1: Exemplos das cinco classes de polaridade. Essa tabela corresponde aos exemplos de polaridades no contexto político para a candidata Dilma

A Tabela 3.1 ilustra as cinco classes de polaridade. Os tweets que expressam uma *aprovação completa* em geral indicam um apoio total ao candidato mencionado, ou seja, uma preferência clara pelo candidato mencionado. Os tweets que expressam uma

opinião *positiva pontual* ou *negativa pontual* em geral mostram uma crítica eventual sobre o candidato onde não é possível afirmar uma total preferência ou reprovação pelo candidato mencionado. E os tweets que expressam uma *rejeição completa* em geral mostram uma reprovação, um repúdio em relação ao candidato.

Na seção a seguir, mostraremos a nossa proposta para classificação de polaridade de cinco níveis.

3.1 Modelo para classificação de polaridade de cinco níveis

A maioria das técnicas de identificação de polaridades classifica as sentenças em três possíveis classes: *positiva, negativa e neutra.* Contudo, em nosso problema, queremos propor um método supervisionado capaz de identificar cinco classes de polaridades de opiniões.

De uma forma geral, seja $T = \{t_1, t_2, t_3, ..., t_n\}$ uma coleção de tweets e $C = \{c_1, c_2, c_3, c_4, c_5\}$ um conjunto de polaridades que representa respectivamente as cinco classes de polaridade: *aprovação completa, opinião positiva pontual, neutro, opinião negativa pontual e rejeição completa.* A tarefa de análise de polaridade proposta neste trabalho pode ser vista como uma tarefa de classificação onde, para cada tweet t_n, pretende-se predizer o rótulo c_j, ou seja, encontrar uma função (modelo) $f : T \Rightarrow C$, tal que $f(t_n) = c_j$

Um dos desafios característicos deste problema é encontrar um conjunto de *features* que melhor discrimine as classes de polaridade e que facilite o aprendizado durante a fase de treino do classificador. A Figura 3.1 mostra uma visão geral da nossa proposta baseada em uma abordagem de aprendizagem supervisionada para solucionar o problema de classificação de cinco níveis.

Dessa forma, para esse trabalho pretendemos usar duas técnicas distintas de extração de *features.* A escolha da melhor técnica será definida através de um conjunto de experimentos para avaliar a acurácia dos classificadores nos diferentes conjuntos de *features* extraídos a partir de um *dataset* rotulado com as cinco classes de polaridade.

Além do modelo de *bag-of-words* que será o nosso modelo de *baseline*, também propomos o uso de evidências de 19 métodos da literatura através da ferramenta iFeel [Gonçalves et al., 2013] e da técnica de extração de meta-características [Canuto et al., 2016] que serão explicadas nas próximas seções.

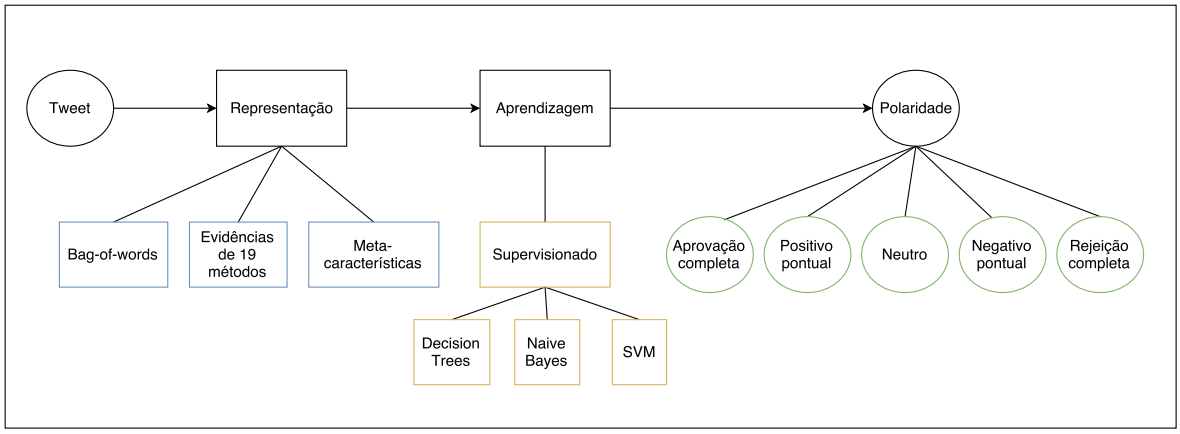

Figura 3.1: Proposta para a aprendizagem de polaridade de cinco níveis

3.1.1 Evidências de 19 métodos da literatura

O trabalho de [Gonçalves et al., 2013] disponibiliza uma ferramenta online que pode ser usada para avaliar a polaridade de sentenças em geral. Seu uso é relativamente simples, basta submeter um csv com os tweets e a ferramenta utiliza 19 técnicas diferentes para classificar cada tweet em uma polaridade positiva, negativa ou neutra. O nome da ferramenta é iFeel [1]. As Tabelas 3.2 e 3.3 listam brevemente cada técnica utilizada.

Um dos objetivos da iFeel é permitir a comparação dos resultados entre diferentes técnicas de análise de polaridade. E para que isso seja possível, a iFeel realiza uma série de adaptações de modo que todos os resultados são transformados para uma representação de polaridade de três níveis (-1, 0, 1) [Ribeiro et al., 2015]. A Tabela 3.4 mostra a estratégia de adaptação de resultado para cada técnica de modo que a cor vermelha mostra a adaptação para a polaridade negativa, a cor azul para a polaridade positiva e a cor preta para a polaridade neutra. Em algumas técnicas léxicas que possuem apenas duas polaridades, a iFeel considera como neutra os casos onde essas técnicas não conseguem detectar qualquer polaridade.

Dado um tweet a ser classificado, a nossa estratégia é usar as classes identificadas pelas 19 técncias do iFeel como features de uma nova etapa de classificação, desta vez para classificar o tweet entre as cinco classes de polaridade. Desta forma, queremos avaliar se os resultados de um conjunto de técnicas de identificação de polaridade em três níveis (positivo, neutro, negativo) pode ajudar a discriminar as cinco classes do nosso trabalho.

A tabela 3.5 mostra um exemplo de classificação feita pela iFeel. Uma classificação -1 representa polaridade negativa, 1 representa positiva e 0 representa polaridade neutra.

[1] iFeel: http://blackbird.dcc.ufmg.br:1210/

Id na iFeel	Referência Bibliográfica
afinn	Léxico de sentimento baseado no Twitter incluido gírias da Internet e palavras obscenas [Nielsen, 2011]
emolex	Léxico de sentimento onde cada elemento do dicionário lista a associação de um token com 8 sentimentos básicos: alegria, raiva, tristeza, etc. O léxico proposto inclui unigramas e bigramas de Macquarie Thesaurus e termos da Wordnet [Mohammad & Turney, 2013].
emoticons	Técnica onde mensagens contendo emoticons positivos/negativos são consideradas positivas/negativas. Mensagens sem a presença de emoticons não são classificadas [Gonçalves et al., 2013].
emoticonds	Cria um léxico pontuado com base em um grande *dataset* de tweets. O escore de pontuação é baseado na freqüência que cada léxico ocorre com emoções positivas ou negativas [Hannak et al., 2012].
happinessindex	Quantifica níveis de felicidade em textos como letras de músicas e blogues. Ele usa termos da ANEW [Bradley et al., 1999] para ranquear documentos Dodds & Danforth [2010].
opinnionfinder	*Framework* para identificar sentenças subjetivas e marcar vários aspectos da subjetividade nestas sentenças, incluindo a fonte (detentora) da subjetividade e as palavras que são incluídas em frases expressando sentimentos positivos ou negativos [Wilson et al., 2005].
nrchashtag	Usa hashtags conhecidas (*#joy, #sadness, #happy etc.*) para classificar o tweet. Depois ele verifica a freqüência de cada n-grama específico que ocorre em uma emoção e calcula a força de associação com essa emoção Mohammad [2012].
opinionlexicon	Com o foco em revisões de produtos. Constrói um léxico para prever a polaridade das frases sobre características do produto. Essas polaridades são somadas para produzir uma pontuação geral por característica do produto [Hu & Liu, 2004].
panas	Versão adaptada de PANAS [Watson et al., 1988], método bem conhecido em psicologia com um grande conjunto de palavras onde cada uma delas está associada a um dos onze modos (surpresa, medo, jovialidade , segurança, serenidade, tristeza, culpa, hostilidade, timidez, fadiga e atenção) [Gonçalves et al., 2013].

Tabela 3.2: Descrição dos métodos usados na iFeel

Ident. na iFeel	Referência Bibliográfica
sann	Inferir classificações adicionais de comentários de usuários através da análise de sentimento de comentários e integrar seus outputs em um modelo de vizinho mais próximo (*Nearest Neighbor(NN)*) que fornece recomendações de multimídia sobre TED talks [Pappas & Popescu-Belis, 2013]
sasa	Sistema em tempo real para a análise de sentimentos do público sobre os candidatos a presidência dos Estados Unidos em 2012. Ele avaliou postagens com opiniões (positivas, negativas e neutras) e pontos de vista sobre os partidos e candidatos [Wang et al., 2012].
senticnet	Usar redução de dimensionalidade para inferir a polaridade de conceitos do senso comum, o SenticNet é uma ferramenta pública para minerar opiniões de textos em linguagem natural através de uma abordagem semântica ao invés de explorar apenas evidências sintáticas [Cambria et al., 2010].
sentiment140	Dicionário léxico de forma semelhante a [Mohammad, 2012], mas os autores usaram a ocorrência de emoticons para classificar o tweet como positivo ou negativo. Em seguida, a pontuação de n-grama foi calculada com base na frequência que ocorre em cada classe de tweets [Mohammad et al., 2013].
sentistrength	Constrói um dicionário léxico anotado por humanos e melhorado com o uso de Aprendizagem de Máquina [Thelwall, 2013].
sentiwordnet	Construção de um recurso léxico para mineração de opinião baseado na WordNet. Os autores agruparam adjetivos, substantivos, etc. em conjuntos de sinônimos *(synsets)* e três escores de polaridade associados (positivos, negativos e neutros) para cada um [Baccianella et al., 2010].
socal	Cria um novo léxico com unigramas (verbos, advérbios, substantivos e adjetivos) e multi-gramas (verbos phrasal e intensificadores). Os autores também incluíram parte do processamento da fala, negação e intensificadores [Taboada et al., 2011].
stanford	Modelo denominado *Recursive Neural Tensor Network* (RNTN) que processa todas as frases relacionadas com as suas estruturas e calcula as interações entre elas [Socher et al., 2013].
umigon	Uma ferramenta gratuita para desambiguar tweets usando dicionário léxico com heurísticas para detectar palavras negativas mais alongadas e avaliação de hashtags lev [2013].
vader	Conjunto de características léxicas (juntamente com suas medidas de intensidade de sentimento) que são especialmente preparadas para detectar sentimentos em contextos de microblogs. Ele combina essas características lexicais com a consideração de cinco regras gerais que incorporam convenções gramaticais e sintáticas para expressar e enfatizar a intensidade do sentimento Hutto & Gilbert [2014].

Tabela 3.3: Descrição dos métodos usados na iFeel

Id na iFeel	Referência	Adaptação do Resultado	Resultado na iFeel
afinn	[Nielsen, 2011]	**-1, 0, 1**	**-1, 0, 1**
emolex	[Mohammad & Turney, 2013].	**-1, 0, 1**	**-1, 0, 1**
emoticons	[Gonçalves et al., 2013].	**-1, 1**	**-1, 0, 1**
emoticonds	[Hannak et al., 2012].	**-1, 0, 1**	**-1, 0, 1**
happinessindex	Dodds & Danforth [2010].	**1, 2, 3, 4, 5, 6, 7, 8, 9**	**-1, 0, 1**
opinnionfinder	[Wilson et al., 2005].	**negative, objective, positive**	**-1, 0, 1**
nrchashtag	Mohammad [2012].	**-1, 0, 1**	**-1, 0, 1**
opinionlexicon	[Hu & Liu, 2004].	**-1, 0, 1**	**-1, 0, 1**
panas	[Gonçalves et al., 2013].	**-1, 0, 1**	**-1, 0, 1**
sann	[Pappas & Popescu-Belis, 2013]	**neg, neu, pos**	**-1, 0, 1**
sasa	[Wang et al., 2012].	**negative, neutral, unsure, positive**	**-1, 0, 1**
senticnet	[Cambria et al., 2010].	**negative, positive**	**-1, 0, 1**
sentiment140	[Mohammad et al., 2013].	**-1, 0, 1**	**-1, 0, 1**
sentistrength	[Thelwall, 2013].	**-1, 0, 1**	**-1, 0, 1**
sentiwordnet	[Baccianella et al., 2010].	**-1, 0, 1**	**-1, 0, 1**
socal	[Taboada et al., 2011].	**[<0), 0, (>0]**	**-1, 0, 1**
stanford	[Socher et al., 2013].	**very negative, negative, neutral, positive, very positive**	**-1, 0, 1**
umigon	lev [2013].	**negative, neutral, positive**	**-1, 0, 1**
vader	Hutto & Gilbert [2014].	**-1, 0, 1**	**-1, 0, 1**

Tabela 3.4: Adaptação dos resultados de todos os 19 métodos para a polaridade de três níveis

Texto	Resultado iFeel (19 métodos)
Eu odeio a dilma	afinn=-1"emolex=-1"emoticons="0"emoticonds="0"happinessindex=-1"opinionfinder=-1" nrchashtag=-1"opinionlexicon=-1"panas="0"sann="1"sasa=-1"senticnet=-1" sentiment140=-1"sentistrength=-1"sentiwordnet=-1"socal=-1"stanford=-1"umigon=-1"vader=-1"
Eu amo a marina	afinn="1"emolex="1"emoticons="0"emoticonds="1"happinessindex="1"opinionfinder="1" nrchashtag="1"opinionlexicon="1"panas="0"sann="1"sasa="1"senticnet="1" sentiment140="1"sentistrength="1"sentiwordnet="1"socal="1"stanford="1"umigon="1"vader="1"

Tabela 3.5: Exemplo de classificação da iFeel. Esta tabela mostra o resultado de classificação para: Eu odeio a Dilma e Eu amo a Marina

3.1.2 Meta-características

Uma tendência recente que surgiu em abordagens supervisionadas para classificações de textos é o uso de meta-características (*meta-features*). As *meta-features* são, em geral estatísticas que descrevem o *dataset* do problema, como por exemplo, o número de sentenças, número de atributos, correlação entre atributos, entropia de classe e etc. O propósito do uso de *meta-features* é reduzir drasticamente o número de atributos do input de entrada e selecionar apenas *features* relevantes que melhoram a acurácia dos algoritmos de classificação. Com a intenção de propor modelos de *meta-features* para análise de sentimentos em mensagens curtas como o Twitter, o trabalho proposto por [Canuto et al., 2016] descreve um conjunto de *meta-features* baseadas no KNN para explorar informações locais de cada instância de um *dataset* rotulado. Ele faz uma busca dos vizinhos mais próximos para extrair as *meta-features* para treinar um modelo de classificação de três níveis.

Aplicando as definições de [Canuto et al., 2016] ao nosso contexto, onde dado o conjunto de tweets T, o vetor de *meta-features* $m_f \in M$ é composto pela concatenação dos vetores (*Rawsim, Cagglex, Maxminlex, Agglex, Rawlex*) para cada exemplo $t_n \in T$ e polaridade $c_j \in C$ para $j = 1, 2, ..., |C|$. Os vetores serão descritos nas próximas seções a seguir.

3.1.2.1 Rawsim

A Figura 3.2 ilustra o processo de extração do vetor de meta-característica *Rawsim* para representar um tweet t_n a partir dos k-vizinhos de cada classe de polaridade.

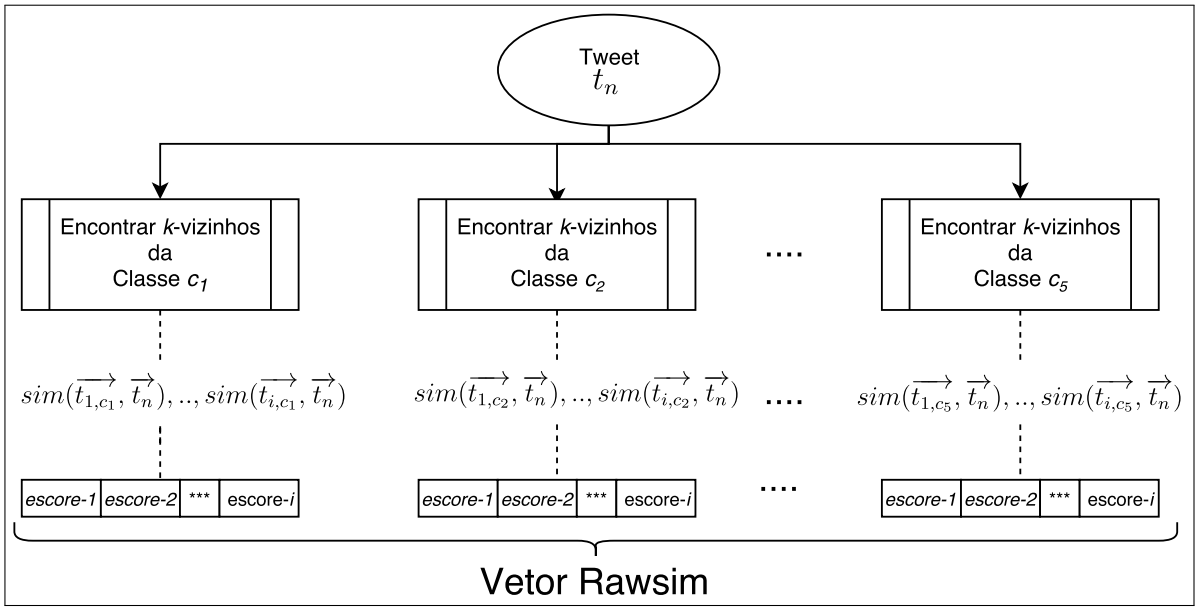

Figura 3.2: Processo de extração de meta-características *Rawsim*

A meta-característica *Rawsim (Raw Similarity)*, é formalmente definida como:

$$\overrightarrow{v}^{rawsim}_{\overrightarrow{t_n}} = [sim(\overrightarrow{t_{1,c_1}}, \overrightarrow{t_n}), sim(\overrightarrow{t_{2,c_1}}, \overrightarrow{t_n}), ..., sim(\overrightarrow{t_{1,c_2}}, \overrightarrow{t_n}), ..., sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n})]$$

Desta forma, a meta-característica *Rawsim* é um vetor $k|C|$-dimensional que considera os k tweets vizinhos mais próximos de cada classe de polaridade para o vetor alvo t_n. Mais especificamente, $\overrightarrow{t_{i,c_j}}$ é o i-ésimo ($i \leq k$) tweet vizinho de classe c_j mais próximo de $\overrightarrow{t_n}$, e $sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n})$ é o escore de similaridade (BM25 ou cosseno) entre eles. Dessa forma, $k|C|$ meta-características são geradas para representar t_n.

Na meta-característica *Rawsim* , cada exemplo de teste é diretamente comparado a um conjunto de exemplos rotulados mais próximos de cada categoria. A intuição por trás dessas meta-características consiste na suposição de que se as distâncias entre um exemplo para os vizinhos mais próximos pertencentes a uma categoria c_j são pequentas, então o exemplo provavelmente pertence a c_j.

3.1.2.2 Cagglex

A Figura 3.3 ilustra o processo de extração do vetor de meta-característica *Cagglex* para representar um tweet t_n a partir dos k-vizinhos de cada classe de polaridade.

A meta-característica *Cagglex* (*Categories Agglomerated Combined with Lexical Scores*), é formalmente definida como:

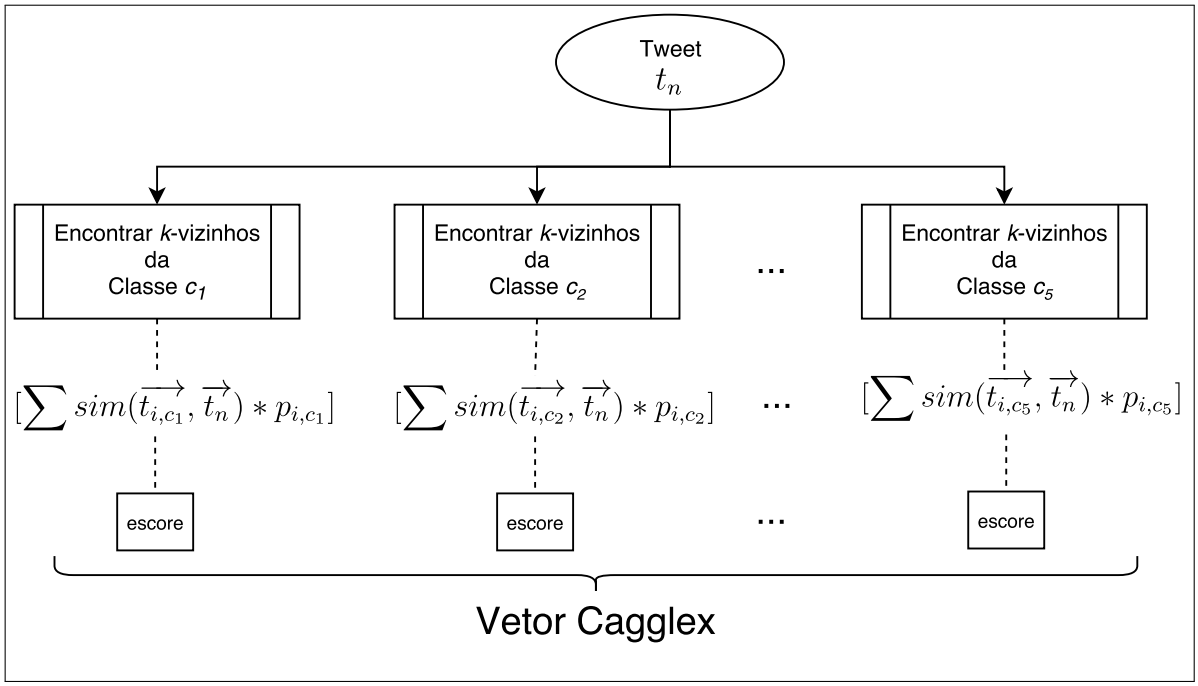

Figura 3.3: Processo de extração da meta-característica *Cagglex*

$$\overrightarrow{v}_{\overrightarrow{t_n}}^{cagglex} = [(\sum sim(\overrightarrow{t_{i,c_1}}, \overrightarrow{t_n}) * p_{i,c_1}), (\sum sim(\overrightarrow{t_{i,c_2}}, \overrightarrow{t_n}) * p_{i,c_2}), \ldots, (\sum sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n}) * p_{i,c_j})]$$

Deste modo, a meta-característica *Cagglex* é um vetor $|C|$-dimensional que considera o somatório dos k tweets vizinhos de t_n que pertencem a cada classe de polaridade de c_j. A similaridade $sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n})$ (BM25 ou cosseno) entre $\overrightarrow{t_n}$ e o seu i-ésimo ($i \leq k$) vizinho mais próximo $\overrightarrow{t_{i,c_j}}$ é usada como fator de ponderação para a polaridade p_{i,c_j} do tweet $\overrightarrow{t_{i,c_j}}$. O escore de polaridade p_{i,c_j} é dado por um método léxico (ex: Sentistrength). A meta-característica *Cagglex* fornece a polaridade aglomerada dos vizinhos de cada categoria de C.

3.1.2.3 Maxminlex

A Figura 3.4 ilustra o processo de extração do vetor de meta-característica *Maxminlex* para representar um tweet t_n a partir dos k-vizinhos mais próximos de cada classe de polaridade.

A meta-característica *Maxminlex* (*Maximum and Minimum Combined with Le-*

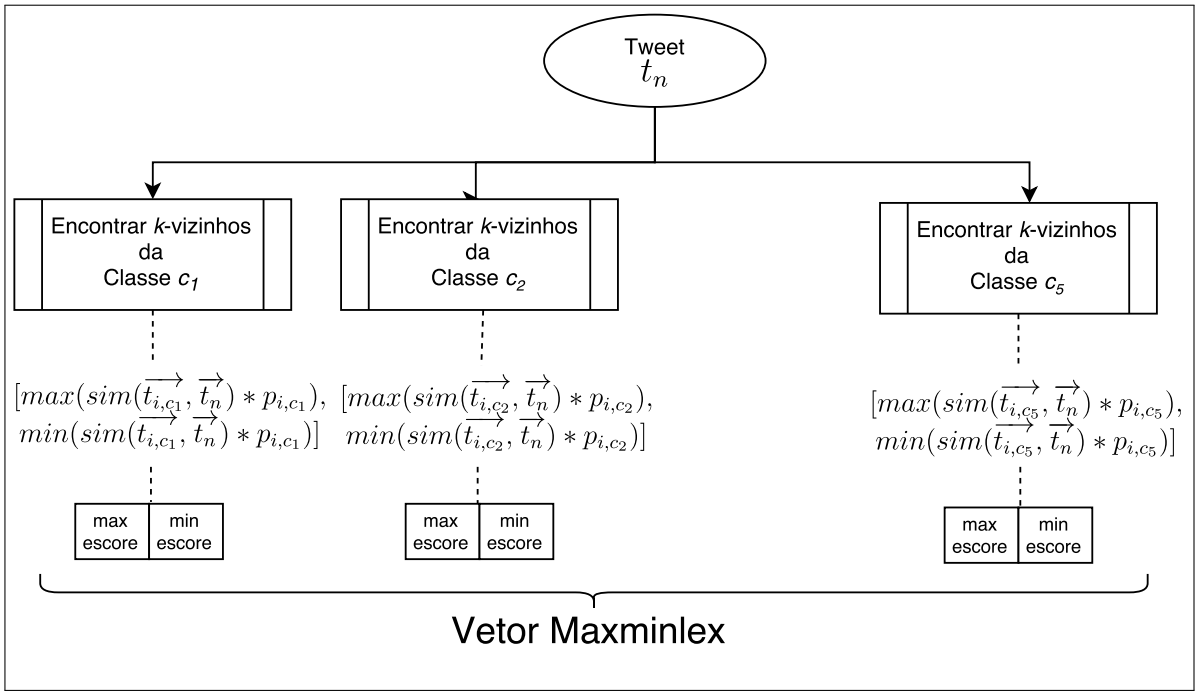

Figura 3.4: Processo de extração da meta-característica *Maxminlex*

xical Scores), é formalmente definida como:

$$\begin{aligned}\overrightarrow{v}_{\overrightarrow{t_n}}^{maxminlex} = [&max(sim(\overrightarrow{t_{i,c_1}}, \overrightarrow{t_n}) * p_{i,c_1}), min(sim(\overrightarrow{t_{i,c_1}}, \overrightarrow{t_n}) * p_{i,c_1}),\\ &max(sim(\overrightarrow{t_{i,c_2}}, \overrightarrow{t_n}) * p_{i,c_2}), min(sim(\overrightarrow{t_{i,c_2}}, \overrightarrow{t_n}) * p_{i,c_2}),\\ &\ldots,\\ &max(sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n}) * p_{i,c_j}), min(sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n}) * p_{i,c_j})]\end{aligned}$$

Portanto, a meta-característica *Maxminlex* é um vetor $2|C|$-dimensional que considera as polaridades ponderadas máxima e mínima dos k tweets vizinhos do vetor alvo t_n que pertencem a cada classe c_j. A similaridade $sim(\overrightarrow{t_{i,c_j}}, \overrightarrow{t_n})$ (BM25 ou cosseno) entre t_n e os seus i-ésimos ($i \leq k$) tweets vizinhos mais próximos t_{i,c_j} é usada para ponderar a polaridade p_{i,c_j} do tweet t_{i,c_j}. O escore de polaridade p_{i,c_j} é dado por um método léxico.

A meta-característica *Maxminlex* extrai os valores máximos e mínimos de cada classe de polaridade que complementam a polaridade aglomerada extraída por *Cagglex*

3.1.2.4 Agglex

A Figura 3.5 ilustra o processo de extração do vetor de meta-característica *Agglex*.

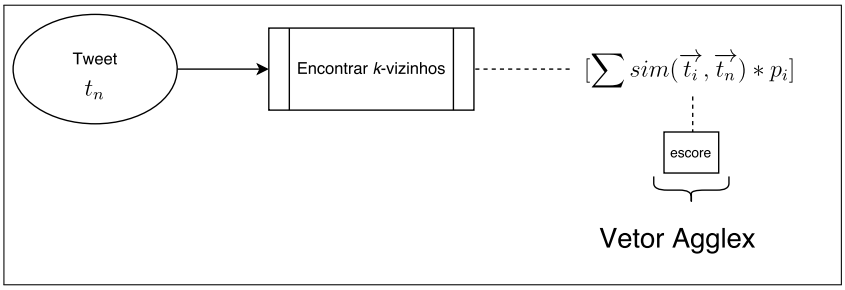

Figura 3.5: Processo de extração da meta-caracterísitca *Agglex*

A meta-característica *Agglex* (*Agglomerated and Combined with Lexical Score*), é formalmente definida como:

$$\overrightarrow{v}^{agglex}_{\overrightarrow{t_n}} = [\sum sim(\overrightarrow{t_i}, \overrightarrow{t_n}) * p_i]$$

Assim sendo, a meta-característica *Agglex* é um vetor unidimensional com a soma das polaridades ponderadas dos k-vizinhos mais próximos. A similaridade $sim(\overrightarrow{t_i}, \overrightarrow{t_n})$ (BM25 ou cosseno) entre t_n e seu i-ésimo ($i \leq k$) vizinho mais próximo t_i é usada para ponderar a polaridade p_i do tweet t_i. O escore de polaridade p_i é dado por um método léxico.

A meta-característica *Agglex* difere das demais pois apenas sumariza a polaridade dos k vizinhos mais próximos sem levar em consideração suas categorias.

3.1.2.5 Rawlex

A Figura 3.6 ilustra o processo de extração do vetor de meta-característica *Rawlex*.

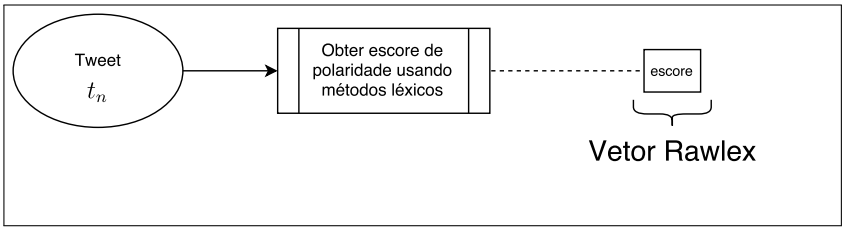

Figura 3.6: Processo de extração da meta-caracterísitca *Rawlex*

A meta-característica *Rawlex* (*Raw Similarity Using Lexical Score*) são escores produzidos pelos resultados de métodos léxicos. Os escores de polaridades positiva e/ou negativa produzidos pelos métodos são combinados em um único escore como resultado.

Implementaremos as cinco meta-características baseadas no kNN para extrair as meta-características *Rawsim, Rawlex, Cagglex, Agglex e Maxminlex* propostas. O objetivo é avaliar se as mesmas meta-informações que tiveram um bom desempenho em problemas onde a polaridade é de três níveis, também podem trazer bons resultados em nosso contexto com cinco níveis. Desta forma, o modelo de meta-características que representa cada tweet, é a concatenação dessas cinco meta-características conforme a figura 3.7

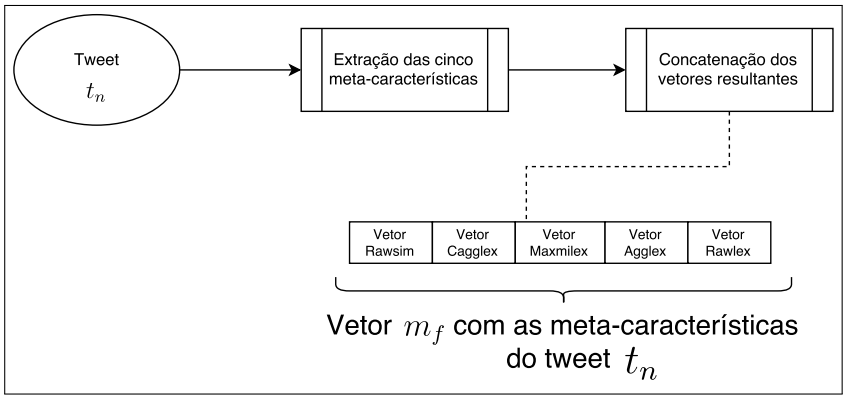

Figura 3.7: Representação do vetor m_f com as meta-características do tweet t_n

3.1.3 Combinação de features

Com a finalidade de avaliar o impacto das *features* nas acurácias dos algoritmos de classificação, e consequentemente permitir a comparação entre os modelos obtidos com o modelo baseline (BoW), elaboramos combinações entre os modelos de *features* implementados neste trabalho. Portanto, combinaremos BoW+iFeel, e dentre os modelos de meta-característica, selecionaremos apenas o modelo que alcançar os melhores resultados de acurácia nos experimentos de classificação.

Logo, as combinações de features realizadas neste trabalho serão: BoW + iFeel, BoW + (Melhor modelo de meta-característica), iFeel + (Melhor modelo de meta-característica) e BoW+iFeel+(Melhor modelo de meta-característica)

Capítulo 4

Experimentos

Neste capítulo apresentamos o passo a passo dos experimentos deste trabalho começando pela descrição do processo de coleta de tweets do *dataset* adotado nesses experimentos. Em seguida, mostramos o procedimento de escolha e avaliação manual dos tweets para então definir um corpus inicial. Logo após, mostramos as etapas de pré-processamento e balanceamento do *dataset* para então iniciar a etapa de extração das *features* e busca de parâmetros. Por fim, apresentamos os resultados obtidos usando os algoritmos de Árvore de Decisão, Naive Bayes e *Support Vector Machine* em cada conjunto de *features*.

4.1 Coleta de Tweets

De forma a obter tweets relacionados ao contexto político, construímos um crawler que usou a api de *streamming* do Twitter para coletar, durante os três meses que antecederam as eleições de 2014, um conjunto de tweets sobre os três principais candidatos à presidência do referido pleito eleitoral: Dilma, Aécio e Marina. No entanto, como essa api carrega tweets em âmbito global, foi necessário especificar filtros usando palavras-chave a fim de coletar somente tweets que tinham a ver os três candidatos. Segundo [Cunha et al., 2011], as hashtags podem ser usadas para classificar mensagens, propagar ideias e também para identificar tweets de temas e pessoas específicas. Com esse propósito, usamos um conjunto de hashtags com referência aos três candidatos à presidência para filtrar o *stream* de dados. Esse conjunto foi obtido a partir do site Top-Hashtags [1], onde basta informar um termo desejado para que o site busque as hashtags mais populares relacionadas ao termo. Em nosso caso, usamos os nomes dos

[1]Top-Hashtags:https://top-hashtags.com/

candidatos para obter as hashtags onde selecionamos as mais relacionadas ao contexto eleitoral.

Dentre as limitações que este método de seleção de hashtags possui, observa-se que o uso de hashtags populares sobre o contexto eleitoral pode levar à coleta de conteúdo enviesado [Jungherr, 2013], com spam [Sedhai & Sun, 2016] e sujeito a ironia [Charalampakis et al., 2015].

As Tabelas 4.1, 4.2 e 4.3 listam as hashtags relacionadas ao contexto eleitoral.

hashtags - Dilma		
#dilma13	*#dilma13maisnordeste*	*#mulherescomdilma13*
#13rasiltodocomdilma	*#maisamor13*	*#dilmadenovo*
#dilma13mudamais	*#lulaedilma13neles*	*#vote13*
#dilma13denovo	*#todoscomdilma*	*#dilma13ideiasnovas*
#dilma13pravencer	*#brasilcomdilma13*	*#felizcomdilma13*
#dilma13maisfuturo	*#dilmais4anos*	*#militânciapelobrasil13*
#dilma13maisbrasil	*#periferiacomdilma*	*#agita13*
#menosodiomaisdilma	*#porissodilma13*	*#dilma13maiseducação*
#brasil13	*#dilmaetarso13denovo*	*#dilmaepadilha13*
#dilma13maisemprego	*#13dilmaconfirma*	*#dilmapresidenta*
#dilmanao	*#foradilma*	*#dilma13bandalarga*
#fimdesemanadilmais	*#dilmapramudarminas*	

Tabela 4.1: Lista de hashtags da Dilma

hashtags - Aécio		
#votoaeciopelobr45il	*#éaecio45confirma*	*#virada45*
#aecio45pelobrasil	*#aeciodevirada*	*#souaecio*
#aécio45	*#éaécio45confirma*	*#aeciopramudar*
#agoraeaecio45confirma	*#eaecio45confirma*	*#juntoscomaecio*
#aecio45	*#viradadoaecio*	*#agoraéaécio*
#somosaecio45	*#euvotoaecio45*	*#agoraeaeciobrasil*
#aeciopelobr45il	*#aéciopelamudança*	*#voto45*
#45aecioconfirma	*#estamoscomaecio*	*#vamosvencer45*
#emtodobrasilaecio45	*#dilmavaiperderaecio45vencer*	*#aecionafrentepelobrasil*
#aeciopelamudança	*#agoraéaeciobrasil*	*#aecioemtodobrasil45*
#aeciopelamudanca	*#aéciopresidente*	*#aécio45confirma*
#euvoudeaecio	*#souaécio*	*#agorabrasilaécio45*
#aecioemtodobrasil	*#eusouaecio*	*#euvoudeaécio*
#aecionever	*#foraaecio*	
#45aécioconfirma	*#vote45*	

Tabela 4.2: Lista de hashtags do Aécio

hashtags - Marina		
#brasilmarina40	*#euvotomarina40*	*#marina40neles*
#soumarina40	*#euvoteimarina40*	*#marinadeverdade*
#foramarina	*#forarede*	*#marinanao*
#marina40	*#marinapresidente*	

Tabela 4.3: Lista de hashtags da Marina

Usando somente essas palavras-chave, capturamos um pouco mais de 11 milhões de tweets no período de 14/08/2014 à 28/10/2014, ou seja, coletados ao longo do primeiro e segundo turno das eleições de 2014.

4.2 Escolha e Rotulagem Manual

Com a coleção de tweets obtidos a partir da coleta descrita na seção anterior, selecionamos aleatoriamente 3 mil tweets. Para montar as coleções de treino e teste, construímos uma aplicação web para disponibilizar os tweets a três avaliadores distintos. O processo de avaliação deu-se através da aplicação web onde cada avaliador rotulou os mesmos 3 mil tweets. Cada tweet poderia ser avaliado apenas uma vez por avaliador e, para cada tweet apresentado ao usuário existiam os botões *'aprovação completa', 'opinião positiva pontual', 'neutro', 'opinião negativa pontual' e 'rejeição completa'* correspondentes as cinco classes de polaridade. Uma vez que o usuário pressionava o botão desejado, o tweet era rotulado com a respectiva classe de polaridade.

Ao final do processo, usando as avaliações obtidas, foi realizado uma votação majoritária para a escolha do rótulo de polaridade de cada tweet. A polaridade vencedora era aquela que havia sido selecionada por pelo menos dois avaliadores. O tweet era descartado quando não havia um consenso em ao menos duas das três votações. A Tabela 4.4 mostra exemplos de tweets do *dataset*.

Tweet	**Polaridade**
"Daqui a pouco começam as apurações. Tds ansiosos,p ver @dilmabr ganhar no primeiro turno. #Dilma13PraVencer"	Aprovação completa
"Aécio Never finalmente deu uma dentro! O sonho do brasileiro é morar na propaganda do PT"	Positivo pontual
"#MarinaSilva está em Porto Alegre e participa de evento de campanha, hoje à noite, na Casa do Gaúcho no Parque Harmonia"	Neutro
"GENTE!!!! A #MarinaSilva não TEM propostas!!!! #ForumDebateBand uma realidade."	Negativo pontual
"Me recuso a não me mexer, a não me manifestar: #ForaMarina! Basta! Não somos idiotas, nem crianças a quem você tenta moldar"	Rejeição completa

Tabela 4.4: Exemplos de tweets do *dataset*. Esta tabela mostra um exemplo de tweet por polaridade.

Desta forma, após a etapa de votação, obteve-se um *dataset* inicial com 1.658 tweets distribuídos entre as cinco classes de polaridade conforme a tabela 4.5

Polaridade	Total de tweets
Aprovação completa	595
Positivo pontual	160
Neutro	229
Negativo pontual	396
Rejeição completa	278

Tabela 4.5: Distribuição dos tweets por classe de polaridade

4.3 Pré-processamento e Balanceamento do Dataset

Após a construção do *dataset* inicial, foi realizado um pré-processamento para a substituição das referências das contas de usuário (@usuário) pelo termo 'AT_USER' tal como sugerido em [Almatrafi et al., 2015]. Além disso, todos os tweets foram submetidos ao tradutor do Google[2] com o objetivo de convertê-los para o inglês pois a maioria das técnicas de análise de sentimentos existentes estão primariamente preparadas para tratar textos no idioma inglês conforme sugerido por [Araujo et al., 2016]. O autor abordou este problema e mostrou que a simples conversão do texto de uma linguagem específica (como o português) para o inglês, para então usar uma técnica preparada para o inglês, contribuiu com os melhores resultados ao invés de usar as técnicas específicas no próprio idioma. Além disso, o autor também mostrou que alguns métodos populares das linguagens específicas que também são todos os mesmos usados pela iFeel, não alcançaram vantagem significante sobre a abordagem de tradução. Por exemplo, até mesmo o original sentistrength com texto traduzido por máquina mostrou ser melhor do que o idioma específico.

Antes	**Depois**
Daqui a pouco começam as apurações. Tds ansiosos,p ver @dilmabr ganhar no primeiro turno. #Dilma13PraVencer	In a little while, the trials begin. All eager to see AT_USER win in the first round. #Dilma13PraVencer

Tabela 4.6: Exemplo de um pré-processamento

Diversos algoritmos tradicionais de aprendizagem de máquina lidam melhor com *datasets* balanceados, isto é, quando não há uma clara desproporção entre número de instâncias de uma ou mais classes em relação às demais. Contudo, o *dataset* inicial

[2]Google Tradutor:https://translate.google.com/

obtido a partir das votações viola essa premissa onde por exemplo a classe majoritária *Aprovação completa* tem mais que o triplo de instâncias da classe *Positivo pontual*.

Para contornar esse problema, foi utilizado uma abordagem híbrida de *Sampling* através da combinação de *under-sampling* e *over-sampling* [García & Herrera, 2009].

Under-sampling e *over-sampling* são técnicas usadas em aprendizagem de máquina para ajustar a distribuição de classes de um *dataset*. A técnica de *over-sampling* consiste em duplicar instâncias de modo a aumentar a representatividade da classe minoritária. De modo oposto, a técnica de *under-sampling* consiste em reduzir a representatividade da classe majoriária.

Este processo foi feito manualmente através de escolha aleatória das instâncias na aplicação de *under-sampling* para classe majoritária e *over-sampling* para as demais classes a fim de igualar a distribuição das classes de polaridade. A tabela 4.7 mostra o resultado desse processo de balanceamento.

Polaridade	Total de tweets
Aprovação completa	400
Positivo pontual	400
Neutro	400
Negativo pontual	400
Rejeição completa	400

Tabela 4.7: Distribuição dos tweets por classe de polaridade após o processo de over-sampling e under-sampling

4.4 Extração de Features

Esta seção descreve a geração das *features* a partir da ferramenta iFeel, do modelo bag-of-words e das meta-características.

4.4.1 Evidências de 19 métodos

Neste modelo, o *dataset* balanceado foi submetido a ferramenta iFeel[Gonçalves et al., 2013] de forma que os resultados retornados pelos 19 métodos possam ser aproveitados como *features* do problema de classificação proposto neste trabalho. Desta forma, cada tweet foi transformado para uma representação vetorial com as 19 *features* onde cada posição do vetor pode ter um valor -1, 0 ou 1. Essas *features* correspondem respectivamente aos métodos de detecção de polaridade implementados pela iFeel: *afinn, emolex,*

emoticons, emoticonds, happinessindex, nrchashtag, opinionfinder, opinionlexicon, panas, sann, sasa, senticnet, sentiment140, sentistrength, sentiwordnet, socal, stanford, umigon, vader.

Tweet	Features iFeel	Polaridade
All I want next Sunday is #Dilma in the 1st round #BrasilComDilma13	[1,0,1,0,0,1,1,0,0,0,1,1,-1,0,1,0,0,0,0]	Aprovação completa
And I do not vote on #Acio, but he's winning the debate, it's a fact.	[1,1,1,0,0,-1,0,1,0,0,1,1,-1,1,-1,0,1,0,1]	Positivo pontual
At Atuser #MarinaSilva announces tomorrow its decision on the second round	[0,0,1,0,0,0,0,0,0,0,0,1,-1,0,-1,0,0,0,0]	Neutro
Vixiiiiii #MarinaSilva is already losing the line on the first question !!!	[-1,0,1,0,0,-1,-1,0,0,0,-1,-1,-1,0,-1,-1,0,0,-1]	Negativo pontual
I do not like either candidate, but I do not want to! #Out Dilma	[1,1,1,0,0,-1,1,1,0,0,0,1,-1,0,-1,-1,-1,-1,0]	Rejeição completa

Tabela 4.8: Exemplos de transformação de tweets para representação de features da iFeel.

4.4.2 Bag-of-words(BoW)

Após um pré-processamento para a remoção de pontuação, stopwords, conversão para caixa baixa, o *dataset* foi transformado na representação de BoW usando tf-idf como fator de ponderação. Desta forma, o conjunto de *features* é o dicionário do corpus. A representação de BoW é o modelo *baseline* para este trabalho.

4.4.3 Meta-características

Com a intenção de conseguir modelos baseados em meta-características, implementamos os vetores de meta-características *Rawsim*, *Cagglex*, *Maxminlex*, *Agglex* e *Rawlex* propostas por [Canuto et al., 2016].

O autor usou Cosseno e BM25 como métricas para encontrar os vizinhos mais próximos em cada meta-característica, exceto para o vetor *Rawlex*. Da mesma forma, as mesmas métricas foram implementadas neste trabalho a fim de gerar os vetores com as meta-características. Como técnica léxica, optamos por usar somente a técnica Sentistrength[Thelwall, 2013] pois a mesma reportou os melhores resultados para análise de textos traduzidos para o idioma inglês [Araujo et al., 2016].

Para entender melhor o número de meta-características gerados por essa técnica na nossa implementação, começamos pelos vetores *Rawsim*. Especificamente, k meta-características são geradas por *Rawsim* derivadas dos k vizinhos mais próximos do *dataset*. Como são duas métricas de similaridade, é gerado um vetor para cada métrica os quais produzem $2k$ meta-características por categoria, ou $2k|C|$ e acrescentando o vetor *Rawlex* que gera uma meta-característica usando o score obtido através da técnica SentiStrength, temos $2k|C|+1$ *features*. Os vetores *Cagglex* e *Maxminlex* geram três meta-características por categoria e o vetor *Agglex* fornece apenas uma meta-

característica, gerando o total de $1+3|C|$ meta-características. Como eles estão usando duas métricas de similaridade, o número de *features* destes vetores são $2(1+3|C|)$.

Portanto, o total de meta-características obtidos por *Rawsim*, *Cagglex*, *Maxminlex*, *Agglex* e *Rawlex* para cada tweet a partir de k vizinhos mais próximos é $2k|C| + 1 + 2(1 + 3|C|)$. A partir do *dataset* balanceado tendo as cinco classes de polaridades como categoria, para fins experimentais, foram gerados quatro modelos de meta-características usando os valores de $k = 5$, $k = 10$, $k = 20$ e $k = 30$ respectivamente. A tabela 4.9 mostra os totais de *features* de cada modelo.

Modelo	**#Features**
Meta(k=5)	83
Meta(k=10)	133
Meta(k=20)	233
Meta(k=30)	333

Tabela 4.9: Quantidade de meta-características por k-vizinhos a partir do *dataset* balanceado

4.5 Ajuste de Parâmetros

O ajuste dos parâmetros de cada classificador foi efetuado através de uma busca de valores em grid a partir de um conjunto de candidatos pré-definidos para então encontrar os melhores valores que maximizam a acurácia dos classificadores através de validação cruzada. Este processo foi realizado em cada *dataset* de *features*.

Para ajustarmos os parâmetros a serem adotados nas *Decision Trees*, adotamos o método *scikit-learn*, que usa uma versão otimizada do algoritmo CART[Breiman et al., 1984]. Ele é muito semelhante ao C4.5 [Quinlan, 1993], mas difere deste último por suportar variáveis-alvo numéricas (regressão) e não computa conjuntos de regras. Esse algoritmo constrói árvores binárias usando a *feature* e o *threshold* que produz o maior ganho de informação em cada nó. Os parâmetros que foram levados em consideração na busca em grid para a escolha de valores por *dataset* que maximizam a acurácia da Árvores de Decisão foram: *class_weight, criterion, max_depth, max_features e min_samples_leaf*. A tabela 4.10 lista os melhores valores encontrados para *Decision Trees*.

Para o Naive Bayes, usamos o classificador MultinomialNB [3] que implementa o Naive Bayes para dados distribuídos multinomialmente [Manning et al., 2008]. Esse

[3]MultinomialNB:http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

	Parâmetros para Decision Trees				
Dataset	**class_weight**	**criterion**	**max_depth**	**max_features**	**min_samples_leaf**
BoW	balanced	entropy	15	auto	2
iFeel	balanced	gini	15	19	2
Meta(k=5)	balanced	entropy	10	19	15
Meta(k=10)	balanced	gini	10	19	2
Meta(k=20)	balanced	gini	10	19	2
Meta(k=30)	balanced	gini	10	19	2
BoW+iFeel	balanced	entropy	15	auto	2
BoW+Meta(k=20)	balanced	entropy	15	auto	2
iFeel+Meta(k=20)	balanced	entropy	10	auto	10
Bow+iFeel+Meta(k=20)	balanced	entropy	15	auto	2

Tabela 4.10: Melhores valores de parâmetros encontrados para *Decision Trees* por *dataset* durante a busca em grid

classificador bayesiano é adequado para classificação de *features* discretas. O parâmetro que foi levado em consideração na busca em grid para a escolha de valores por *dataset* que maximizam a acurácia do algoritmo Naive Bayes foi o parâmetro *alpha*. A tabela 4.11 lista os melhores valores encontrados.

	Parâmetro para Naive Bayes
Dataset	**alpha**
BoW	0.01
iFeel	1.0
Meta(k=5)	1.0
Meta(k=10)	0.01
Meta(k=20)	0.01
Meta(k=30)	0.01
BoW+iFeel	0.01
BoW+Meta(k=20)	0.01
iFeel+Meta(k=20)	1.0
BoW+iFeel+Meta(k=20)	1.0

Tabela 4.11: Melhores valores encontrados para o parâmetro alpha para o Naive Bayes por *dataset* durante a busca em grid

Para o SVM, usamos o classificador SVC [4] que é uma implementação baseada no LIBSVM [Chang & Lin, 2011]. O tempo de treino é mais que quadrático com o número de amostras o que faz com que seja difícil escalar para um *dataset* com mais de 10.000 amostras. O suporte a multiclasse é realizado seguindo o esquema one-vs-one. Os parâmetros que foram levados em consideração na busca em grid para a escolha de valores por *dataset* que maximizam a acurácia do SVM foram: *C, degree, kernel*. A tabela 4.12 lista os melhores valores encontrados.

[4]SVC: http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

	Parâmetros para SVM		
Dataset	**C**	**degree**	**kernel**
BoW	7	1	linear
iFeel	7	1	rbf
Meta(k=5)	7	1	linear
Meta(k=10)	3	1	linear
Meta(k=20)	1	1	linear
Meta(k=30)	3	1	linear
BoW+iFeel	7	1	linear
BoW+Meta(k=20)	1	1	linear
iFeel+Meta(k=20)	7	1	linear
BoW+iFeel+Meta(k=20)	1	1	linear

Tabela 4.12: Melhores valores de parâmetros encontrados para SVM por *dataset* durante a busca em grid

4.6 Avaliação

Para avaliar a capacidade preditiva dos classificadores durante os experimentos deste trabalho, usamos a métrica de acurácia [Witten & Frank, 2005; Baeza-Yates & Ribeiro-Neto, 2011]. Essa métrica representa a porcentagem de exemplos que foram rotulados corretamente pelo classificador. A fórmula para o cálculo é $Accuracy = TA/TC$, onde TA é o total de documentos corretamente classificados e TC o número total de documentos do dataset.

Como metodologia para obtenção da acurácia final de cada classificador, utilizamos o procedimento estatístico conhecido como validação cruzada de 10 partições[Witten & Frank, 2005]. Essa técnica consiste em dividir o *dataset* em dez partições. Onde, em seguida são realizadas dez rodadas de experimentos de forma que na i-ésima rodada, a i-ésima partição é usada como conjunto de teste e as demais são usadas como conjunto de treino. Deste modo, cada documento é avaliado apenas uma vez. A acurácia final do classificador é a média das dez rodadas.

Ao final, em todas as comparações entre as *baselines* e as alternativas de cada algoritmo, para termos a convicção de que a diferença observada nas acurácias não é consequência de uma escolha aleatória dos dados, aplicaremos o teste-T[Witten & Frank, 2005] pareado para estabeler a confiança estatística da diferença observada. Tal métrica é utilizada para verificar se há uma hipótese nula quando a estatística de teste segue uma distribuição t de *Student*. Em particular, para este trabalho, a hipótese nula é quando os métodos comparados tem desempenho similar (empate). O teste-T mede a probabilidade p de que a diferença observada supera a hipótese nula.

Neste trabalho, em todas as comparações nos algoritmos descritos, iremos identi-

ficar com um asterístico (*) quando o valor observado é significativo considerando um nível de 95%, ou seja, a chance da hipótese nula ser a correta é inferior a 5%. Também marcaremos como **negrito** os melhores resultados obtidos em cada algoritmo.

4.7 Resultados

Nesta seção apresentamos os resultados obtidos nos experimentos. Primeiro, mostramos a acurácia obtida por classificador nos modelos de *features* de forma separada para então avaliar a acurácia dos classificadores nos modelos combinados.

Os experimentos foram implementados em Python usando os algoritmos de classificação de Árvore de Decisão(DecisionTreeClassifier), Naive Bayes(MultinomialNB) e Support Vector Machines(SVC) oferecidos pela biblioteca scikit-learn [5] usando validação cruzada conforme estabelecido na seção sobre avaliação.

4.7.1 Modelos Separados

Inicialmente, nós comparamos a acurácia dos classificadores *Decision Trees*, Naive Bayes e SVM nos modelos de *features* BoW, iFeel e Meta-características. A tabela 4.13 mostra a acurácia obtida em cada modelo de *features*. Nos resultados é possível observar que o modelo *baseline* BoW teve um desempenho satisfatório somente no Naive Bayes e SVM, sendo bem pior somente nas *Decision Trees*. O modelo iFeel se mostrou ineficaz em todos os classificadores, tendo os piores resultados de acurácia. Os melhores resultados podem ser observados nos modelos de *meta-features*.

Em especial, os modelos Meta(k=5), Meta(k=20) e Meta(k=30) conseguiram os melhores scores de acurácia nos classificadores Naive Bayes, Árvores de Decisão e SVM. De fato, as meta-informações dos k vizinhos mais próximos de cada instância mostrou-se eficaz em descobrir características relevantes a partir dos dados. Um outro ponto importante é o espaço de *features* que cada modelo gerou. Apesar de conseguir ter somente 19 *features*, o modelo iFeel que usa 19 resultados de diversos algoritmos de detecção de polaridade não foi suficiente para discriminar as características entre as cinco classes de polaridade. O modelo baseline BoW apesar de obter resultados satisfatórios para o Naive Bayes e SVM, teve o maior espaço de *features* entre os modelos.

[5] scikit-learn: http://scikit-learn.org/stable/

		Acurácia		
Dataset	**Nro de Features**	**Decision Trees**	**Naive Bayes**	**SVM**
BoW (baseline)	3985	26.45 (b)	73.70 (b)	80.35 (b)
iFeel	19	44.0 ±5.99 *	28.65 ±2.31 *	31.80 ±2.48 *
Meta(k=5)	83	**80.8 ±5.19 ***	**75.7 ±5.93 ***	82.8 ±4.92
Meta(k=10)	133	77.85 ±4.91 *	73.85 ±5.69 *	83.65 ±4.11 *
Meta(k=20)	233	76.55 ±6.08 *	69.9 ±5.43 *	**85.3 ±4.35 ***
Meta(k=30)	333	72.55 ±6.88 *	64.35 ±5.02 *	**85.25 ±4.24 ***

Tabela 4.13: Resultados parcial das acurácias nos modelos separados e com os respectivos intervalos de confiança de 95% usando os diferentes datasets nos algoritmos Árvores de Decisão, Naive Bayes e SVM. Os resultados marcados com (b) são os valores de baseline. Os resultados estatísticamente significativos estão marcados com asterístico (*). Os melhores resultados obtidos em cada algoritmo estão marcados em **negrito**.

4.7.2 Modelos Combinados

Nesta seção, nós avaliamos se vale a pena combinar os diferentes modelos de *features* para melhorar o desempenho dos classificadores. Por esse motivo, elaboramos as combinações de *features* BoW+iFeel, e escolhemos apenas os modelos de meta-features que apresentaram o melhores resultados de acurácia nos três classificadores durante os experimentos em separado. Dessa forma, temos as combinações: BoW+Meta(k=5), BoW+Meta(k=20), iFeel+Meta(k=5), iFeel+Meta(k=20), BoW+iFeel+Meta(k=5) e BoW+iFeel+Meta(k=20). A tabela 4.14 mostra a acurácia obtida dos classificadores em cada combinação de *features*.

Nos resultados, é possível observar que houve uma melhora significativa apenas para os classificadores Naive Bayes através da combinação de BoW+Meta(k=5). As combinações envolvendo *features* iFeel mostraram em geral que esse conjunto de *features* gera impacto negativo nas acurácias dos classificadores. Na figura 4.1 é possível observar que o SVM teve o melhor desempenho em todos os modelos exceto no modelo iFeel.

		Acurácia		
Dataset	**Nro de Features**	**Decision Trees**	**Naive Bayes**	**SVM**
BoW (baseline)	3985	26.45 (b)	73.70 (b)	80.35 (b)
iFeel	19	44.0 ±5.99 *	28.65 ±2.31 *	31.80 ±2.48 *
Meta(k=5)	83	**80.8 ±5.19 ***	75.7 ±5.93 *	82.8 ±4.92
Meta(k=10)	133	77.85 ±4.91 *	73.85 ±5.69 *	83.65 ±4.11 *
Meta(k=20)	233	76.55 ±6.08 *	69.9 ±5.43 *	**85.3 ±4.35 ***
Meta(k=30)	333	72.55 ±6.88 *	64.35 ±5.02 *	**85.25 ±4.24 ***
BoW+iFeel	4004	33.6 ±3.17 *	72.55 ±3.66	79.20 ±4.57)
BoW+Meta(k=5)	4068	57.40 ±7.95 *	**81.35 ±4.46 ***	84.15 ±4.45 *
BoW+Meta(k=20)	4218	63.45 ±6.89 *	78.05 ±3.98 *	84.05 ±4.40 *
iFeel+Meta(k=5)	102	79.30 ±5.01 *	73.35 ±6.25	82.40 ±4.77
iFeel+Meta(k=20)	252	77.65 ±5.40 *	67.70 ±5.73 *	83.85 ±4.70 *
BoW+iFeel+Meta(k=5)	4087	56.35 ±8.11 *	79.7 ±4.42 *	83.20 ±4.79 *
BoW+iFeel+Meta(k=20)	4237	62.10 ±6.08 *	77.75 ±4.35 *	83.95 ±4.56 *

Tabela 4.14: Resultados final das acurácias dos modelos separados e combinados com os respectivos intervalos de confiança de 95% usando os diferentes datasets nos algoritmos Árvores de Decisão, Naive Bayes e SVM. Os resultados marcados com (b) são os valores de baseline. Os resultados estatísticamente significativos estão marcados com asterístico (*). Os melhores resultados obtidos em cada algoritmo estão marcados em **negrito**.

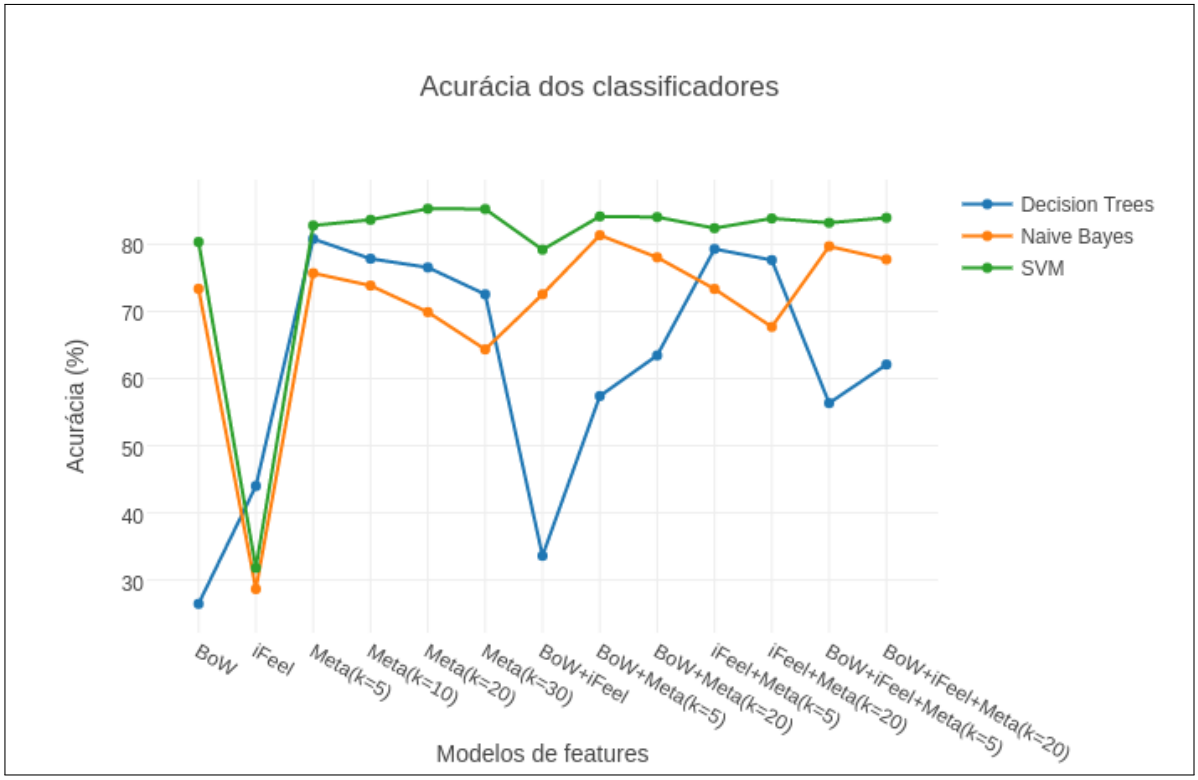

Figura 4.1: Comparativo entre as acurácias obtidas por cada classificador nos diferentes modelos de *features*

Capítulo 5

Conclusões

Neste capítulo, apresentamos as conclusões do nosso trabalho, o que inclui limitações da pesquisa e direções futuras que podem ser exploradas

5.1 Resultados Obtidos

Neste trabalho, investigamos como estabelecer um aprendizado supervisionado para identificar polaridades de cinco níveis. Em particular, avaliamos como diferentes modelos de extrações de features impactam na acurácia de cada classificador. Entre os métodos de extração de features que experimentados, consideramos o modelo bag-of-words, os resultados da ferramenta iFeel que usa 19 métodos da literatura e o método de extração de meta-informações.

Em geral, observamos que o modelo obtido a partir dos resultados da iFeel foi o que resultou no menor espaço de features, além disso, as acurácias alcançadas pelos classificadores usando esse modelo foram insatisfatórias e ficaram abaixo de 50%. Neste modelo, onde cada dimensão (feature) pode conter apenas um valor entre três possíveis (-1,0 ou 1), não foi eficiente em discriminar o modelo proposto de cinco polaridades.

Também observamos que o modelo de bag-of-words foi o que resultou no maior espaço de features e as acurácias obtidas foram boas em apenas dois classificadores (SVM e Naive Bayes), sendo bem pior no modelo de Decision Trees. De fato, Decision Trees são bem sensíveis a ruídos no conjunto de dados.

Além disso, observamos que o uso de meta-informações trouxeram os melhores resultados nos três classificadores tanto nos modelos separados quanto nos combinados. Em geral, os modelos de meta-informações obtiveram um número de features menor que o modelo de bag-of-words, porém com um melhor poder de captura de informações relevantes. Dessa forma, as meta-informações foram capazes de capturar particularida-

des dos cinco níveis de polaridade, no entanto, também observamos que à medida que k aumentava, houve um decaimento na acurácia principalmente nas Decision Trees e Naive Bayes.

O melhor escore de 85.30% de acurácia foi alcançado com o SVM fazendo uso do modelo de meta-informações extraídos a partir dos cinco vizinhos mais próximos de cada instância do *dataset*. As combinações envolvendo as *features* iFeel não representaram qualquer ganho de acurácia e apenas levaram a resultados inferiores à utilização dos modelos separados.

Desta forma, concluímos que o modelo mais promissor para a classificação dos tweets em cinco níves de polaridade proposto por este trabalho é resultante do treino do SVM junto com o modelo de meta-características, comprovando que as conclusões de [Canuto et al., 2016] para a classificação de três níveis de polaridade, também se aplicam ao contexto de cinco níveis de polaridade.

5.2 Limitações

A nossa coleção foi formada específicamente por tweets do contexto político. Embora as conclusões que alcançamos possam ser válidas para outros domínios, se faz necessário o estudo e avaliação em outras coleções, como *reviews* de produtos, filmes e etc.

As análises se concentraram em um número pequeno e pré-determinado de entidades, no qual usamos apenas três candidatos, enquanto que, em um cenário real onde pretende-se usar este método como ferramenta de avaliação de opinião, o número de entidades pode ser bem maior e necessita ser determinado de forma automática ou ser baseado em uma pré-seleção inicial maior.

Uma outra limitação em nossos experimentos foi que cada tweet de nossa coleção poderia expressar opinião sobre apenas um candidato enquanto que no cenário real, é muito comum encontrar tweets com opiniões sobre mais de um candidato.

5.3 Trabalhos Futuros

Como trabalhos futuros temos as seguintes propostas:

- Verificar as nossas conclusões em coleções de outros domínios como *reviews* de produtos e filmes.
- Estudar métodos de detecção de entidade como pessoas, organizações, localidades e produtos em textos de microblogs como o Twitter. A partir deste estudo,

estabelecer a técnica que irá realizar o reconhecimento prévio sobre entidades mencionadas nos tweets para então classificar as opiniões sobre tais entidades nos cinco níveis de polaridade.

- Investigar métodos para tratar tweets com mais de uma opinião pois em muitos casos, os usuários emitem em um único tweet, diferentes opiniões sobre as entidades. Por exemplo, no contexto político, podemos ter o seguinte tweet "A Marina tem boas idéias, mas prefiro Aécio e é nele que eu vou votar". Neste exemplo, podemos afirmar que existem duas opiniões sobre dois candidatos. A primeira opinião seria *positiva pontual* sobre a Marina e a segunda seria uma *aprovação completa* sobre o Aécio.

- Implementação do método de extração de meta-informações para a classificação dos cinco níveis de polaridade usando a GPU da mesma forma que [Canuto et al., 2016]. Desta forma, será possível extrapolar a quantidade de vizinhos avaliadas em nossos experimentos e também permitir a avaliação da nossa proposta de cinco níveis de polaridades em datasets com grande volumes de tweets.

Referências Bibliográficas

(2013). *Umigon: sentiment analysis for tweets based on lexicons and heuristics*. Zenodo.

Almatrafi, O.; Parack, S. & Chavan, B. (2015). Application of location-based sentiment analysis using twitter for identifying trends towards indian general elections 2014. Em *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, IMCOM '15, pp. 41:1--41:5, New York, NY, USA. ACM.

Araujo, M.; Reis, J.; Pereira, A. & Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. Em *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, SAC '16, pp. 1140--1145, New York, NY, USA. ACM.

Baccianella, S.; Esuli, A. & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. Em Chair), N. C. C.; Choukri, K.; Maegaard, B.; Mariani, J.; Odijk, J.; Piperidis, S.; Rosner, M. & Tapias, D., editores, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Baeza-Yates, R. & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition) (ACM Press Books)*. Addison-Wesley Professional, 2 edição. ISBN 0321416910.

Beasley, A. & Mason, W. (2015). Emotional states vs. emotional words in social media. Em *Proceedings of the ACM Web Science Conference*, WebSci '15, pp. 31:1--31:10, New York, NY, USA. ACM.

Bradley, M. M.; Lang, P. J.; Bradley, M. M. & Lang, P. J. (1999). Affective norms for english words (anew): Instruction manual and affective ratings.

Breiman, L.; Friedman, J. H.; Olshen, R. A. & Stone, C. J. (1984). *Classification and Regression Trees*. Chapman & Hall, New York, NY.

Cambria, E.; Speer, R.; Havasi, C. & Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. Em *AAAI Fall Symposium: Commonsense Knowledge*, volume FS-10-02 of *AAAI Technical Report*. AAAI.

Canuto, S.; Gonçalves, M. A. & Benevenuto, F. (2016). Exploiting new sentiment-based meta-level features for effective sentiment analysis. Em *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pp. 53--62, New York, NY, USA. ACM.

Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. Em *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pp. 161--168, New York, NY, USA. ACM.

Chang, C.-C. & Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1--27:27. ISSN 2157-6904.

Charalampakis, B.; Spathis, D.; Kouslis, E. & Kermanidis, K. (2015). Detecting irony on greek political tweets: A text mining approach. Em *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*, EANN '15, pp. 17:1--17:5, New York, NY, USA. ACM.

Cunha, E.; Magno, G.; Comarela, G.; Almeida, V.; Gonçalves, M. A. & Benevenuto, F. (2011). Analyzing the dynamic evolution of hashtags on twitter: A language-based approach. Em *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pp. 58--65, Stroudsburg, PA, USA. Association for Computational Linguistics.

De Smedt, T. & Daelemans, W. (2012). Pattern for python. *J. Mach. Learn. Res.*, 13(1):2063--2067. ISSN 1532-4435.

Dodds, P. S. & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441--456. ISSN 1573-7780.

Fellbaum, C., editor (1998). *WordNet: an electronic lexical database*. MIT Press.

García, S. & Herrera, F. (2009). Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. *Evol. Comput.*, 17(3):275--306. ISSN 1063-6560.

Gonçalves, P.; Benevenuto, F. & Cha, M. (2013). Panas-t: A pychometric scale for measuring sentiments on twitter. *CoRR*, abs/1308.1857.

Gonçalves, P.; Araújo, M.; Benevenuto, F. & Cha, M. (2013). Comparing and combining sentiment analysis methods. Em *Proceedings of the First ACM Conference on Online Social Networks*, COSN '13, pp. 27--38, New York, NY, USA. ACM.

González-Ibáñez, R.; Muresan, S. & Wacholder, N. (2011). Identifying sarcasm in twitter: A closer look. Em *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, pp. 581--586, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hannak, A.; Anderson, E.; Barrett, L. F.; Lehmann, S.; Mislove, A. & Riedewald, M. (2012). Tweetin'in the rain: Exploring societal-scale effects of weather on mood. Em *ICWSM*.

Heerschop, B.; Goossen, F.; Hogenboom, A.; Frasincar, F.; Kaymak, U. & de Jong, F. (2011). Polarity analysis of texts using discourse structure. Em *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pp. 1061--1070, New York, NY, USA. ACM.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews. Em *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pp. 168--177, New York, NY, USA. ACM.

Hutto, C. J. & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Em *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014*.

Jungherr, A. (2013). Tweets and votes, a special relationship: The 2009 federal election in germany. Em *Proceedings of the 2Nd Workshop on Politics, Elections and Data*, PLEAD '13, pp. 5--14, New York, NY, USA. ACM.

Krieger, M. & Ahn, D. (2010). Tweetmotif: exploratory search and topic summarization for twitter. Em *In Proc. of AAAI Conference on Weblogs and Social*.

Makrehchi, M. (2014). The correlation between language shift and social conflicts in polarized social media. Em *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies*

(IAT) - Volume 02, WI-IAT '14, pp. 166--171, Washington, DC, USA. IEEE Computer Society.

Mann, W. C. & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243--281.

Manning, C. D.; Raghavan, P. & Schütze, H. (2008). Text classification and naive bayes. Em *Introduction to Information Retrieval:*, pp. 234–265. Cambridge University Press, Cambridge.

Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D. & Miller, K. (1990). Introduction to WordNet: An online lexical database. *International Journal of Lexicography*, 3(4):235--244.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39--41. ISSN 0001-0782.

Mohammad, S. M. (2012). #emotional tweets. Em *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, SemEval '12, pp. 246--255, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mohammad, S. M.; Kiritchenko, S. & Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. Em *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA.

Mohammad, S. M. & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. 29(3):436--465.

Mukherjee, S.; Malu, A.; A.R., B. & Bhattacharyya, P. (2012). Twisent: A multistage system for analyzing sentiment in twitter. Em *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pp. 2531--2534, New York, NY, USA. ACM.

Nielsen, F. Å. (2011). A new ANEW: evaluation of a word list for sentiment analysis in microblogs. Em Rowe, M.; Stankovic, M.; Dadzie, A.-S. & Hardey, M., editores, *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, volume 718 of *CEUR Workshop Proceedings*, pp. 93--98.

Pappas, N. & Popescu-Belis, A. (2013). Sentiment analysis of user comments for one-class collaborative filtering over ted talks. Em *36th ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.

Park, J.; Barash, V.; Fink, C. & Cha, M. (2013). Emoticon Style: Interpreting Differences in Emoticons Across Cultures. *The 7th International AAAI Conference on Weblogs and Social Media*.

Pla, F. & Hurtado, L. F. (2014). Political tendency identification in twitter using sentiment analysis techniques. Em *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pp. 183--192.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of emotion*, 1:3--31.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 1-55860-238-0.

Ribeiro, F. N.; Araújo, M.; Gonçalves, P.; Benevenuto, F. & Gonçalves, M. A. (2015). A benchmark comparison of state-of-the-practice sentiment analysis methods. *CoRR*, abs/1512.01818.

Schinas, E.; Papadopoulos, S.; Diplaris, S.; Kompatsiaris, Y.; Mass, Y.; Herzig, J. & Boudakidis, L. (2013). Eventsense: Capturing the pulse of large-scale events by mining social media streams. Em *Proceedings of the 17th Panhellenic Conference on Informatics*, PCI '13, pp. 17--24, New York, NY, USA. ACM.

Sedhai, S. & Sun, A. (2016). Effect of spam on hashtag recommendation for tweets. Em *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pp. 97--98, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D.; Ng, A. Y. & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. Em *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631--1642, Stroudsburg, PA. Association for Computational Linguistics.

Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K. & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Comput. Linguist.*, 37(2):267--307. ISSN 0891-2017.

Thelwall, M. (2013). Heart and soul: Sentiment strength detection in the social web with sentistrength 1.

Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F. & Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle. Em *Proceedings of the ACL 2012 System Demonstrations*, ACL '12, pp. 115--120, Stroudsburg, PA, USA. Association for Computational Linguistics.

Watson, D.; Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of Personality and Social Psychology*, 54:1063–1070.

Wilson, T.; Hoffmann, P.; Somasundaran, S.; Kessler, J.; Wiebe, J.; Choi, Y.; Cardie, C.; Riloff, E. & Patwardhan, S. (2005). Opinionfinder: A system for subjectivity analysis. Em *Proceedings of HLT/EMNLP on Interactive Demonstrations*, HLT-Demo '05, pp. 34--35, Stroudsburg, PA, USA. Association for Computational Linguistics.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 0120884070.