

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

CÁSSIO CASTALDI ARAUJO BLAZ

**Análise de Sentimentos em Tíquetes para o
Suporte de TI**

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

Orientador: Prof^a. Dr^a. Karin Becker

Porto Alegre
2017

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Blaz, Cássio Castaldi Araujo

Análise de Sentimentos em Tíquetes para o Suporte de TI / Cássio Castaldi Araujo Blaz. – Porto Alegre: PPGC da UFRGS, 2017.

96 f.: il.

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

1. Análise de Sentimentos. 2. Mineração de Opinião. 3. Tíquetes para TI. 4. Dicionário de Domínio. I. Becker, Karin. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof^a. Jane Fraga Tutikian

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

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. João Luiz Dihl Comba

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

*“For myself, losing is not coming second.
It’s getting out of the water knowing you could have done better.
For myself, I have won every race I’ve been in.”*

— IAN THORPE

AGRADECIMENTOS

Agradeço a todos que, direta ou indiretamente, contribuíram e me ajudaram a realizar e a concluir este mestrado. Em especial:

À professora Karin Becker, por me orientar ao longo de todo o trabalho. Como aluno, não teria conhecimento suficiente para concluí-lo sozinho, mas agradeço principalmente por direcionar o desenvolvimento conforme meu perfil e por me auxiliar e me incentivar a superar minhas dificuldades. Saliento o ótimo relacionamento interpessoal e agradeço por compreender a questão de saúde que enfrentei ao longo do curso.

A Daniela de Almeida, ao Paulo Krieser, ao Gabriel Moser e ao Centro Nacional de Supercomputação (CESUP), por cederem as informações que geraram a base de dados que foram o cerne deste trabalho.

A Maria Evelise Blaz, ao Rafael Blaz e a Daisy Iris Castaldi, por anotarem parte da base de dados, o que serviu para avaliar os resultados obtidos nos experimentos deste trabalho.

A Denise Ewald, por permitir, e ao Glaucio Souza, por me substituir, durante minhas eventuais ausências no CESUP, necessárias para a participação das aulas e das reuniões do mestrado.

RESUMO

Análise de Sentimentos/Mineração de Opinião é adotada na engenharia de software para questões como usabilidade e sentimentos de desenvolvedores em projetos. Este trabalho propõe métodos para avaliar os sentimentos presentes em tíquetes abertos à área de suporte de TI.

Há diversos tipos de tíquetes abertos à TI (e.g. infraestrutura, software), que envolvem erros, incidentes, requisições, etc. O maior desafio é automaticamente distinguir entre a necessidade em si, a qual é intrinsecamente negativa (por exemplo, a descrição de um erro), de um sentimento embutido na descrição.

Nossa abordagem automaticamente cria um dicionário de domínio que contém termos que expressam sentimentos no contexto de TI, utilizados para filtrar expressões em um tíquete para análise de sentimentos. Nós criamos e avaliamos três métodos de classificação para calcular a polaridade em tíquetes. Nosso estudo utilizou 34.895 tíquetes de cinco organizações.

Para polaridade, 2.333 tíquetes foram selecionados aleatoriamente para compor nosso *gold standard*. Nossos melhores resultados apresentam uma precisão e revocação de 82,83% e 88,42%, respectivamente, o que supera outras soluções de análise de sentimentos comparadas.

De forma complementar, emoções em tíquetes foram estudadas considerando os modelos de Ekman e VAD. Um dos três métodos de classificação criados foi adaptado para também identificar emoções nos tíquetes. Possíveis correlações entre polaridade e emoções foram verificadas via regras de associação. Resultados correlacionam tíquetes positivos com valência e dominância altas e excitação baixa, além de presença de alegria e surpresa e ausência de medo. Tíquetes negativos correlacionam com valência, excitação e dominância neutras, além de ausência de alegria e presença de medo. Contudo os resultados para a polaridade negativa não são precisos.

Palavras-chave: Análise de Sentimentos. Mineração de Opinião. Tíquetes para TI. Dicionário de Domínio.

Sentiment Analysis in Tickets for IT Support

ABSTRACT

Sentiment Analysis/Opinion Mining has been adopted in software engineering for problems such as software usability and sentiment of developers in projects. This work proposes methods to evaluate the sentiment contained in tickets for IT (Information Technology) support.

IT tickets are broad in coverage (e.g. infrastructure, software), and involve errors, incidents, requests, etc. The main challenge is to automatically distinguish between factual information, which is intrinsically negative (e.g. error description), from the sentiment embedded in the description.

Our approach is to automatically create a domain dictionary that contains terms with sentiment in IT context, used to filter terms in tickets for sentiment analysis. We created and evaluate three classification methods for calculating the polarity of terms in tickets. Our study was developed using 34,895 tickets from five organizations.

For polarity, we randomly selected 2,333 tickets to compose a gold standard. Our best results display an average precision and recall of 82.83% and 88.42%, respectively, which outperforms the compared sentiment analysis solutions.

Complementarily, emotions in tickets were studied considering the models of Ekman and VAD. One of the three classification methods created has been adapted to also identify emotions in the tickets. Possible correlations between polarity and emotions were verified through association rules. Results correlate positive tickets with valence and dominance high and low excitement, besides presence of joy and surprise and absence of fear. Negative tickets correlate with valence, neutral excitement and dominance, besides absence of joy and presence of fear. However the results for negative polarity are not accurate.

Keywords: Sentiment Analysis. Opinion Mining. IT Tickets. Domain Dictionary.

LISTA DE FIGURAS

Figura 4.1 Exemplos de Tíquetes e de suas Estruturas	37
Figura 4.2 Visão Geral da Abordagem Proposta.....	38
Figura 4.3 Exemplo de Classificação de Sentimentos dos Métodos de Dicionário, de Estrutura e Híbrido.....	40
Figura 4.4 Exemplo de Expansão de Semente	43
Figura 4.5 Exemplo de Expansão de Semente Composta de Palavras Expansíveis	43
Figura 4.6 Exemplo de Pré-Processamento de Tíquete	45
Figura 4.7 Exemplo de Classificação de Sentimentos do Método de Dicionário	48
Figura 4.8 Exemplo de Classificação de Sentimentos do Método de Estrutura.....	50
Figura 4.9 Exemplo de Classificação de Sentimentos do Método Híbrido.....	52
Figura 4.10 Tela Principal da POC	56
Figura 4.11 Cadastro de Fontes e Tíquetes na POC.....	57
Figura 4.12 Métodos de Classificação para Execução na POC	58
Figura 4.13 Manipulação do Dicionário de Domínio na POC.....	59
Figura 4.14 Sumarização e Filtros na POC.....	60
Figura 4.15 Visualização de Tíquetes Filtrados	62
Figura 4.16 Visualização do Texto de Tíquete Selecionado	62
Figura 4.17 Visualização dos Tokens de um Tíquete.....	63
Figura 4.18 Nuvens de Tokens na POC	64
Figura 4.19 Gráficos com Estatísticas de Sentimentos na POC.....	65
Figura 5.1 Diferenças de Pontuações de Polaridade entre Traduções Manual e Automática..	78
Figura 5.2 Emoções de Ekman nos Tíquetes Positivos do Gold Standard	83
Figura 5.3 Emoções VAD nos Tíquetes Positivos do Gold Standard	84
Figura 5.4 Emoções de Ekman nos Tíquetes Negativos da POC.....	86
Figura 5.5 Emoções VAD dos Tíquetes Negativos da POC.....	87

LISTA DE TABELAS

Tabela 2.1	Matriz de Confusão	23
Tabela 3.1	Trabalhos Relacionados.....	33
Tabela 4.1	Descrição das Fontes dos Tíquetes.....	34
Tabela 4.2	Tipos de Tíquetes e Exemplos.....	35
Tabela 4.3	Exemplo de Cálculos da Polaridade Padrão	47
Tabela 4.4	Exemplo de Cálculos da Polaridades de Categoria	50
Tabela 5.1	Exemplos de Tíquetes Positivos para o Tutorial de Anotação	68
Tabela 5.2	Divergências entre as Anotações dos Anotadores	69
Tabela 5.3	Sementes do Dicionário de Domínio.....	71
Tabela 5.4	Comparação entre os Baselines e os Métodos Propostos.....	73
Tabela 5.5	Comparação entre o Método de Dicionário, o Método de Estrutra e o Método Híbrido	76
Tabela 5.6	Comparação entre as Medidas Média e Mediana.....	76
Tabela 5.7	Comparação entre a Aplicação e a Não Aplicação do Espelhamento.....	77
Tabela 5.8	Exemplos de Regras Redundantes.....	79
Tabela 5.9	Correlações entre Polaridade e Emoções em Tíquetes.....	80

LISTA DE ABREVIATURAS E SIGLAS

ANEW	Affective Norms for English Words
BLEU	Bilingual Evaluation Understudy
DD	Dicionário de Domínio
LDA	Latent Dirichlet Allocation
NRC	National Research Council Canada
PLN	Processamento de Linguagem Natural
POS	Part of Speech
TI	Tecnologia da Informação
VAD	Valence-Arousal-Dominance

SUMÁRIO

1 INTRODUÇÃO	12
2 FUNDAMENTAÇÃO TEÓRICA	17
2.1 Sentimentos, Polaridade e Emoções	17
2.2 Análise de Sentimentos	18
2.3 Processamento de Linguagem Natural para Análise de Sentimentos.....	18
2.4 Classificação de Sentimentos.....	19
2.5 Dicionários de Sentimentos	20
2.6 Análise de Sentimentos Multilíngue	22
2.7 Métricas para Avaliação	23
2.8 Regras de Associação	24
3 TRABALHOS RELACIONADOS	26
3.1 Análise de Sentimento em Artefatos de TI	26
3.2 Considerações Finais	30
4 ABORDAGEM PROPOSTA	34
4.1 Características dos Tíquetes	34
4.2 Visão Geral da Abordagem para Análise de Sentimento	37
4.3 Dicionário de Domínio.....	40
4.3.1 Pré-processamento de Sementes	41
4.3.2 Expansão	42
4.3.3 Poda.....	44
4.4 Análise de Polaridade	44
4.4.1 Pré-Processamento	44
4.4.2 Métodos para Classificação de Polaridade.....	45
4.4.2.1 Método de Dicionário	45
4.4.2.2 Método de Estrutura.....	48
4.4.2.3 Método Híbrido.....	51
4.4.3 Agregação das Pontuações de Polaridade.....	52
4.5 Análise de Emoções.....	52
4.5.1 Classificação de Emoções	53
4.5.2 Agregação de Emoções.....	54
4.6 Prova de Conceitos.....	54
4.6.1 Funcionalidades de Cadastro e Importação de Tíquetes.....	57
4.6.2 Manipulação dos Dicionários	58
4.6.3 Método de Classificação	59
4.6.4 Sumarizações e Filtros	61
4.6.5 Visualizações.....	62
4.7 Considerações Finais	66
5 EXPERIMENTOS	67
5.1 Gold Standard para Polaridade	67
5.2 Configuração dos Experimentos para Polaridade	69
5.2.1 Objetivos	69
5.2.2 Limite de Iterações no Processo de Expansão do Dicionário de Domínio	70
5.2.3 Baselines para Polaridade	70
5.2.4 Implementação dos Métodos de Classificação de Polaridade.....	72
5.3 Polaridade: Resultados e Discussões	72
5.3.1 Comparação com os Baselines.....	72
5.3.2 Comparação entre Métodos	74
5.3.3 Influência da Tradução.....	77

5.4 Configuração dos Experimentos para Emoção	78
5.4.1 Objetivos	78
5.4.2 Geração de Regras de Associação entre Polaridade e Emoções	78
5.5 Emoção: Resultados e Discussões.....	80
5.5.1 Correlação entre Polaridade e Emoções	80
5.5.2 Análise das Relações entre Polaridade e Emoções na POC	82
5.5.3 Discussão	88
6 CONCLUSÃO E TRABALHOS FUTUROS	89
REFERÊNCIAS.....	92

1 INTRODUÇÃO

Tecnologia da Informação (TI) é um assunto amplo que envolve diferentes aspectos de processamento, análise e gerência de informações em organizações, como computadores, software, redes. A utilização de softwares via computadores pessoais é crucial em todos os níveis, do suporte às atividades cotidianas no nível operacional, ao acesso a dados cruciais para tomada de decisões estratégicas. Usuários necessitam desde o simples uso emails, até a interação com softwares complexos e customizados que requerem infraestruturas próprias. As organizações podem prover equipes internas de TI ou delegar essa responsabilidade a empresas externas.

O atendimento de suporte a usuários é um ponto crucial à área de TI. Defeitos em sistemas e problemas de infraestrutura prejudicam o trabalho de usuários, muitas vezes inviabilizando suas atividades. Portanto, devem ser resolvidos no menor tempo possível. A satisfação de clientes também pode ser medida via a qualidade do suporte de uma empresa ou área de tecnologia. Frameworks como o Information Technology Infrastructure Library (ITIL) e o Control Objectives for Information and Related Technologies (COBIT) (SAHIBUDIN; SHARIFI; AYAT, 2008) são utilizados para gestão de TI e sugerem critérios objetivos para medir a qualidade do serviço provido e para priorizar e atender chamados. Por exemplo, ITIL recomenda a priorização de tíquetes baseado na importância e no impacto. Já COBIT considera o tempo de entrega/atendimento baseado em um acordo de nível de serviço.

As empresas gerenciam as demandas de muitas fontes, como email, telefonemas, reuniões ou sistemas de suporte. Neste trabalho, nós denominaremos "tíquete" como qualquer solicitação de um usuário cadastrada/enviada via sistema e direcionada ao time de TI.

Quando usuários relatam um problema ou solicitam uma requisição, eles podem expressar sentimentos que provêm informações adicionais nesse contexto. Tíquetes contêm uma descrição textual de um problema ou de uma requisição, possivelmente com informações adicionais estruturadas, como data, prioridade, impacto, sistema, etc. Sentimentos podem revelar o estado emocional do usuário em relação ao pedido ("Eu preciso urgentemente 6 notebooks para a próxima sexta-feira!!!") ou podem indicar um prévio *feedback* sobre a qualidade dos serviços prestados pela TI ("Eu estou sem acesso ao meu e-mail novamente. É a terceira vez que isso acontece nessa semana!"). A consideração e avaliação deste sentimento criam uma oportunidade única para analisar e manter a satisfação dos clientes antes mesmo do atendimento do chamado.

A análise de sentimentos envolve a identificação automática de opiniões, sentimentos,

avaliações, atitudes e emoções de pessoas, expressas em linguagem escrita (LIU, 2012). A maioria dos trabalhos em análise de sentimentos foca em identificar uma opinião medida em termos de um valor de polaridade, que varia de negativo a positivo. Outros estudos trabalham com emoções, no entanto não há consenso sobre qual modelo deve ser considerado para identificar e mensurar de emoções (LIU, 2012). Um dos modelos mais populares é o de Ekman, que considera a existência de seis emoções básicas: surpresa (surprise), alegria (joy), tristeza (sadness), raiva (anger), medo (fear) e nojo (disgust) (EKMAN, 1992). Já o modelo Valence-Arousal-Dominance (VAD) classifica emoções conforme três dimensões ortogonais: valência (quanto um estímulo é positivo ou negativo, em termos de prazer), excitação (intensidade da emoção provocada por um estímulo) e dominância (o grau de controle exercido pela emoção em uma pessoa) (WARRINER; KUPERMAN; BRYLSBAERT, 2013).

Para a identificação/classificação dos sentimentos, as abordagens baseadas em léxicos são muito populares (TSYTSARAU; PALPANAS, 2012; MOHAMMAD; TURNEY, 2010; WARRINER; KUPERMAN; BRYLSBAERT, 2013) e precisam de um dicionário relacionando termos a alguma medida de sentimento ou categorização de uma ou mais emoções. Já a abordagem de aprendizado de máquina envolve um corpus anotado para treinamento em algoritmos de classificação/regressão. Aplicações clássicas são a identificação de opiniões em revisões de produtos e de serviços (LIU, 2012), notícias (BECKER; MOREIRA; SANTOS, 2017), emails (MOHAMMAD, 2012), gerenciamento de marcas (BALAHUR; PEREA-ORTEGA, 2014; GODBOLE; SRINIVASAIAH; SKIENA, 2007; MOHAMMAD, 2016) e previsões baseadas em sentimentos (BOLLEN; MAO; ZENG, 2011), (TUMITAN; BECKER, 2014).

A importância de análise de sentimentos também é reconhecida no campo da engenharia de software. Sentimentos são estudados em postagens e interações entre desenvolvedores em tíquetes de sistemas para cadastros de erros (MURGIA et al., 2014; PLETEA; VASILESCU; SEREBRENIK, 2014; JURADO; RODRIGUEZ, 2015; ORTU et al., 2015; ORTU et al., 2016; ISLAM; ZIBRAN, 2017; GACHECHILADZE et al., 2017), em *commits* de arquivos submetidos em repositórios/sistemas de versionamento (GUZMAN; AZÓCAR; LI, 2014; SINHA; LAZAR; SHARIF, 2016) e em artefatos variados de projetos (e.g. emails, Twitter, anexos, descrições de demandas, relatórios de usabilidade, etc.) (GUZMAN; BRUEGGE, 2013; EL-HALEES, 2014). A mineração de opinião em tíquetes criados por usuários ao suporte de TI é um cenário que ainda não foi explorado. A maioria dos trabalhos emprega soluções genéricas de análises de sentimentos, que produzem resultados diferentes e conflitantes (JONGELING; SEREBRENIK, 2015), possivelmente por não considerarem a forma como um sentimento é transmitido em um contexto de engenharia de software.

O principal desafio em analisar sentimentos em tíquetes é separar o objeto da requisição, o qual envolve relatar algum problema ou necessidade, de sentimentos incorporados ao texto ao relatá-lo. Jargões técnicos utilizados para descrever um tíquete (e.g. “problema”, “erro”, “defeito”) são intrinsecamente negativos, mas não necessariamente expressam algum sentimento. Nesse contexto, sentimentos são transmitidos por saudações (“bom dia”), observações de conclusão (“obrigado”, “aguardo retorno brevemente”), expressões que denotam intensidade (“novamente”, “urgentemente”), entre outras. Um desafio similar foi identificado em notícias, ao distinguir notícias ruins de opiniões negativas (BALAHUR; PEREA-ORTEGA, 2014). Por exemplo, considere o tíquete “Bom dia, favor verificar por que meu computador está lento. Obrigado.”. Apesar do usuário relatar um problema de performance, as expressões “bom dia”, “favor” e “obrigado” demonstram sentimentos positivos de cortesia. A mesma questão poderia ser escrita como “Verificar por que meu computador está lento” (neutro), ou “Mais uma vez eu peço para verificarem por que meu computador está extremamente lento. Aguardo retorno.” (negativo).

Assim o estudo de sentimentos de usuários em relação a sistemas, a infraestrutura e a área de TI como um todo é um problema atual que carece de pesquisa. A criação e utilização de métodos para identificação de polaridade e de emoções, bem como suas possíveis correlações, provêm um *feedback* antecipado a respeito da qualidade de serviços de TI prestados. Esse *feedback* permite que sejam analisadas mudanças recentes em sistemas, infraestrutura, na equipe e na área de TI como um todo. Assim, ações proativas podem ser realizadas para melhorar a qualidade do serviço de TI de forma breve. A identificação de sentimentos nos tíquetes também permite uma nova forma de classificação de atendimento destes. Por exemplo, um usuário com sentimentos negativos pode ser um dos critérios para priorização, pois pode indicar uma pessoa estressada, impaciente ou insatisfeita, a qual carece de uma maior atenção.

O objetivo deste estudo é desenvolver métodos para análise de sentimentos, considerando polaridade e emoções, focados em tíquetes para a TI. Por definição, soluções genéricas de análise de sentimentos não consideram a distinção do objeto do relato de um problema de possíveis sentimentos, portanto não produzem resultados consistentes (JONGELING; SEREBRENİK, 2015). Esta pesquisa busca uma melhor compreensão de como sentimentos são expressos na comunicação entre usuários e a equipe de TI, além de como a polaridade e emoções se relacionam nesse contexto.

Trabalhos anteriores focaram em textos escritos por pessoas com conhecimentos técnicos (e.g. desenvolvedores (JURADO; RODRIGUEZ, 2015; MURGIA et al., 2014; ORTU et al., 2015), testadores (EL-HALEES, 2014)). Completamos essas tentativas ao abordar como sen-

timentos são expressos por usuários leigos em áreas além da TI. Desenvolvemos uma solução que aborda os desafios inerentes aos tíquetes como forma de produzir resultados mais precisos. Também contribuímos abordando os tíquetes de diferentes tipos (e.g. problemas de infraestrutura, requisições para extração de dados, necessidades administrativas, *feedbacks*), coletados de cinco diferentes organizações, resultando em uma considerável generalização em nossas descobertas dentro do contexto de TI, enquanto que outros trabalhos de análise de sentimentos no domínio da engenharia de software focam apenas em sistemas/aplicativos/software.

Neste trabalho são propostas técnicas para avaliar sentimentos presentes em tíquetes, classificando-os conforme a polaridade, as seis emoções de Ekman e os três componentes VAD. Nossa abordagem inclui:

- a) a criação e expansão automática de um dicionário de domínio (DD) que contém palavras e expressões com polaridade e/ou emoções no domínio de TI;
- b) três estratégias diferentes para análise de polaridade em tíquetes, sendo uma léxica, uma estrutural e outra híbrida. As estratégias variam na forma como atribuem pontuações para palavras e para expressões de tíquetes, agregando as pontuações individuais para calcular a polaridade geral de um tíquete;
- c) uma estratégia léxica para análise de emoções, a qual identifica palavras que estão presentes nos tíquetes e que possuem as emoções segundo dois modelos (de Ekman e/ou de VAD), agregando os valores individuais para estabelecer os totais de cada tíquete;
- d) a verificação de possíveis correlações entre polaridade e os dois modelos de emoções nesse contexto;
- e) Uma prova de conceito que mostra como explorar a análise de diferentes sentimentos no contexto de tíquetes, compilando os resultados dos experimentos.

Nosso estudo foi desenvolvido utilizando 34.895 tíquetes de cinco organizações diferentes, os quais englobam diferentes tipos de solicitações (defeitos, incidentes, requisições, sugestões, etc.). Para avaliar o desempenho da solução proposta, nós criamos um *gold standard* anotado para polaridade, composto de 2.333 tíquetes. Nossos melhores resultados na análise de polaridade apresentam precisão e revocação de 82.83% e de 88.42%, respectivamente, o que supera em muito as soluções de análise de sentimento comparadas. Sobre emoções, a aplicação de regras de associação (AGRAWAL; IMIELNISKI; SWAMI, 1993) mostra correlação entre tíquetes negativos com valência, excitação e dominância neutras, além da ausência de alegria e a presença de medo. Já os tíquetes positivos estão relacionados com valência e dominância

altas e excitação baixa, bem como a presença de alegria e de surpresa e a ausência de medo. Resultados preliminares deste trabalho são relatados em (BLAZ; BECKER, 2016).

O restante deste trabalho é estruturado como segue: o Capítulo 2 apresenta a fundamentação teórica necessária para entender este estudo. O Capítulo 3 descreve trabalhos relacionados. O Capítulo 4 apresenta e detalha a abordagem e as técnicas propostas, também apresentando a aplicação desenvolvida como prova de conceito. O Capítulo 5 provê descrições detalhadas dos experimentos realizados e seus resultados. Finalizando, o Capítulo 6 apresenta as conclusões e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo descreve conceitos e definições necessários para a compreensão deste trabalho.

2.1 Sentimentos, Polaridade e Emoções

Sentimento e outros conceitos relacionados, como emoção, avaliação, postura e humor, estão vinculados a crenças e sensações das pessoas (LIU, 2015). Não existe consenso na área de mineração de opinião sobre uma forma padrão para medir sentimentos.

Uma forma popular é a polaridade. A polaridade pode ser atribuída de forma discreta, como as categorias positiva, neutra ou negativa, ou corresponder a uma pontuação (por exemplo, -0,7) em uma escala que representa a avaliação em uma dessas três categorias (TSYTSARAU; PALPANAS, 2012).

Já emoções são classificadas conforme outros modelos, sem que haja um consenso entre os autores (LIU, 2012). Diferentes teorias propõem quais seriam as emoções básicas (EKMAN, 1992; PLUTCHIK, 1991; PARROTT, 2001). Ekman (EKMAN, 1992) sugere que são seis: surpresa (surprise), alegria (joy), tristeza (sadness), raiva (anger), medo (fear) e nojo (disgust). Já Plutchik (PLUTCHIK, 1991) apresenta duas adicionais: confiança (trust) e expectativa (anticipation). Parrott (PARROTT, 2001) sugere três níveis de emoções, sendo as seis do primeiro nível as seguintes: amor (love), alegria (joy), surpresa (surprise), raiva (anger), tristeza (sadness), e medo (fear). Já o modelo VAD caracteriza emoções sobre três componentes: valência, excitação e dominância (WARRINER; KUPERMAN; BRYSSBAERT, 2013). Valência corresponde ao quanto agradável um estímulo é, variando de infeliz a feliz. Excitação é relacionada à intensidade. Dominância denota o poder de um estímulo, variando de fraco/submisso a forte/dominante. Cada estímulo tem uma pontuação atribuída para cada um dos componentes, a qual varia de 0 (baixo) a 10 (alto).

Em nosso trabalho, optamos por estudar sentimento na forma de pontuações de polaridade e de emoções. Para emoções, consideramos tanto os seis tipos básicos de Ekman quanto o modelo VAD.

2.2 Análise de Sentimentos

Análise de sentimentos é uma área que estuda opiniões, sentimentos, avaliações, atitudes e emoções direcionadas a entidades, como produtos, serviços, organizações, indivíduos, problemas, eventos, tópicos em relação aos seus atributos e as suas características (LIU, 2012).

Um das principais etapas no processo de análise de sentimento é a classificação (BECKER; TUMITAN, 2013), na qual são identificados sentimentos presentes intrinsecamente em um documento. O sentimento pode ser identificado no documento como um todo, individualmente para cada sentença, ou em nível de aspecto. Antes dessa classificação, deve ser realizado um pré-processamento, no qual o texto original sofre transformações via técnicas de processamento de linguagem natural (PLN) necessárias à classificação de sentimentos em si. Para classificação, existem diferentes abordagens, como a léxica, a de aprendizado de máquina e a estatística (TSYTSARAU; PALPANAS, 2012). Maiores detalhes sobre PLN e técnicas de classificação são dadas nas seções 2.3 e 2.4, respectivamente.

O objetivo deste trabalho é analisar tíquetes escritos por usuários em sistemas de suporte à área de TI. Tíquetes apresentam variadas entidades (sistemas, funcionalidades, computadores, pessoas) e cada um é analisado como um todo. Portanto adotou-se o nível de documento para este trabalho.

2.3 Processamento de Linguagem Natural para Análise de Sentimentos

O PLN propõe diversas técnicas para manipulação e transformação de documentos, as quais são essenciais ao pré-processamento de textos para classificação de sentimentos na mineração de opiniões. O PLN está relacionado à compreensão da linguagem, oral ou escrita, e pode ser dar em diferentes níveis (JURAFSKY; MARTIN, 2000), sendo os principais:

- Morfológico: conhecimento das construções e dos componentes de palavras;
- Sintático: conhecimento das relações estruturais entre palavras;
- Semântico: conhecimento do significado de palavras em sentenças.

As principais técnicas realizadas neste trabalho são: tokenização, lematização e etiquetagem morfosintática.

Tokenização é uma técnica para identificação dos termos (*tokens*) de um texto via remoção de caracteres separadores (espaços em branco, pontuações, quebras de linha). É utilizada

neste trabalho para listar os termos/expressões presentes nos textos dos tíquetes.

Lematização consiste em deflexionar uma palavra, reduzindo-a a sua forma essencial. Por exemplo: “errar”, “erroneamente” e “errado” derivam do lema “erro”. As palavras presentes nos dicionários de sentimentos existentes e o dicionário de domínio (DD) criado em nosso estudo estão em suas formas lematizadas.

A etiquetagem morfosintática (part-of-speech ou POS) é o processo de identificar a classe gramatical de um termo, baseado em sua definição e no contexto no qual se encontra. Neste trabalho, foram consideradas as classes substantivo, adjetivo, advérbio e verbo.

Essas três técnicas pertencem ao nível morfológico.

2.4 Classificação de Sentimentos

As abordagens mais populares para a classificação do sentimento são a léxica e a de aprendizado de máquinas supervisionado (TSYTSARAU; PALPANAS, 2012).

A primeira é totalmente não supervisionada e requer léxicos (ou dicionários) nos quais os termos estão associados a um rótulo ou a uma pontuação que denote intensidade. Assim, para analisar quais sentimentos uma palavra ou expressão apresentam, deve ser verificada se a mesma está presente no dicionário. Uma função de agregação calcula estas pontuações para o conjunto de palavras contidas no texto.

Já aprendizado de máquina envolve o treinamento sobre um corpus anotado. Essa abordagem busca identificar regras gerais em grandes conjuntos de dados, de modo que informações implícitas existentes possam ser extraídas de forma automática, compondo um modelo. Em estudos de análise de sentimentos, há um predomínio de métodos supervisionados de aprendizagem de classificação e de regressão (TAN; STEINBACH; KUMAR, 2005). Essas regras também são utilizadas para identificar estruturas em sentenças associadas a sentimentos (PRABOWO; THELWALL, 2009; WU; CHUANG; LIN, 2006). O sucesso desta abordagem é vinculado ao tamanho do corpus anotado, bem como sua vinculação com o domínio.

Outra abordagem não tão popular é a estatística, também não supervisionada (TSYTSARAU; PALPANAS, 2012; BECKER; TUMITAN, 2013). Baseia-se na proximidade de termos, ou seja, a polaridade de um termo pode ser calculada a partir da coocorrência com outros, cujas polaridades são conhecidas. Um modelo estatístico é o Latent Dirichlet allocation (LDA) (BLEI; NG; JORDAN, 2003), que apresenta uma hierarquia de três níveis do método de Bayesian. É muito utilizada para a extração de tópicos em textos.

Decidimos utilizar a abordagem baseada em léxicos, pois a de aprendizado de máquina

necessitaria um considerável corpus de treinamento anotado com sentimentos.

Sentimentos vêm sendo estudados em diversos domínios. Um domínio comum é referente a revisões de produtos, nos quais características de um item são classificadas como positivas ou negativas de acordo com a opinião de consumidores (BOLLEGALA; MU; GOULERMAS, 2016).

Alguns domínios apresentam dificuldades específicas, como a análise de sentimentos em notícias (GODBOLE; SRINIVASIAH; SKIENA, 2007). Atribuir polaridades negativas pode ser complicado nessa área, já que é necessário separar possíveis sentimentos negativos do simples relato em uma matéria sobre uma tragédia, um roubo, um escândalo, etc. O mesmo acontece em tíquetes, que essencialmente descrevem erros ou necessidades de usuários.

Embora muitos trabalhos utilizem ferramentas gerais para análise de sentimentos, no domínio de Engenharia de Software, os resultados apresentam imprecisões. Trabalhos recentes têm criado classificadores específicos para domínios (GACHECHILADZE et al., 2017). Essas novas ferramentas podem ser baseadas ou expandir léxicos já existentes, mesmo as de propósito geral, e melhoram a qualidade da mineração de sentimentos no domínio de TI (ISLAM; ZIBRAN, 2017).

2.5 Dicionários de Sentimentos

Um dicionário de sentimentos popular é o SentiWordNet¹ (BACCIANELLA; ESULI; SEBASTIANI, 2010), no qual cada entrada apresenta o objeto (uma palavra ou expressão, considerando sua morfossintaxe), uma pontuação numérica positiva e outra negativa. Sua criação utilizou o WordNet² (FELLBAUM, 1998), que é uma combinação de léxico genérico com thesaurus, contendo palavras na língua Inglesa e as agrupando em conjuntos de sinônimos, nomeados *synsets*, além de prover outras informações. Primeiramente definiram-se termos sementes para o SentiWordNet, com pontuações positiva e negativa conhecidas. Sua expansão foi realizada a partir da busca dos *synsets* de cada semente no léxico WordNet.

Uma solução léxica avançada para análise de polaridade amplamente utilizada é o SentiStrength³ (THELWALL et al., 2010), composta de uma fórmula de atribuição de polaridade e um léxico próprio. Essa ferramenta divide o texto em vários fragmentos, que consistem em palavras ou sentenças. Para cada fragmento, são identificados termos que denotam sentimen-

¹<http://sentiwordnet.isti.cnr.it/>

²<https://wordnet.princeton.edu/>

³<http://sentistrength.wlv.ac.uk/>

tos. Cada um desses termos possui um valor negativo e um positivo, variando de [-5,0] e de [0,5] respectivamente. Então são calculadas as pontuações totais de cada fragmento e do texto como um todo, considerando os extremos negativo e positivo. Outras soluções genéricas para polaridade são IBM Alchemy⁴, HP Haven OnDemand⁵, Stanford NLP⁶, entre outras.

Um léxico para a língua Portuguesa de Portugal é o Sentilex-PT⁷ (SILVA; CARVALHO; SARMENTO, 2012). É composto 6.321 lemas, masculinos e no singular, e mais 25.406 formas flexionadas. Dentre outras informações, cada entrada apresenta uma polaridade (positiva, negativa ou neutra), e a forma que essa polaridade foi atribuída: manualmente ou via uma ferramenta de anotação desenvolvida pela mesma equipe que gerou o dicionário.

Para análise de emoções, o léxico NRC (MOHAMMAD; TURNEY, 2010) baseia-se nas oito emoções de Plutchik. Para cada palavra presente, o dicionário indica se apresenta ou não cada uma dessas emoções, além de sua polaridade. O léxico foi gerado a partir do serviço Mechanical Turk⁸, da Amazon, que é uma plataforma online com variados serviços que necessitam de esforço humano, como a anotação de textos.

Outro dicionário voltado a emoções é o ANEW (BRADLEY; LANG, 1999), o qual contém valores das três dimensões VAD para palavras. Esse léxico foi expandido em um trabalho posterior (WARRINER; KUPERMAN; BRYLSBAERT, 2013), aumentando a quantidade de registros de 1.034 palavras para quase 14.000. Foram inclusas palavras de outras categorias mais diversificadas (tipos de doenças, profissões, termos sexuais, palavras de baixo calão). Tal qual o original, o novo dicionário apresenta pontuações para os componentes de valência, excitação e dominância para cada registro.

Estudos que usam dicionários podem utilizar abordagens variadas, por exemplo: a) adaptar dicionários existentes para seus estudos (BOLLEGALA; MU; GOULERMAS, 2016); b) expandir dicionários existentes (QIU et al., 2009); c) gerar seus dicionários baseados em outros existentes (STEINBERGER et al., 2011; MÄNTYLÄ et al., 2017); d) e criar seus próprios dicionários (SILVA; CARVALHO; SARMENTO, 2012). Léxicos gerais (livres de contexto) podem não ter boa performance em certos domínios. Por exemplo: o verbo “salvar” está presente no SentiWordNet com a polaridade positiva, pois apresenta o sentido de “salvar da ruína, destruição ou dano”. Contudo, na área de TI, é um termo comum relacionado à ação de salvar (gravar) um arquivo, o que não denota sentimento ou emoção.

Visto que nosso estudo é específico ao contexto de TI, optamos por gerar um dicionário

⁴www.alchemyapi.com/

⁵<https://www.havenondemand.com/>

⁶<https://nlp.stanford.edu/sentiment/>

⁷<http://xldb.fc.ul.pt/wiki/SentiLex-PT01>

⁸<https://www.mturk.com/mturk/welcome>

de domínio. De forma semelhante a outras propostas de léxicos orientados para o domínio (GODBOLE; SRINIVASIAH; SKIENA, 2007; HU; LIU, 2004; LIU, 2012), o processo inclui três etapas: a) pré-processamento das sementes; b) expansão das sementes e c) poda.

Nossas análises, métodos e técnicas criadas e experimentos basearam-se no SentiWordNet, WordNet, NRC (apenas considerando as seis emoções de Ekman) e na expansão do ANEW. Optamos por criar um dicionário de domínio, específico à área de TI, e desenvolver uma técnica própria para classificação para análise de sentimentos baseada em dicionários, devido ao baixo desempenho das soluções genéricas existentes para o contexto de TI.

2.6 Análise de Sentimentos Multilíngue

De forma geral, recursos para manipular e analisar textos são escassos em línguas diferentes do Inglês. Para Português, os melhores dicionários de sentimentos e ferramenta de PLN são, respectivamente, Sentilex-PT e Palavras⁹, os quais não têm bom desempenho com o Português do Brasil (TUMITAN; BECKER, 2014).

Diferentes estudos demonstram que a tradução automática é uma técnica madura para análise de sentimentos (BALAHUR; TURCHI, 2014; BECKER; MOREIRA; SANTOS, 2017; SANTOS A. G. L. DOS; BECKER; MOREIRA, 2017). Um extenso estudo experimental (BALAHUR; TURCHI, 2014) envolvendo traduções automáticas do Inglês para Francês, Alemão e Espanhol revelou que a qualidade da classificação de polaridade não correlaciona com a qualidade das traduções. Para mensuração da qualidade das traduções, os textos-base foram manualmente traduzidos, e o estudo compara quatro diferentes tradutores automáticos. Os melhores resultados de tradução foram obtidos com a língua Espanhola, parecida com Português, medidos via o método BLEU.

Tradução automática com o uso de dicionário de sentimentos em diferentes idiomas foi estudada por SANTOS A. G. L. dos; BECKER e MOREIRA (2014). O objetivo foi verificar qual produziria um melhor resultado: a tradução de textos do corpus, ou a tradução do dicionário. No caso, os textos foram escritos em Português do Brasil, e o dicionário apresentava palavras em Inglês. A tradução do corpus alcançou melhores resultados, pois os textos apresentam um contexto, enquanto que as palavras do dicionário, não.

Diferentes combinações de aprendizado de máquina para classificação de emoções e tradução automática também foram comparadas por SANTOS A. G. L. dos; BECKER e MOREIRA (2017). Foram criados classificadores de emoção separados por idiomas, chamados

⁹<http://beta.visl.sdu.dk/>

monolíngues, e criado um classificador composto do idioma original e das traduções, chamado multilíngue. O estudo envolve textos em Inglês e Português, os quais foram traduzidos para Espanhol, Francês e Português/Inglês. Os resultados de experimentos demonstram que a classificação de emoções em textos traduzidos apresenta resultados comparáveis aos do texto original.

A partir das constatações desses trabalhos, os textos coletados para nosso trabalho foram traduzidos do Português do Brasil para o Inglês, a fim de aproveitar ferramentas existentes para o processamento de linguagem natural e para determinação de sentimentos que auxiliaram em nossos experimentos.

2.7 Métricas para Avaliação

Para analisar os resultados de um classificador, podem ser considerados os valores que compõem a matriz de confusão para um problema de classificação binário. Esses valores representam a quantidade de instâncias corretas e incorretas classificadas pelos algoritmos. A Tabela 2.1 mostra os valores que compõem a matriz de confusão: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN).

Tabela 2.1 – Matriz de Confusão

		Classe Prevista	
		C_1	C_2
Classe Real	C_1	Verdadeiro Positivo (VP) C_1	Falso Negativo (FN) C_2
	C_2	Falso Positivo (FP) C_1	Verdadeiro Negativo (VN) C_2

Fonte: O Autor

Com base nos resultados da matriz de confusão, é possível calcular as medidas precisão, revocação e medida-f para cada classe, que representam o desempenho do classificador para diferentes classes $C_1, C_2 \dots C_n$, sendo $n \geq 2$.

Precisão representa a quantidade de instâncias que foram corretamente classificadas, como pertencentes a uma determinada classe pelo algoritmo:

$$Precisão_{ci} = \frac{VP_{ci}}{VP_{ci} + FP_{ci}}$$

Dada uma classe específica, revocação representa as instâncias classificadas corretamente para essa classe:

$$Revocação_{ci} = \frac{VP_{ci}}{VP_{ci} + FN_{ci}}$$

Medida-f representa a média harmônica entre precisão e revocação, na qual a mesma relevância é considerada para ambas as métricas:

$$Medida - F_{ci} = 2 \times \frac{Precisão_{ci} \times Revocação_{ci}}{Precisão_{ci} + Revocação_{ci}}$$

Uma forma de comparar o desempenho de métodos de classificação diferentes é realizar o teste T bicaudal. O aplicativo Weka Experimenter¹⁰ provê esse tipo de teste. Aplicamos esse teste para comparar os métodos criados neste trabalho entre si e com os resultados SentiStrength, adotando um nível de significância de 0,05.

2.8 Regras de Associação

Regras de associação envolvem técnicas de mineração de dados descoberta de relações entre elementos que ocorrem comumente em uma base de dados (HAN, 2005). Seu objetivo é encontrar relações de implicação entre elementos (se/então) em uma mesma transação. Cada regra identificada tem um antecedente (se) e um conseqüente (então). Foi introduzido analisando relações entre transações de produtos em sistemas de supermercados, como por exemplo, “90% das transações de compra de pão e manteiga implicam na compra de leite também”.

Uma regra de associação é uma expressão de implicação no formato $X \Rightarrow Y$, na qual X e Y são conjuntos disjuntos de itens de uma base de dados. Para medir a qualidade de regras, existem várias métricas baseadas na estrutura dos padrões descobertos e nas estatísticas subjacentes. O suporte representa a frequência de transações em uma base de dados. Sendo $P(X)$ e $P(Y)$ as frequências/probabilidades do conjuntos de itens X e Y , respectivamente, na base de dados, suporte é a probabilidade $P(X \cup Y)$, na qual $X \cup Y$ indica que uma transação contém X e Y , ou seja, a união de conjuntos de itens X e Y . Outra métrica é a confiança, que avalia o grau de certeza da associação detectada. É a probabilidade condicional $P(Y|X)$, ou seja, a probabilidade de uma transação contendo X também conter Y . Em resumo, suporte e confiança são definidos como:

$$suporte(X \Rightarrow Y) = P(X \cup Y)$$

$$confiança(X \Rightarrow Y) = P(Y|X)$$

Lift é uma métrica de correlação. A ocorrência do conjunto de itens X é independente

¹⁰<http://www.cs.waikato.ac.nz/ml/weka/>

da ocorrência do conjunto de itens Y se $P(X \cup Y) = P(X)P(Y)$. Senão, os conjuntos de itens X e Y são dependentes e correlacionados. É definida como a razão entre a confiança da regra e o suporte do conjunto de itens consequente:

$$lift(X, Y) = \frac{confiança(X \Rightarrow Y)}{suporte(Y)}$$

Se o valor resultante do lift for inferior a 1, a ocorrência de X é negativamente correlacionada com a ocorrência de Y . Se o valor resultante for maior que 1, então X e Y estão positivamente correlacionados, o que significa que a ocorrência de um aumenta a probabilidade da ocorrência do outro. Se o valor resultante for igual a 1, então X e Y são independentes e não há correlação entre eles.

Um dos algoritmos para gerar regras de associação é o Apriori¹¹. Apesar de popular, esse algoritmo pode gerar regras redundantes. A definição de uma regra redundante é: sendo $X \Rightarrow Y$ e $X' \Rightarrow Y'$ duas regras com confiança cf e cf' , respectivamente. $X \Rightarrow Y$ dita uma regra redundante para $X' \Rightarrow Y'$ se X pertence a X' ; Y' pertence a Y , e $cf \leq cf'$ (ZAKI, 2000). Regras redundantes dificultam a análise dos resultados e podem ser removidas via poda.

Neste trabalho, regras de associação são usadas para identificar possíveis correlações entre polaridade e emoções, segundo os dois modelos de emoção estudados, a fim de entender quais emoções podem ser consideradas positivas e quais negativas ao analisar sentimentos em tíquetes. Para aplicar regras de associação, foi utilizada a ferramenta RStudio¹², que é um ambiente de desenvolvimento para a linguagem de programação R, linguagem de programação para computação estatística e gráfica.

¹¹ <https://www.rdocumentation.org/packages/arules/versions/1.5-4/topics/apriori>

¹² <https://www.rstudio.com/>

3 TRABALHOS RELACIONADOS

Este capítulo apresenta trabalhos relacionados a este estudo, que abordam análise de sentimentos no domínio de TI em geral, e de software em particular.

3.1 Análise de Sentimento em Artefatos de TI

Os trabalhos descritos nesta seção abordam a análise de sentimentos em diversos artefatos relacionados a TI, e serviram como base e referência para este estudo. A Tabela 3.1 sumariza os principais aspectos dos estudos detalhados abaixo, em particular qual sentimento focado (i.e. polaridade, emoção), técnicas de análise de sentimentos empregadas (i.e. dicionário, aprendizagem supervisionada, ou outra), ferramentas de apoio (e.g. SentiStrength, ferramentas de mineração de dados), artefatos sobre os quais foram feitas as análises (e.g. postagem sobre erro, *commits* de código), autor do documento com sentimento (e.g. usuário, desenvolvedor de software), e características do corpora, tais como volume e *gold standard*. Assim as diferentes abordagens dessas pesquisas podem ser comparadas entre si, bem como com o presente estudo.

Guzman e Bruegge (2013) propõem uma abordagem que automaticamente extrai e agrega emoções de textos de variadas fontes e tipos (e não apenas em um tipo de texto), presentes em artefatos colaborativos produzidos por times de desenvolvimento. Seu corpus é composto por 1.018 textos escritos por desenvolvedores, extraídos de emails, do Twitter e de artefatos produzidos em uma ferramenta colaborativa (*commits* em repositórios de arquivos, cadastros de erros e textos de repositórios de conhecimento) durante um curso de laboratório na Technische Universität München. A análise da polaridade combinou as abordagens léxica, via a ferramenta SentiStrength, e de LDA, via a ferramenta Matlab Topic Modeling Toolbox¹. Seus resultados sugerem que os sentimentos dos colaboradores têm relação com a performance e com a motivação no projeto. Também, que os artefatos apresentam mais conteúdo neutro quando comparados com emails. Entretanto obtiveram suas conclusões a partir de entrevistas com os gestores, nas quais os resultados dos classificadores e gráficos foram discutidos.

Já El-Halees (2014) estuda textos criados por usuários, e não de desenvolvedores. Esse trabalho propõe o uso de mineração de opinião como técnica para automaticamente avaliar usabilidade de software, tal como um modelo para melhorá-la. O modelo restringe-se a três fatores de usabilidade: eficácia, eficiência e satisfação. É composto por três métodos: um léxico, com o uso da ferramenta SentiStrength, e dois de classificação: máxima entropia, utilizando o

¹http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm/

software Maxent² (MAKHOUL et al., 1999), e k-nearest, utilizando Rapidminer³ (PHILLIPS; DUDÍK; SCHAPIRE, 2004). O modelo foi aplicado em 345 revisões positivas e 220 negativas, alcançando uma acurácia de 85% e precisão, revocação e f-measure que os autores consideraram aceitáveis.

Uma das primeiras questões ao estudar mineração de opinião no contexto de TI é se textos nesse domínio apresentam sentimentos. Murgia et al. (2014) verificam se postagens de desenvolvedores em relatos de erros de sistemas apresentam emoções a respeito dos projetos nos quais estão inseridos. Esses autores consideram o estudo como um primeiro passo antes de utilizar/criar alguma ferramenta para detecção de sentimentos automaticamente, pois se humanos não conseguem determinar sentimentos nesses textos, ferramentas também não conseguirão. Sua análise considera as emoções de Parrott e seu corpus é composto por quase 392 textos do sistema de cadastro de erros da Apache Software Foundation, os quais foram analisados manualmente por anotadores em um estudo exploratório. O estudo confirma que desenvolvedores expressam emoções, principalmente amor, alegria e tristeza, sobre seus projetos (e.g. sobre escolhas de padrões, atividades de manutenção, colegas) em textos referentes a erros.

A validação das ferramentas de mineração de opinião a partir de um *gold standard* é fundamental para verificar a qualidade dos classificadores. Um dos objetivos de Jongeling e Serebrenik (2015) é verificar se os resultados das seguintes ferramentas de análise de sentimentos coincidem com os valores anotados nos textos por avaliadores: SentiStrength, Alchemy⁴, Stanford NLP e NLTK. Como *gold standard*, utilizou a base de Murgia et al. (2014), composta por 392 postagens de desenvolvedores anotadas com as seis emoções de Parrott. As quatro ferramentas classificam os textos em valores de polaridade (positivo, negativo ou neutro), portanto estabeleceram as emoções amor, alegria e surpresa como positivas, e raiva, tristeza e medo como negativas. Consideraram uma postagem positiva/negativa quando pelo menos três avaliadores indicaram emoções positivas/negativas e nenhum indicou negativas/positivas; e neutro quando pelo menos três avaliadores rotularam como neutra. Assim o corpus é composto por 24 textos negativos, 54 positivos e 217 neutros. O estudo descartou 97 textos que foram rotulados com emoções consideradas contraditórias (como medo e surpresa para o mesmo texto). Os experimentos do estudo apresentaram inconsistências. A decisão de correlacionar subjetivamente algumas emoções como positivas e outras como negativas prejudicou a qualidade resultados, pois nem todas surpresas são positivas, por exemplo. Ainda, os resultados das ferramentas apresentaram baixo desempenho quando comparados com o *gold standard* e também

²https://biodiversityinformatics.amnh.org/open_source/maxent/

³<https://rapidminer.com/>

⁴<http://www.alchemyapi.com/products/alchemy/language/sentiment-analysis>

foram contraditórios entre si. Assim, os autores sugerem a necessidade de uma ferramenta para análise de sentimentos focada no domínio de engenharia de software.

A correlação entre sentimentos e emoções também é um objeto de estudo na mineração de opinião. A pesquisa de Ortu et al. (2015) busca correlacionar sentimentos e emoções, além do conceito de educação (politeness). Considera educação como a capacidade de tornar todas as partes descontraídas e confortáveis umas com as outras. Utilizou a ferramenta Stanford Politeness⁵, construída com técnicas de aprendizado de máquina e que, dentre outras funções, avalia a educação de requisições do site Stack Overflow⁶, o qual é bastante popular na engenharia de software. Para sentimentos, usou a ferramenta SentiStrength. Já para emoções, como não há disponíveis, construiu uma própria com a abordagem de aprendizado de máquina, considerando apenas quatro emoções básicas (das seis de Parrott): alegria, amor, raiva e tristeza. Seu corpus é composto por mais de 560.000 textos do sistema de cadastro de erros da Apache Software Foundation. A fim de verificar correlações entre sentimentos, emoções e educação, construíram um modelo de regressão a partir de métricas para os três conceitos. Então calcularam o coeficiente de correlação de Pearson sobre suas métricas, considerando como uma correlação fraca um valor menor que 0,4, moderada entre 0,4 e 0,7, e forte acima de 0,7. De forma geral, concluí que os conceitos são independentes, com uma correlação fraca, de no máximo 0.36.

Corpora anotados são fundamentais para pesquisas de análise de sentimentos, contudo há escassez de textos anotados com emoções no domínio de engenharia de software. O trabalho de Ortu et al. (2016) gerou um corpus anotado com 2.000 comentários e 4.000 sentenças escritas por desenvolvedores. Os textos foram extraídos de sistemas de cadastro de erros dos repositórios de quatro comunidades que desenvolvem projetos de código-aberto: Apache, Spring, JBoss, e CodeHaus. Foram anotadas as emoções de Parrott por dezesseis avaliadores: quatro mestrandos, dez doutores e dois pesquisadores.

O trabalho de Gachechiladze et al. (2017) verifica a viabilidade de identificar automaticamente a emoção raiva, presente nos modelos de Ekman e de Parrott, na comunicação entre desenvolvedores. Como *gold standard*, anotou 723 textos de postagens referentes a interação entre desenvolvedores no sistema de cadastro de erros da Apache Software Foundation. Considerou um modelo no qual a emoção raiva, quando presente em um texto, pode ser direcionada a um de três alvos: ou ao próprio desenvolvedor que escreveu o texto, ou a outro(s) desenvolvedor(es), ou a objeto(s). Essa divisão auxilia: na prevenção da desmotivação e da perda de produtividade, quando o alvo é o próprio desenvolvedor; na orientação sobre o comportamento dos desenvolvedores em suas interações e a resolução de problemas em equipe, quando o alvo

⁵<https://www.cs.cornell.edu/cristian/Politeness.html>

⁶<http://stackoverflow.com>

são outros desenvolvedores; e problemas de frameworks, linguagens de programação e documentação, quando o alvo são objetos. Para identificação dos alvos, extraiu *tokens* e calculou os termos mais importantes via filtro *StringToWordVector* da ferramenta Weka, estabelecendo os alvos manualmente a partir desses termos. O *gold standard* mostrou-se não balanceado, com 73% dos textos com emoções apresentando objetos como alvo. Para classificação de sentimentos, considerou os algoritmos de aprendizado de máquina Support Vector Machines (SVM), J48, e Naive Bayes, os executando também na ferramenta Weka. O melhor desempenho pertence ao algoritmo SVM, com uma alta precisão (entre 80% e 89%) para os três tipos de alvo, contudo com uma baixa revocação (18%) quando o alvo são outros desenvolvedores.

Islam e Zibran (2017) conduziram um estudo quantitativo para identificar o motivo do baixo desempenho das ferramentas de análise de sentimentos no domínio de engenharia de software e constataram que grande parte das dificuldades ocorre porque essas ferramentas são genéricas. Também desenvolveram uma ferramenta específica para o domínio de engenharia de software, o SentiStrength-SE, a partir da popular SentiStrength. Primeiramente construíram um dicionário de domínio a partir de 490.000 *commits* de 50 projetos de código-aberto presentes no GitHub. Extraíram os *tokens* desses *commits*, os converteram para seus respectivos lemas e verificaram quais estão presentes no dicionário do SentiStrength. Por fim, identificaram subjetivamente 460 lemas que apresentam sentimentos no domínio de TI, os quais compõem seu dicionário. O classificador desenvolvido usa a mesma abordagem e técnica do SentiStrength. A fim de avaliar seus resultados, criaram um *gold standard* com 5.992 postagens e interações de desenvolvedores, extraídos do sistema de erros do Jira⁷ e anotados com as emoções de Parrott. Já que tanto o SentiStrength, quanto o SentiStrength-SE classificam no formato de polaridade, consideraram as emoções amor e alegria como positivas, e raiva, tristeza e medo como negativa. Já surpresa pode ser considerada como positiva ou negativa, dependendo do texto. As comparar os resultados das ferramentas SentiStrength e SentiStrength-SE, a criada nesse estudo é superior, alcançando 73,85% de precisão, 85% de revocação e 77,48% de medida-f.

O trabalho de Pletea, Vasilescu e Serebrenik (2014), apesar de focar no domínio de TI, atua apenas em um contexto específico. Esse estudo analisa sentimentos em textos que tratam sobre tópicos de segurança em TI, a partir de 115.550 *commits* e requisições de desenvolvedores presentes no GitHub, provenientes do MSR 2014 Mining Challenge Dataset. Para análise de sentimentos, utilizou a ferramenta NLTK⁸. Dado um texto, o aplicativo retorna as probabilidades do texto ser positivo, negativo ou neutro. Para classificação dos sentimentos, a ferramenta foi treinada com textos de revisões de filmes e executada com os algoritmos Naive Bayes e

⁷<https://www.atlassian.com/software/jira>

⁸<http://www.nltk.org/>

Hierárquico. Um estudo de caso foi realizado em 30 *commits*, como forma de avaliação dos resultados. Ao comparar com textos de outros assuntos, seus resultados apresentam mais emoções negativas em textos sobre segurança (72,52% contra 54,2% em *commits* e 81,00% contra 69,58% para requisições).

Já Guzman, Azócar e Li (2014) não restringem a análise a *commits* de um contexto específico. Estudam emoções em *commits* de desenvolvedores de projetos abertos e de locais diferentes e as relacionam com outros fatores, como linguagem de programação, hora e dia da semana, distribuição do time e aprovação do projeto. Para sua análise, coletaram 60.425 textos do Github e os classificaram com a ferramenta SentiStrength, considerando o *commit* positivo se sua pontuação ficasse na faixa (1,5], negativo em (-1,-5] e neutro em [-1,1]. Concluem que existem relações entre emoções e os fatores analisados: projetos desenvolvidos em Java e *commits* escritos na segunda-feira tendem a ser mais negativos, enquanto que projetos com times distribuídos têm conteúdos mais positivos. Contudo contestam que seus resultados devem ser confirmados com uma amostra maior. Não houve validação dos resultados gerados pelo classificador nesse trabalho.

A pesquisa de Sinha, Lazar e Sharif (2016) apresenta uma análise de 2.251.585 *commits* de desenvolvedores em projetos do GitHub ao longo de sete anos. Os textos foram agrupados com base no dia da semana que o *commit* foi realizado, a fim de determinar se há relação entre sentimentos e o dia. Para classificação de sentimentos, usou a ferramenta SentiStrength. A maioria dos textos foram classificados como neutros, 74,74%, enquanto que 18,05% como negativos e 7,20% como positivos. Terça-feira é o dia com mais sentimentos negativos, apesar que os projetos que possuem mais *commits* apresentaram quarta-feira e quinta-feira como os dias mais negativos.

3.2 Considerações Finais

Esses trabalhos proveram técnicas que expandimos, usamos como base para criação de novas e comparamos os resultados obtidos em nossa pesquisa.

As conclusões de Murgia et al. (2014), de que desenvolvedores expressam sentimentos em textos referentes a erros, dão subsídio para o desenvolvimento de ferramentas para análise automática de sentimentos no domínio de TI, como realizamos em nosso trabalho. Também expandimos esse trabalho ao verificar a presença de sentimentos em textos criados por usuários, e não por desenvolvedores.

Apesar do estudo de El-Halees (2014) analisar textos criados por usuários, no domínio

de TI, limita-se ao contexto de usabilidade, enquanto que Pletea, Vasilescu e Serebrenik (2014), restringe-se a tópicos de segurança em TI. O corpus anotado gerado por Ortu et al. (2016) possui textos de várias organizações, mas todos escritos por desenvolvedores. Já nosso trabalho abrange um contexto maior, ao analisar tíquetes de diferentes tipos: erros (de infraestrutura e sistema), requisições (para extração de dados, manutenção de sistema, infraestrutura e necessidades administrativas) e *feedback* (da área de TI, sistemas e infraestrutura), que foram extraídos de cinco organizações de áreas/negócios variados: telecomunicações, desenvolvimento de sistemas, seguros e previdência e pesquisa.

A constatação de Pletea, Vasilescu e Serebrenik (2014), que textos sobre tópicos de segurança em TI apresentam mais sentimentos negativos quando comparados a textos gerais, é interessante quando comparada ao nosso trabalho. Nossos textos são, na maioria, tíquetes relatando erros, o que pode remeter a ideia de serem negativos na maioria também. Já os resultados de Sinha, Lazar e Sharif (2016) apontam que a maioria dos textos é neutro, o que também ocorreu em nossas análises.

O trabalho de Guzman, Azócar e Li (2014) não validou os resultados de seu classificador, enquanto que o estudo de Guzman e Bruegge (2013) realizou apenas algumas entrevistas demonstrando seus resultados a gestores. Já os resultados de nossos métodos foram validados com um *gold standard* e seus desempenhos comparados e analisados com as métricas precisão, revocação e medida-f.

O baixo desempenho da ferramenta genérica usada por Jongeling e Serebrenik (2015), e a conclusão de que são necessárias de soluções específicas ao domínio de TI para análise de sentimentos converge com a análise inicial que realizamos em nosso estudo. Assim optamos pelo desenvolvimento de métodos novos, focados na área de TI.

O estudo de Gachechiladze et al. (2017) gerou um *gold standard* e executou diferentes algoritmos de classificação de sentimentos. No entanto utilizou métodos de classificação genéricos e restringiu-se à emoção raiva, enquanto que nossos métodos de classificação foram criados especificamente para o contexto de TI e, sobre emoções, consideramos as seis básicas de Ekman e o modelo VAD.

Um dos objetivos de Ortu et al. (2015) foi verificar possíveis correlações entre sentimentos e emoções via modelos de regressão. Já Jongeling e Serebrenik (2015), Islam e Zibran (2017) consideram algumas emoções como positivas e outras como negativa, mas de forma subjetiva. Nosso estudo também tem o objetivo de verificar possíveis correlações entre emoções e polaridade, contudo consideramos os conceitos de regras de associação para essa análise.

O trabalho de Islam e Zibran (2017) é o mais similar ao nosso, criando um dicionário de

domínio para engenharia de software e um método para classificação léxico próprio, baseados em uma ferramenta existente para análise de sentimentos. Baseia-se em textos criados por desenvolvedores, enquanto que abordamos textos escritos por usuários, além de criarmos três métodos de classificação.

Tabela 3.1 – Trabalhos Relacionados

Trabalho	Tipo de Análise	Correlaciona Tipos de Análise	Técnica	Ferramentas para Análise de Sentimentos	Artefato Analisado	Corpus	Autor dos Textos Analisados	Gold Standard
Guzman e Bruegge (2013)	Polaridade	-	Dicionário; LDA	SentiStrength	Email; Commit; Postagem de erro (sobre sistema); Repositório de conhecimento Wiki; Postagem no Twitter	1.018 textos de emails, do Twitter e de artefatos produzidos em uma ferramenta colaborativa durante o curso de laboratório iPraktikum da Technische Universität München	Desenvolvedor	Não
El-Halees (2014)	Polaridade	-	Aprendizado de Máquina; Dicionário	SentiStrength; Rapidminer; Maxent	Revisão (sobre usabilidade)	565 textos de revisões de sistema	Usuário	Sim (todo o corpus)
Murgia et al. (2014)	Emoção (Parrott)	-	Manual	-	Postagem/Interação sobre erro	392 textos do sistema de cadastro de erros da Apache	Desenvolvedor	Não se aplica
Jongeling e Serebrenik (2015)	Polaridade; Emoção (Parrott)	Sim (manualmente)	Aprendizado de Máquina; Dicionário	SentiStrength; Alchemy; Stanford NLP; NLTK	Postagem/Interação sobre erro	392 textos extraídos de Murgia et al. (2014)	Desenvolvedor	Sim (todo o corpus, anotado em Murgia et al. (2014))
Ortu et al. (2015)	Polaridade; Emoção (alegria, amor, raiva e tristeza); Educação	Sim (coeficiente de correlação de Pearson)	Aprendizado de Máquina; Dicionário	SentiStrength; Stanford Politeness; Própria (criada)	Postagem/Interação sobre erro	Mais de 560.000 textos do sistema de cadastro de erros da Apache	Desenvolvedor	Sim (4.000 textos)
Ortu et al. (2016)	Emoção (Parrott)	-	Manual	-	Postagem/Interação sobre erro	6.000 textos de sistemas da Apache, Spring, JBoss e CodeHaus	Desenvolvedor; Usuário	Sim (todo o corpus)
Gachechiladze et al. (2017)	Emoção (Raiva)	-	Aprendizado de Máquina	Weka	Postagem/Interação sobre erro	723 textos do sistema de cadastro de erros da Apache	Desenvolvedor	Sim (todo o corpus)
Islam e Zibran (2017)	Polaridade; Emoção (Parrott)	Sim (manualmente)	Dicionário	SentiStrength; Própria (criada)	Postagem/Interação sobre erro	5.922 textos do sistema de cadastro de erros do Jira	Desenvolvedor	Sim (todo o corpus)
Pletea, Vasilescu e Serebrenik (2014)	Polaridade	-	Aprendizado de Máquina	NLTK	Commit (sobre segurança); Requisição (para revisão de código)	115.550 textos do MSR 2014 Mining Challenge Dataset (presentes no GitHub)	Desenvolvedor	Sim (30 textos)
Guzman, Azócar e Li (2014)	Polaridade	-	Dicionário	SentiStrength	Commit	60.425 textos do GitHub	Desenvolvedor	Não
Sinha, Lazar e Sharif (2016)	Polaridade	-	Dicionário	SentiStrength	Commit	2.251.585 textos do GitHub	Desenvolvedor	Não
Este trabalho	Polaridade; Emoção (Ekman e VAD)	Sim (regras de associação)	Dicionário; Estrutura; Híbrida	SentiWordNet; Própria (criada)	Postagem de erro (infraestrutura, sistema); Requisição (extração de dados, manutenção de sistema, infraestrutura, administrativa); Feedback (área de TI, sistema, infraestrutura)	34.895 textos de sistemas de suporte de TI de diferentes empresas	Usuário	Sim (2.333 textos, apenas com polaridade)

Fonte: O Autor

4 ABORDAGEM PROPOSTA

Este capítulo descreve a abordagem e as técnicas utilizadas e criadas neste trabalho para análise de sentimentos em tíquetes. Primeiramente, o corpus coletado para o estudo é descrito. Em seguida, é apresentada uma visão geral das técnicas. O DD e a classificação de sentimentos são detalhadas em seguida.

4.1 Características dos Tíquetes

Nosso estudo foi baseado em um corpus composto por 34.895 tíquetes de cinco organizações reais que atuam em diferentes áreas de negócios, conforme detalhado na Tabela 4.1. Esses tíquetes foram extraídos dos respectivos sistemas de suporte dessas empresas, relatados por usuários internos ou externos, e com conteúdos variados, conforme a cobertura de serviços prestados pela equipe de TI. Todos os tíquetes são escritos em Português do Brasil.

Tabela 4.1 – Descrição das Fontes dos Tíquetes

Negócio	Usuário	Cobertura do Suporte de TI	Tíquetes
Telecomunicação	Interno	Infraestrutura e sistemas desenvolvidos e adquiridos para uso interno.	29.748
Desenvolvimento de Software	Interno	Tarefas administrativas e sistemas desenvolvidos para clientes.	2,898
Desenvolvimento de Software	Externo	Sistemas desenvolvidos para clientes.	1,745
Seguros e Previdência	Interno	Sistemas desenvolvidos para uso interno.	362
Pesquisa	Externa	Sugestões à infraestrutura fornecida a usuários e à área de TI.	142

Fonte: O Autor

Primeiramente foi realizada uma análise preliminar dos tíquetes para estudar suas especificidades. Como resultado dessa análise, foram identificadas três categorias de tíquetes, ilustradas na Tabela 4.2. Tíquetes de erros relatam problemas de infraestrutura ou defeitos em softwares. Requisições envolvem demandas relativas à extração de dados, manutenção de sistemas, necessidades de infraestrutura ou tarefas administrativas. *Feedbacks* são direcionados para

equipe ou departamento de TI e a respeito de softwares, como um meio para sugerir melhorias. A Tabela 4.2 mostra também subtipos para cada uma destas categorias, junto com um exemplo representativo.

Tabela 4.2 – Tipos de Tíquetes e Exemplos

Tipo	Subtipo	Exemplo
Erro	Infraestrutura	<i>“Estou com problemas para enviar e-mails externos, preciso urgentemente de ajuda, pois preciso enviar propostas comerciais.”</i>
	Software	<i>“Bom dia, por favor verifiquem por que não está exibindo o contrato coberto nos certificados.”</i>
Requisição	Extração de Dados	<i>“Por favor, nos enviem a lista de contribuições canceladas desse mês.”</i>
	Manutenção de Sistema	<i>“Precisamos criar um relatório de pagamentos em um determinado período. Atualmente geramos manualmente.”</i>
	Infraestrutura	<i>“Instalar o CRM em meu computador.”</i>
	Administrativo	<i>“Registrar as horas do João imediatamente.”</i>
Feedback	Área de TI	<i>“Excelente suporte da equipe responsável pela manutenção do ambiente.”</i>
	Software	<i>“O que acham de automatizar a importação dos dados do sistema antigo?”</i>
	Infraestrutura	<i>“Habilitar a submissão de mais jobs no cluster poderia diminuir a espera de usuários no pool.”</i>

Fonte: O Autor

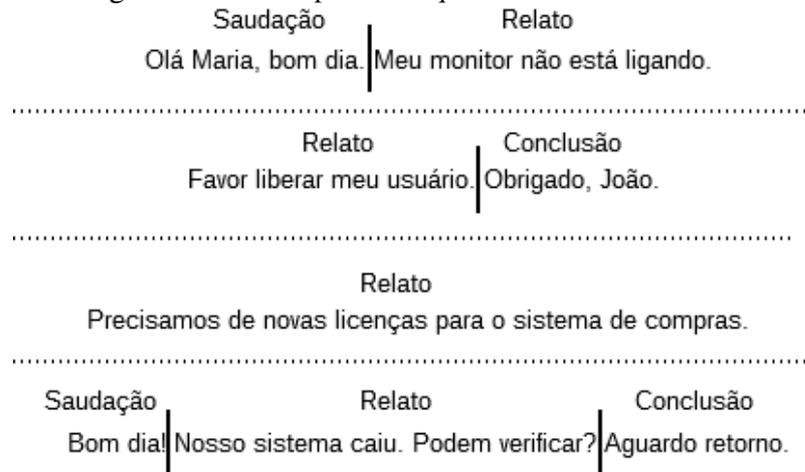
Nota-se que os tíquetes têm uma estrutura semelhante quando comparados a emails, composta geralmente por: a) saudação, onde os usuários escrevem cumprimentos; b) relato, onde o usuário solicita algo ou relata um problema; e c) conclusão, onde o usuário escreve uma mensagem de despedida ou expressa algum senso de ansiedade. Existem algumas expressões que caracterizam os cumprimentos (por exemplo, “prezado”, “olá”, “boa tarde”) e segmentos de fechamento (por exemplo, “obrigado”, “abraços”, “aguardo retorno”). A parte do relato é a mais importante e está sempre presente no texto, enquanto que as outras são opcionais. A Figura 4.1 ilustra os tíquetes com base nas quatro combinações possíveis desses segmentos.

Inicialmente, realizamos uma análise superficial em uma amostra do nosso corpus para identificar se sentimentos na forma de polaridade são expressos em tíquetes de TI. Apesar de a maioria dos tíquetes serem neutros, observamos formas sutis ou explícitas de expressão do sentimento de acordo com os termos empregados. Saudações e expressões de conclusão são

formas mais fracas de sentimento, enquanto expressões emocionais dentro do segmento de relato são indicadores mais fortes de estados afetivos. Por exemplo, no texto “*Desbloqueiem meu usuário. Obrigado, João*”, a palavra final “obrigado” é sutil, mas traz um sentimento positivo ao tíquete. Em “*Temos um problema grave no módulo de fornecedores, não conseguimos fazer pagamentos.*”, a palavra “grave” tem uma clara conotação negativa. Quanto à polaridade, é possível haver termos positivos e negativos e, em certos casos, é difícil determinar a polaridade do tíquete. Em “*O sistema tem um leiaute lindo, mas o menu é horrível*”, não é possível detectar qual sentimento é dominante devido a uma intensidade semelhante, portanto é considerado neutro. Por outro lado, em “*Prezado João, temos um problema de segurança perigoso no nosso armazenamento de dados*”, embora “prezado” seja positivo, “perigoso” é um termo negativo mais forte, então o tíquete denota um sentimento negativo. Esta intuição foi confirmada quando três revisores geraram manualmente um *gold standard*, conforme detalhado na Seção 5.1.

Já a análise preliminar de emoções mostrou-se bem mais complexa. A partir de outros modelos de emoções já utilizados (Ekman e VAD), em muitos casos é difícil concluir subjetivamente quais emoções estão expressas em um tíquete. As características intrínsecas dos modelos de emoção fazem com que as interpretações variem bastante. O uso de VAD, por exemplo, implica avaliar a intensidade das emoções em três dimensões, o que dificulta uma avaliação/anotação subjetiva de quais emoções um tíquete apresenta. Essa avaliação seria mais simples se fosse realizada com valores discretos. Mesmo analisando com emoções básicas tais como as de Ekman, a identificação das emoções presentes em documentos é bem complexa e subjetiva. Por exemplo, a emoção *surpresa*, é ambígua, conforme conclusões dos estudos de Jongeling e Serebrenik (2015) e Islam e Zibran (2017). Em “*Obrigado pelo rápido retorno! Mas o erro no relatório continua.*”, o trecho “*Obrigado pelo rápido retorno!*” indica uma surpresa positiva. Já em “*Urgente! O sistema de vendas caiu.*” é transmitida uma emoção de surpresa negativa. Devido a esse cenário, o foco deste trabalho e da abordagem proposta foi a análise de polaridade, enquanto que no tocante à análise de emoções, nos restringimos à verificação da correlação entre polaridade e emoções como um aspecto complementar.

Figura 4.1 – Exemplos de Tíquetes e de suas Estruturas



Fonte: O Autor

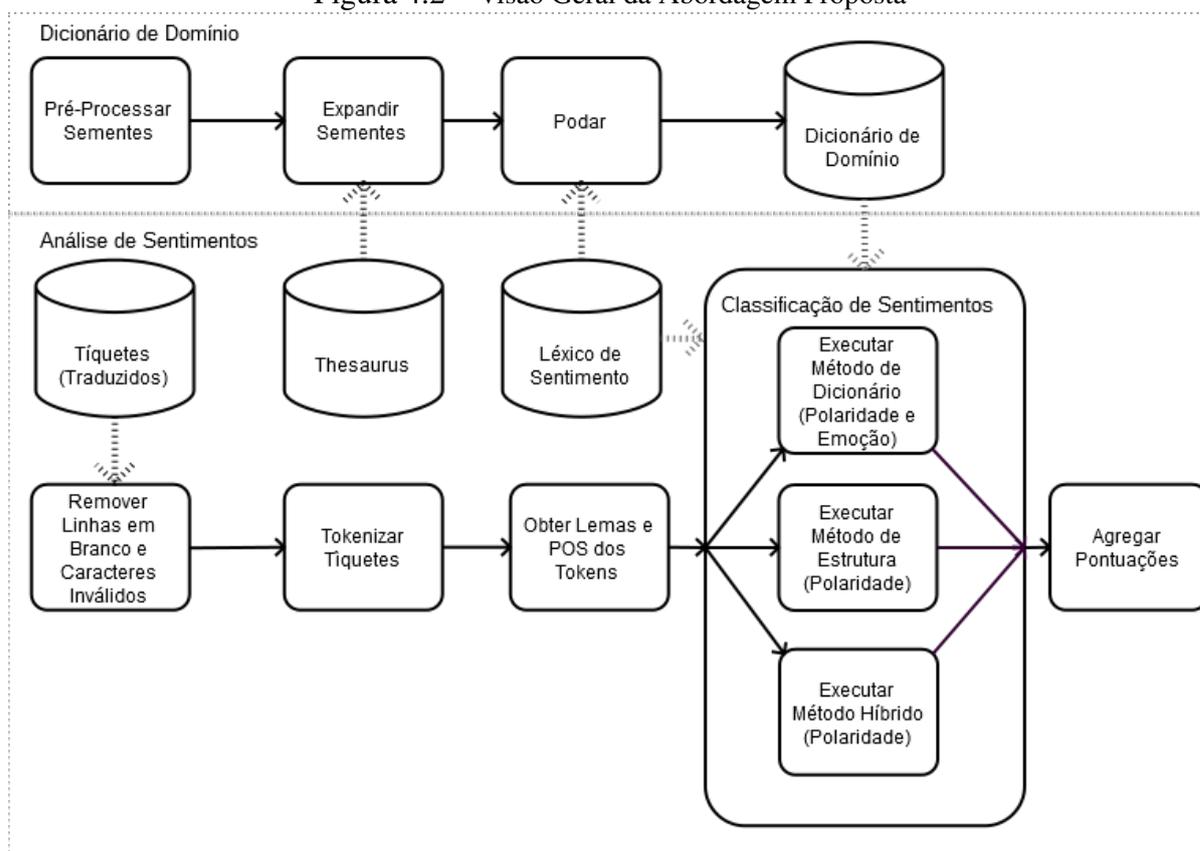
4.2 Visão Geral da Abordagem para Análise de Sentimento

Como já mencionado, o principal desafio na análise de sentimentos contidos em tíquetes de TI é detectar automaticamente possíveis sentimentos embutidos na narração dos fatos objetivos sobre um problema relatado pelo usuário. Quando um usuário escreve um tíquete, na maioria das vezes está relatando algo intrinsecamente negativo: um defeito, um erro, um incidente, um pedido de ajuda. No entanto, isso não significa que está expressando sentimentos negativos sobre esse problema. Desenvolvemos uma abordagem que pode identificar e separar automaticamente essas duas informações. A abordagem proposta está ilustrada na Figura 4.2, e inclui dois subprocessos: a criação automática de um Dicionário de Domínio personalizado e o processo de análise de sentimentos. Decidimos usar uma abordagem baseada em dicionário para mineração de sentimentos, uma vez que a adoção do aprendizado de máquina requer um corpus anotado de tamanho considerável.

Utilizamos neste processo léxicos e ferramentas de PLN disponíveis para a língua Inglesa. Assim, as ferramentas de PLN, os léxicos utilizados bem como o DD criado são todos voltados à língua inglesa. Assumimos que caso os documentos não estejam escritos em Inglês, como é o caso do corpus de tíquetes explorado neste estudo, eles podem ser traduzidos automaticamente como uma etapa preliminar. Com efeito, os sistemas de tradução automática atingiram um bom nível de maturidade para fins de análise de sentimentos (BALAHUR; TURCHI, 2014; MOLINA-GONZALEZ et al., 2013; STEINBERGER et al., 2011). Os tíquetes de nosso corpus foram traduzidos para o Inglês via a ferramenta Google Translator¹, uma solução

¹<https://translate.google.com/>

Figura 4.2 – Visão Geral da Abordagem Proposta



Fonte: O Autor

com baixo custo e que produz ótimos resultados. Ao longo deste trabalho, todos os exemplos (e.g. textos, tokens, figuras) estão escritos em Português apenas para aumentar a legibilidade do texto.

O processo de *criação de um dicionário de domínio* visa obter um léxico que permita a distinção entre palavras objetivas e de sentimento no domínio de TI. Basicamente, o DD contém apenas palavras e expressões que incorporam algum sentimento no domínio, de modo que possa ser usado como um filtro prévio para determinar seu sentimento, bem como auxiliar à atribuição de uma pontuação de sentimento. O DD é portanto complementar em relação aos léxicos genéricos de sentimento. O DD personalizado é criado pela expansão e poda automáticos de termos/expressões semente, usando um thesaurus e um léxico de polaridade. Adotamos o thesaurus WordNet e o dicionário de sentimento SentiWordNet, ainda que outros dicionários possam ser usados.

Cada palavra/expressão fornecida como semente deve ser associada a uma orientação de polaridade (i.e. positivo ou negativo), e possivelmente uma informação de uso estrutural (e.g. as palavras de cortesia “Bom dia” e “Atenciosamente” são tradicionalmente utilizadas no início e no fim dos tiquetes, respectivamente). O processo de expansão e poda agrega novas entradas

ao DD, as quais herdam estas informações das sementes a partir das quais foram expandidas. Mais detalhes sobre a criação deste dicionário são fornecidos na Seção 4.3.

O processo de *análise do sentimento* é realizado de acordo com três etapas: pré-processamento, classificação de palavras/expressões individuais e agregação do sentimento. Realizamos a análise de sentimento em nível do documento porque o tíquete é avaliado como um todo e comumente se refere a uma única entidade.

Na primeira etapa, *pré-processamento*, os tíquetes são pré-processados utilizando ferramentas de PLN que extraem e preparam informações textuais para análise de sentimentos (tokenização, etiquetagem morfosintática, etc.), extraindo do texto termos e/ou expressões que doravante denominaremos *tokens*. Como mencionado, traduzimos todos os tíquetes do nosso corpus para o Inglês usando o Google Translate.

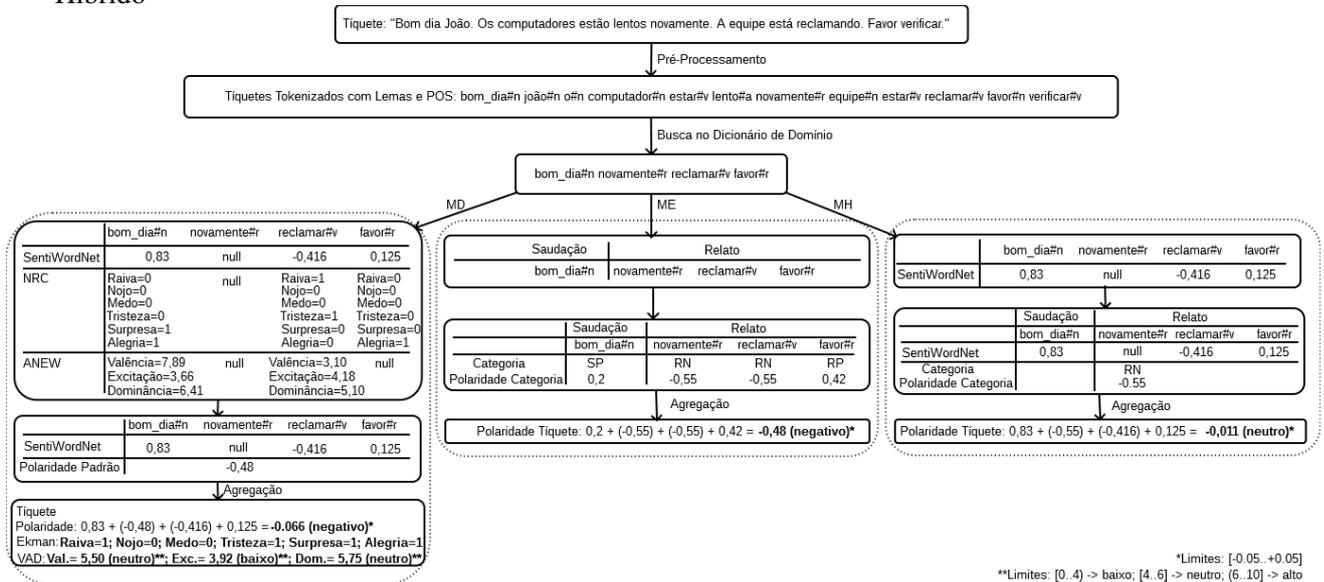
Na etapa de *classificação*, todos os *tokens* contidos nos tíquetes pré-processados recebem uma medida/pontuação de polaridade e de emoção. O DD tem duplo uso: a) filtrar os *tokens* candidatos a sentimentos, e b) fornecer informações que permitem mensurar automaticamente o sentimento de acordo com um algoritmo de classificação de sentimento. Mais especificamente, as informações associadas à cada entrada do DD, junto com os dicionários de sentimento genéricos, são utilizadas pelos métodos de classificação de polaridade propostos para gerar automaticamente uma pontuação.

Desenvolvemos e testamos três algoritmos para atribuir uma medida de sentimento às palavras/expressões contidas em tíquetes: Método de Dicionário (MD), Método de Estrutura (ME) e Método Híbrido (MH), detalhados na Seção 4.4.2. MD é um método puramente léxico, enquanto que os outros consideram, além disso, informações estruturais. O MD classifica tanto polaridade quanto emoções. Já os outros dois são voltados apenas à polaridade.

A *agregação* é o passo final do processo que combina as pontuações de polaridade e de emoções atribuídas aos *tokens* individuais, gerando uma avaliação de polaridade/emoções do tíquete como um todo.

O processo de análise do polaridade está detalhado na Seção 4.4, e o de emoções, na Seção 4.5. A Figura 4.3 ilustra um exemplo com as principais etapas do processo e os resultados distintos de cada algoritmo, os quais serão detalhados na Seção 4.4.

Figura 4.3 – Exemplo de Classificação de Sentimentos dos Métodos de Dicionário, de Estrutura e Híbrido



Fonte: O Autor

4.3 Dicionário de Domínio

Experiências iniciais de uso do SentiWordNet em nosso corpus nos permitiram detectar os seguintes problemas com este léxico de propósito geral:

- Apesar do léxico associar uma polaridade a certos termos ou expressões, eles são neutros no domínio de TI, pois correspondem a jargões neutros comuns na área (e.g. “salvar”, “diretório”, “processo”), ou termos subjacentes à descrição de problemas ou pedidos em um tíquete (e.g. “erro”, “problema”, “defeito”);
- Muitos termos utilizados para expressar o sentimento neste contexto não estão presentes no dicionário (e.g. “urgentemente”, “priorizar”, “extremamente lento”);
- A polaridade de alguns termos/expressões é oposta àquela registrada no léxico, quando se considera seu uso no domínio de TI para questões de relatos de erros. Exemplos são expressões como “guardo retorno” (positivo em SentiWordNet, mas negativo em domínio de TI), ou “importante” (positivo em SentiWordNet, mas negativo em domínio de TI).

Por isso, propomos um processo para criar um dicionário personalizado de sentimentos de TI, que é esboçado no quadro superior da Figura 4.2 (“Dicionário de Domínio”). O processo depende de três entradas: a) um conjunto de sementes; b) um léxico de sentimento; e c) um

thesaurus de sinônimos. O processo inclui três etapas: a) pré-processamento das sementes; b) expansão das sementes e c) poda, os quais são detalhados no restante desta seção.

Cada entrada no DD é identificada de forma exclusiva por <token>#<POS>, e associada a uma orientação de polaridade (i.e. positiva e negativa), e classificação posicionais (i.e. delimitador inicial/final). Os métodos de análise de sentimento propostos neste trabalho apenas classificam o sentimento de *tokens* que são encontrados no DD. Eles exploram a orientação de polaridade e a classificação posicional para determinar dinamicamente uma pontuação de polaridade de um *token* extraído de um tíquete, conforme será discutido na Seção 4.4.

Nossa implementação utilizou o SentiWordNet como léxico de sentimentos e o WordNet como thesaurus de sinônimos. Para acessar o WordNet, foi usada a ferramenta JWI².

4.3.1 Pré-processamento de Sementes

O pré-processamento das sementes consiste em selecionar e preparar termos/expressões para criar o dicionário. As sementes são palavras ou expressões extraídas de tíquetes de TI, que são preparadas manualmente da seguinte forma:

- a) Atribuir manualmente uma orientação de polaridade à semente de acordo com o domínio (por exemplo, “horrrível” é negativo, e “excelente” é positivo);
- b) Identificar entre as sementes as que delimitam segmentos iniciais (saudação) e finais (conclusão) de um tíquete (por exemplo, “prezado” é um delimitador inicial e “aguardando retorno” é um delimitador final);
- c) Substituir espaços em branco por “_” em uma expressão que deve ser considerada como um todo (por exemplo, “bom dia” é representado como “bom_dia”);
- d) Classificar cada termo candidato para expansão de acordo com uma das seguintes POS: “a” (adjetivo), “n” (substantivo), “r” (advérbio) ou “v” (verbo) e adicionando-o ao *token* usando a marca especial “#” (por exemplo, “excelente#a”, “extremamente#r lento#a” e “boa_tarde#n”). A etiquetagem POS é explorada tanto na etapa de poda (descrita em seguida), quanto na classificação de sentimentos.

Em nosso trabalho, escolhemos as sementes usando uma lista expressões de saudações/conclusões e um conjunto de palavras de sentimento determinadas manualmente, extraídas das mais frequentes encontradas no *gold standard*.

²<http://projects.csail.mit.edu/jwi/>

4.3.2 Expansão

Expansão do vocabulário relacionado ao domínio a partir de sementes é uma abordagem comum para a criação de léxicos de sentimentos personalizados, nos quais *synsets* são usados para encontrar termos similares de forma recursiva, considerando um limite máximo de iterações nessa recursividade (TSYTSARAU; PALPANAS, 2012).

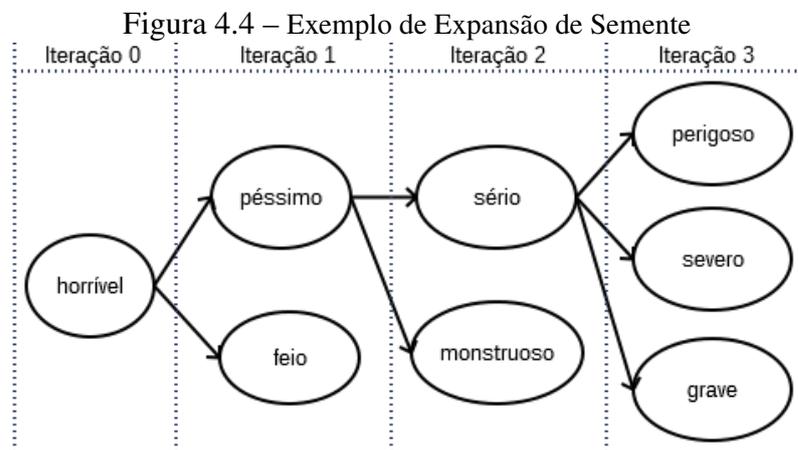
Através de uma análise preliminar, observamos que nossas sementes continham três padrões específicos a serem expandidos:

- a) Palavras, que poderiam ser substituídos por sinônimos para expressar sentimentos (por exemplo, “excelente”);
- b) Termos/Expressões com mais de uma palavra em que o todo é responsável pela polaridade (por exemplo, “bom dia”, “aguardo retorno”);
- c) Termos/Expressões com mais de uma palavra em que apenas alguns termos poderiam ser substituídos por sinônimos independentes, formando todas as combinações possíveis (por exemplo, em “impossível de usar”, tanto o termo “impossível” quanto o “usar” são expansíveis, assim a expressão como um todo é semelhante a “impossível de utilizar”, a “impraticável de usar” e a “impraticável de utilizar”).

Dado um limite de iterações e um conjunto de sementes pré-processadas, o passo de expansão explora recursivamente os *synsets* para gerar mais *tokens* (GODBOLE; SRINIVASAIAH; SKIENA, 2007; LIU, 2012). Um exemplo é fornecido na Figura 4.4. Se o limite é definido como 1, somente os sinônimos diretos são recuperados (por exemplo, “péssimo”), caso contrário, caminhos mais longos são explorados (por exemplo, “grave” e “perigoso” para o limite 3). A cada iteração realizada, os *tokens* gerados que já foram identificados anteriormente são desconsiderados.

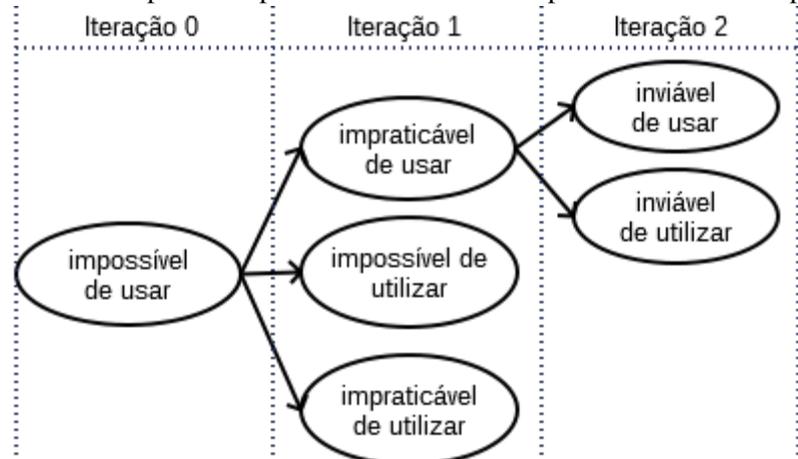
Expressões compostas são indicadores importantes de sentimentos em tíquetes. Algumas são comuns em tíquetes, como “muito/extremamente lento”, “muito/extremamente importante”, ou “mais breve/rápido possível”. Em muitas dessas expressões, uma ou mais palavras que as compõe podem variar e seus sinônimos podem formar outras expressões compostas com o mesmo significado da original. Esse cenário é coberto no método de expansão, no qual palavras dessas expressões são expandidas individualmente e todas as combinações possíveis são formadas e adicionadas no DD. A Figura 4.5 apresenta um exemplo de parte da expansão da semente “impossível de usar”.

Durante a expansão, um *token* de sinônimo herda todas as informações de polaridade e de estrutura da sua semente correspondente. Por exemplo, o termo semente “prezado” foi classificado manualmente como positivo e como delimitador inicial. Portanto, o seu sinônimo “estimado” será classificado de forma semelhante. Somente *tokens* com o mesmo POS da semente são considerados durante a expansão, pois o SentiWordNet apresenta pontuações de polaridade diferentes para o mesmo termo, com POS diferentes. Por exemplo: em Inglês, “dear” está presente no SentiWordNet tanto como adjetivo (significando “prezado”) quanto como adjetivo (significando “à alto preço”). Os dois casos apresentam pontuações diferentes, e identificamos “dear” apenas no formato de adjetivo em tíquetes.



Fonte: O Autor

Figura 4.5 – Exemplo de Expansão de Semente Composta de Palavras Expansíveis



Fonte: O Autor

4.3.3 Poda

A etapa de poda é necessária porque o WordNet não é um dicionário apropriado para o domínio de TI, nem um léxico de sentimento. Portanto, nem todos os *tokens* retornados pelo passo de expansão são relevantes para nossos propósitos. Por exemplo: ao expandir o *token* “prejudicar#v”, um possível sinônimo é “matar#v”, que é um termo comum em TI (por exemplo, em “matar um processo”), o qual não possui sentimento relacionado (ou seja, um termo neutro). Assim, recuperamos a polaridade associada ao *token* expandido no SentiWordNet e descartamos o *token* expandido se uma das seguintes condições for verificada:

- a) O *token* não está presente no léxico do sentimento ou tem uma pontuação de polaridade igual a 0;
- b) O *token* expandido tem uma polaridade inversa em comparação com a semente original (por exemplo, “lentamente#r” é uma semente negativa, mas ao buscar seus sinônimos no WordNet, o *token* “levemente#r” é encontrado, o qual é positivo no domínio de TI), o que faz referência à questão das polaridades opostas que foram feitas no uso SentiWordNet no domínio de TI, apresentado no início desta seção; ou
- c) O *token* expandido tem uma pontuação de polaridade entre -0,1 e 0,1, o que exclui termos com sentimento muito fraco (GODBOLE; SRINIVASIAH; SKIENA, 2007).

Como resultado, o DD tem entradas geradas a partir das sementes que foram escolhidas após a análise do corpus, as quais apresentam sentimentos em tíquetes. Seu uso no processo de análise de sentimentos será discutido na próxima seção.

4.4 Análise de Polaridade

4.4.1 Pré-Processamento

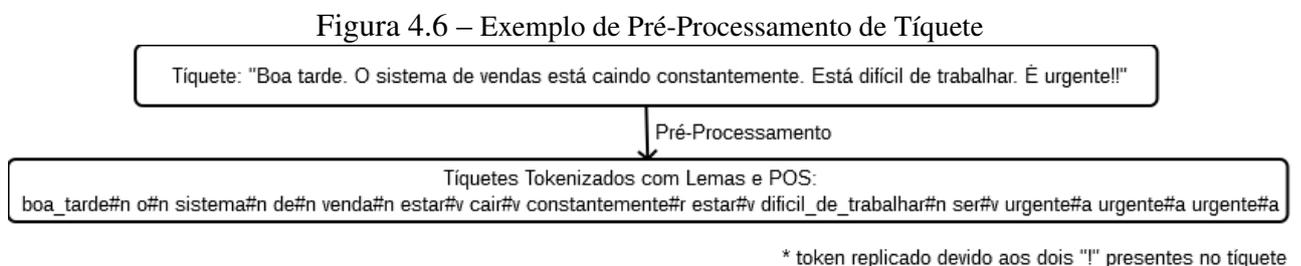
O pré-processamento é o primeiro passo do processo (Figura 4.2), necessário para derivar de cada tíquete uma lista de *tokens*, com o respectivo POS e lema, para ser usado como chave para pesquisa no DD e no SentiWordNet. O método pressupõe textos em Inglês, considerando os léxicos e ferramentas de PLN utilizados. Assim, devem ser traduzidos se estiverem em outro idioma.

Primeiramente, ações básicas de normalização são realizadas (por exemplo, remoção de quebras de linha e de caracteres especiais), de modo que o texto possa ser tokenizado corretamente. O texto após a normalização contém apenas caracteres alfanuméricos, espaços em branco e pontuação. Espaços em branco e pontuação são usados para dividir o texto em *tokens*. As sequências de *tokens* que correspondem a uma expressão no DD são combinadas em um único *token* (e.g., “bom_dia”). Os textos foram tokenizados usando a API Java `java.text.BreakIterator`³.

Em seguida, cada *token* é lematizado e associado a sua POS. O resultado é um conjunto de *tokens* no formato <lema>#<POS>. Para esse processo, utilizamos o TT4J⁴, uma implementação Java do TreeTagger⁵.

Os sinais de exclamação são indicadores de intensidade para sentimentos positivos/negativos (MOHAMMAD, 2016). Como passo final do pré-processamento, sempre que um *token* for seguido por um ponto de exclamação, ele é adicionado duas vezes à lista de *tokens* de um tíquete. Se há mais de um sinal de exclamação (e.g. “!!”, “!!!!!!”), o *token* é adicionado tantas vezes quanto aparecer.

A Figura 4.6 exibe um exemplo de pré-processamento de um tíquete.



Fonte: O Autor

4.4.2 Métodos para Classificação de Polaridade

4.4.2.1 Método de Dicionário

O Método de Dicionário (MD), para o qual o pseudocódigo é fornecido no Algoritmo 1, é baseado puramente em dicionários. Para cada *token* de um tíquete, o método procura se esse termo existe no DD, usando o lema e o POS do *token* como chave para a pesquisa (linha 5). Se

³<http://docs.oracle.com/javase/tutorial/i18n/text/about.html>

⁴<https://reckart.github.io/tt4j/>

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

for encontrado, é um termo candidato para sentimento; caso contrário, é um termo neutro no domínio. As linhas 6 a 15 lidam com a polaridade e serão detalhadas abaixo. As linhas 16-17 referem-se à classificação de emoção, e serão discutidas na Seção 4.5.

```

1: tokensClassificados ← inicializarNovaLista()
2: pontuacaoPolaridade ← null
3: enquanto tiqueteTokenizado.temProximo() fazer
4:   token ← tiqueteTokenizado.proximo()
5:   se DicionarioDominio.contem(token.lema,token.pos) entao
6:     se SentiWordNet.contem(token.lema,token.pos) e SentiWordNet.buscarPontuacaoPolaridade(token.lema,token.pos) != 0 entao
7:       pontuacaoPolaridade ← SentiWordNet.buscarPontuacaoPolaridade(token.lema,token.pos)
8:     else
9:       tokenDicionarioDominio ← DicionarioDominio.buscarToken(token.lema,token.pos)
10:      se tokenDicionarioDominio.orientacaoPolaridade == "POSITIVA" entao
11:        pontuacaoPolaridade ← PolaridadePadrao.positiva
12:      else
13:        pontuacaoPolaridade ← PolaridadePadrao.negativa
14:      fim se
15:    fim se
16:    pontuacoesEkman ← Nrc.buscarToken(token.lema)
17:    pontuacoesVad ← Anew.buscarToken(token.lema)
18:    tokensClassificados.adicionar(token,pontuacaoPolaridade,pontuacoesEkman,pontuacoesVad)
19:  fim se
20: fim enquanto
21: retornar tokensClassificados

```

Algoritmo 1: Pseudocódigo do Método de Dicionário

Para atribuir uma pontuação de polaridade, o algoritmo busca essa chave no SentiWordNet (linha 6): se encontrada, recupera a pontuação correspondente (linha 7). Caso contrário (ou seja, não está incluído no SentiWordNet), o método determina automaticamente uma pontuação. Essa pontuação é referida como *polaridade padrão* (linhas 10-14). Note que este caso ocorre apenas para termos fornecidos como semente, dado que na etapa de poda, os *tokens* gerados pela expansão só serão incluídos no DD se tiverem um sentimento associado no SentiWordNet, conforme descrito na Seção 4.3.

Devido à dificuldade de definir um valor de polaridade padrão manualmente, experimentamos diferentes maneiras de usar os valores disponíveis no SentiWordNet para realizar o cálculo automaticamente. Utilizamos como base do cálculo um conjunto de *tokens* do DD também encontrados no SentiWordNet. Experimentamos o uso da mediana e da média para calcular a polaridade padrão. Assim, a polaridade padrão é dada pela média/mediana de todos os *tokens* positivos/negativos com pontuação diferente de zero presentes no DD. Existe uma polaridade padrão negativa e outra positiva. Portanto, os *tokens* do SentiWordNet são agrupados conforme a sua orientação de polaridade (positiva/negativa), e a média/mediana de cada grupo calculada.

Também criamos uma estratégia complementar, referida como *espelhamento*. Com o uso do espelhamento, existe uma única polaridade padrão calculada sobre o conjunto de *tokens* com pontuação de sentimento presentes no DD, tal como extraídas do SentiwordNet. Ou seja, os *tokens* não são agrupados conforme a sua orientação de polaridade (positiva/negativa). O

cálculo é feito igualmente com média ou mediana, de acordo com a medida adotada, mas é utilizado o módulo das pontuações negativas.

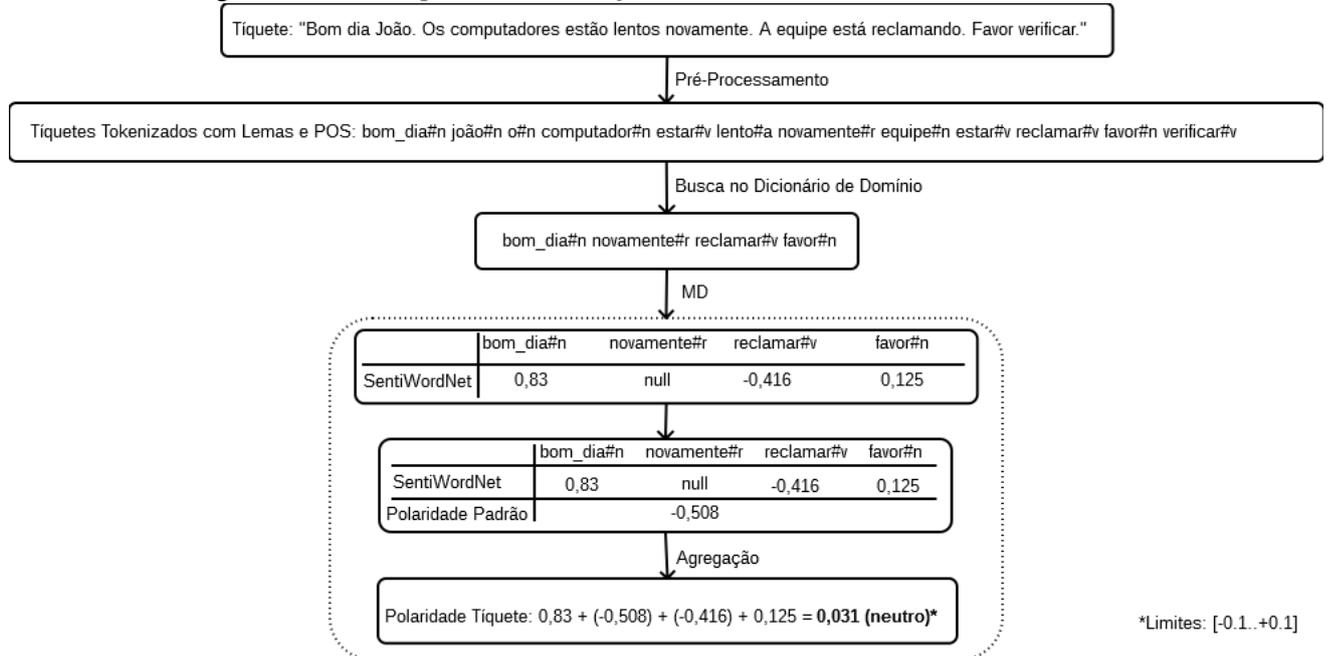
A Tabela 4.3 ilustra o cálculo da polaridade padrão conforme as estratégias adotadas. Nesse exemplo fictício, o DD seria composto apenas pelos seis *tokens* presentes na tabela. Assim, o MD definiria um valor para a polaridade padrão que seria calculado a partir da escolha entre média/mediana, e do uso ou não do espelhamento. As quatro combinações possíveis das duas estratégias estão na coluna *Estratégias p/ Calcular Polaridade de Categoria*. No Capítulo 5, são demonstrados testes e comparações realizadas com as possíveis variações.

Token DD	Polaridade	SentiWordNet	Estratégias p/ Calcular Polaridade Padrão			
			Com Espelhamento		Sem Espelhamento	
			Média	Mediana	Média	Mediana
reclamar#v	Negativa	-0,416	0,508/-0,508	0,459/-0,459	-0,562	-0,562
difícil#a	Negativa	-0,708				
novamente#r	Negativa	Ausente				
favor#n	Positiva	0,125			0,471	0,459
bom_dia#n	Positiva	0,83				
prezado#a	Positiva	0,459				

Tabela 4.3 – Exemplo de Cálculos da Polaridade Padrão

A Figura 4.7 ilustra as principais etapas do método no tocante à polaridade. Inicialmente os tíquetes são pré-processados, e *tokens* relevantes encontrados (i.e. palavras/expressões). Para cada *token* do texto, usa-se o DD para determinar se o mesmo será considerado, ou descartado. O exemplo utiliza o mesmo DD fictício da Tabela 4.3, e quatro *tokens* foram encontrados no DD. Para estes *tokens*, consulta-se o SentiWordNet para extrair as respectivas pontuações. Os *tokens* “bom dia”, “reclamar” e “favor” estão presentes no SentiWordNet, e suas pontuações são recuperadas. Já o *token* “novamente” não é encontrado, e portanto lhe é atribuída a polaridade padrão, que no exemplo foi calculado usando a estratégia de espelhamento e utilizando a média. Por fim, a agregação é realizada para determinar a polaridade, resultando em um tíquete neutro.

Figura 4.7 – Exemplo de Classificação de Sentimentos do Método de Dicionário



Fonte: O Autor

4.4.2.2 Método de Estrutura

Muitos tíquetes seguem um modelo semelhante aos emails (Figura 4.1). O Método de Estrutura (ME) explora a estrutura do documento (KOBAYASHI; INUI; MATSUMOTO, 2007; ORTU et al., 2015). A estratégia subjacente ao ME é que *tokens* na mesma posição no tíquete recebem uma pontuação comum, chamada de *polaridade de categoria*. A combinação de polaridade e da posição resulta em seis categorias: Saudação Positiva (SP), Saudação Negativa (SN), Relato Positivo (RP), Relato Negativo (RN), Conclusão Positiva (CP) e Conclusão Negativa (CN). Os *tokens* de saudação estão presentes na posição inicial do tíquete, os de conclusão no segmento final e os *tokens* de relato estão localizados no meio.

O Algoritmo 2 apresenta o pseudocódigo para o método ME. Conforme explicado na Seção 4.3.2, algumas sementes usadas para criar o DD são classificadas como delimitadores de início/fim, além de orientação de polaridade positiva/negativa, informação esta que é herdada por todos os termos resultantes da etapa de expansão. Assim, o primeiro passo no ME é analisar esta estrutura (linha 2): primeiro, todos os *tokens* iniciais e finais são localizados na lista de *tokens*, onde o último *token* inicial delimita o *segmento de saudação* e o primeiro *token* final delimita o *segmento de conclusão*. O texto restante entre esses delimitadores é considerado como o *segmento de relato*.

Em seguida, o método itera sobre a lista tokenizada recebida como entrada e procura cada *token* no DD, com base no respectivo lema e POS (linha 6). Para cada *token* encontrado,

```

1: tokensClassificados ← newinicializarNovaLista()
2: tiqueteSeparadoEstrutura ← Estrutura.separarTiquete(tiqueteTokenizado)
3: enquanto tiqueteTokenizado.temProximo() fazer
4:   token ← tiqueteTokenizado.proximo()
5:   pontuacaoPolaridade ← null
6:   se DomainDictionary.contem(token.lemma,token.pos) entao
7:     tokenDicionarioDominio ← DomainDictionary.buscarToken(token.lemma,token.pos)
8:     orientacao ← tokenDicionarioDominio.orientacaoPolaridade
9:     posicao ← tiqueteSeparadoEstrutura.buscarPosicaoToken(token)
10:    pontuacaoPolaridade ← PolaridadeCategoria.buscarPontuacaoPolaridade(orientacao,posicao)
11:    tokensClassificados.add(token,pontuacaoPolaridade)
12:  fim se
13: fim enquanto
14: retornar tokensClassificados

```

Algoritmo 2: Pseudocódigo do Método de Estrutura

identifica a respectiva orientação de polaridade (linha 8) e sua posição dentro do tíquete (linha 9) e atribui uma das seis polaridades de categorias correspondente (linha 10); Caso contrário, o *token* é considerado neutro.

Assim, como a polaridade padrão, a pontuação de cada uma das seis polaridades de categorias é calculada automaticamente considerando as pontuações de polaridade de *tokens* do SentiWordNet que estão presentes no DD. Os *tokens* no DD classificados como delimitadores iniciais e como delimitadores finais fornecem pontuações para as categorias de saudação (SP e SN) e de conclusão (CP e CN), respectivamente. Os *tokens* restantes são usados para calcular pontuações para as categorias de relatos (RP e RN).

Da mesma forma que no MD, pode ser utilizada para o cálculo a média ou a mediana, bem como ser considerado ou não o *espelhamento*. Sem o uso do espelhamento, os *tokens* positivos são usados para as categorias SP, CP e RP, e os negativos, para as categorias SN, CN e RN. Ao usar a estratégia de espelhamento, essa distinção não é feita por polaridade, e são considerados os módulos dos valores negativos no cálculo. No Capítulo 5, comparamos os resultados das diferentes estratégias.

A Tabela 4.4 ilustra os cálculos das polaridades de categorias conforme as estratégias adotadas. A tabela é a mesma do exemplo exibido na Tabela 4.3 e representa um DD fictício, composto apenas por seis *tokens*. O ME definiria as polaridades de categorias a partir do uso/não uso de espelhamento, e da escolha entre média/mediana. As quatro combinações possíveis das duas estratégias estão na coluna *Estratégias p/ Calcular Polaridade Padrão*. No Capítulo 5, são demonstrados testes e comparações realizadas com as variações dessas duas estratégias.

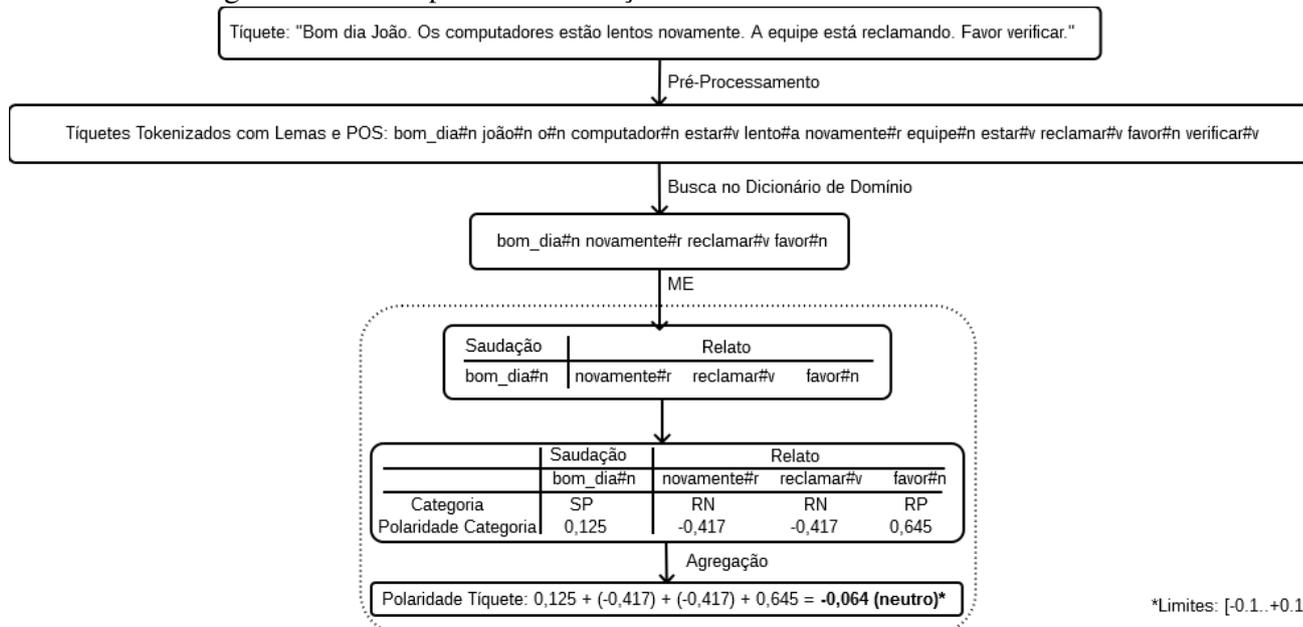
A Figura 4.8 apresenta o mesmo tíquete da Figura 4.7 e o DD da Tabela 4.4, agora classificado com o ME. Para cada *token* que está presente no DD, é identificada sua posição. O *token* “bom dia” é classificado na área de saudação, enquanto que os outros três são classificados no relato. Então, também associando às orientações de polaridade (positiva/negativa), são classificados em uma categoria de polaridade. Assim, o *token* “bom dia” fica na categoria

SP, “novamente” e “reclamar” na RN e “favor” na RP. Cada categoria possui uma pontuação de polaridade, previamente calculada usando a estratégia de espelhamento e utilizando a média como medida. Finalmente, a agregação é realizada, resultando em um tíquete neutro.

Token DD	Polaridade	Delimitador	SentiWordNet	Categoria	Estratégias p/ Calcular Polaridade de Categoria			
					Com Espelhamento		Sem Medida	
					Média	Mediana	Média	Mediana
reclamar#v	Negativa	Não	-0,416	RN	0,417/-0,417	0,416/-0,416	-0,562	-0,562
difícil#a	Negativa	Não	-0,708	RN				
novamente#r	Negativa	Não	Ausente	RN			0,125	0,125
favor#n	Positiva	Não	0,125	RP	0,645/-0,645	0,645/-0,645	0,645	0,645
bom_dia#n	Positiva	Inicial	0,83	SP				
prezado#a	Positiva	Inicial	0,459	SP				

Tabela 4.4 – Exemplo de Cálculos da Polaridades de Categoria

Figura 4.8 – Exemplo de Classificação de Sentimentos do Método de Estrutura



Fonte: O Autor

4.4.2.3 Método Híbrido

O Método Híbrido (MH) combina propriedades dos dois métodos anteriores. O Algoritmo 3 mostra seu pseudocódigo. Como em MD, o DD é usado para detectar uma palavra de sentimento de candidato (linha 6), e o SentiWordNet é usado para atribuir uma pontuação positiva/negativa (linha 8). Se o *token* não for encontrado no SentiWordNet, o *token* será classificado de acordo com sua posição e orientação de polaridade, de acordo com uma das seis polaridades de categorias, como no ME (linhas 10 a 13).

A Figura 4.9 exemplifica a classificação de um tíquete pelo MH. A diferença para o MD, é que o *token* “novamente”, que não está presente no SentiWordNet, tem sua pontuação de polaridade atribuída conforme sua posição no tíquete e sua orientação de polaridade, da mesma forma que é realizado no ME. Os outros *tokens* têm sua polaridade definida conforme as regras do MD. Diferente do MD e do ME, o mesmo tíquete agora foi classificado como positivo.

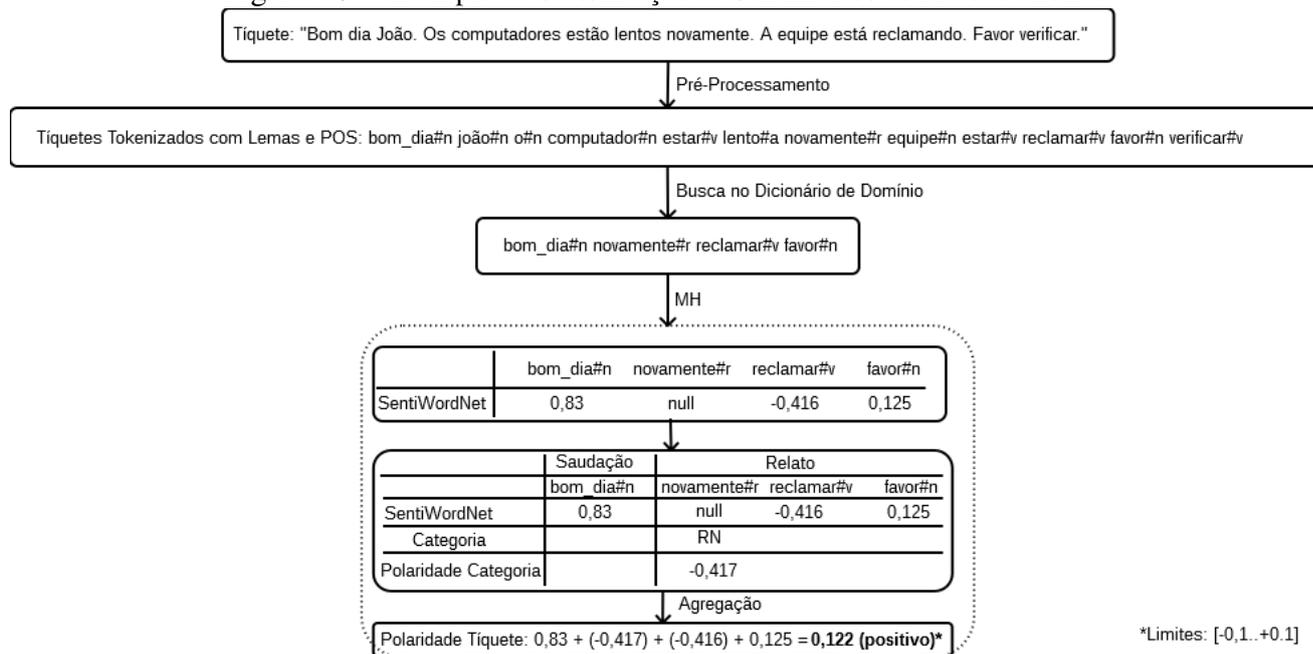
```

1: tokensClassificados ← newinicializarNovaLista()
2: pontuacaoPolaridade ← null
3: tiqueteSeparadoEstrutura ← Estrutura.separarTiquete(tiqueteTokenizado)
4: enquanto tiqueteTokenizado.temProximo() fazer
5:   token ← tiqueteTokenizado.proximo()
6:   se DomainDictionary.contains(lema,pos) entao
7:     se SentiWordNet.contem(token.lema,token.pos) e SentiWordNet.buscarPontuacaoPolaridade(token.lema,token.pos) != 0 entao
8:       pontuacaoPolaridade ← SentiWordNet.buscarPontuacaoPolaridade(token.lema,token.pos)
9:     else
10:      tokenDicionarioDominio ← DomainDictionary.buscarToken(token.lema,token.pos)
11:      orientacao ← tokenDicionarioDominio.orientacaoPolaridade
12:      posicao ← tiqueteSeparadoEstrutura.buscarPosicaoToken(token)
13:      pontuacaoPolaridade ← PolaridadeCategoria.buscarPontuacaoPolaridade(orientacao,posicao)
14:     fim se
15:     tokensClassificados.add(token,pontuacaoPolaridade)
16:   fim se
17: fim enquanto
18: retornar tokensClassificados

```

Algoritmo 3: Pseudocódigo do Método de Híbrido

Figura 4.9 – Exemplo de Classificação de Sentimentos do Método Híbrido



Fonte: O Autor

4.4.3 Agregação das Pontuações de Polaridade

O passo final agrega as pontuações de polaridade e de emoções individuais dos *tokens*, a fim de estabelecer os valores dos *tokens*.

Para polaridade, as pontuações individuais são somadas para calcular a polaridade total do tíquete. Dados limites positivos/negativos (TSYTSARAU; PALPANAS, 2012), se o resultado for inferior ao limite negativo, o tíquete é negativo; se for superior a um limiar positivo, positivo. Caso contrário, é neutro, o que significa que ele contém apenas palavras neutras em relação ao domínio de TI, ou termos contraditórios que se anulam de tal forma que não é possível detectar uma tendência.

4.5 Análise de Emoções

Há variados modelos estudados para emoções, portanto uma das primeiras questões é qual(is) modelo(s) escolher. A ausência de ferramentas para classificação de emoções também é um empecilho. Os modelos de Ekman e de VAD possuem léxicos gerados e utilizados em outros trabalhos: o NRC e o ANEW, respectivamente. Esses dois léxicos estão disponíveis para pesquisa, então optamos por adotá-los neste trabalho.

4.5.1 Classificação de Emoções

Quanto às abordagens dos métodos de classificação de polaridade criados neste trabalho (MD, ME, MH), apenas o MD se enquadra para classificação de emoções. As abordagens que utilizam estruturas (ME, MH) foram desconsideradas. Para o modelo VAD, esta decisão se justifica pela necessidade de atribuir pesos conforme o segmento estrutural do tíquete, os quais não teríamos como validar. No caso do modelo de Ekman, o dicionário NRC apresenta para cada uma de suas entradas valores discretos (binários, no caso) para cada emoção, isto é, cada *token* está associado (ou não) a cada uma das seis emoções. Portanto, independente da posição do termo no tíquete, não há como apresentar uma pontuação “maior” ou “menor” de uma emoção: ou o termo está ou não está associado a essa emoção.

As linhas 16-17 do algoritmo MD (Algoritmo 1) mostram a classificação de emoções. Para os *tokens* associados a uma polaridade, são pesquisadas também as possíveis emoções que possam representar. Para as emoções de Ekman, é verificada a existência do lema no léxico NRC (linha 16): caso exista, significa que o termo apresenta pelo menos uma das seis emoções básicas, e todos os rótulos de emoções correspondentes são recuperados. Para o modelo VAD, uma busca na expansão do ANEW é realizada também utilizando o lema (linha 17): caso seja encontrado um registro nesse dicionário, os três valores VAD são recuperados.

Ao contrário do que é feito com *tokens* com polaridade zero ou muito fraca, *tokens* sem emoção (independente do modelo) não são descartados, pois emoções são consideradas informações complementares.

Todos os *tokens* presentes no NRC e do ANEW são compostos por apenas uma palavra, enquanto que o DD possui expressões com mais de uma palavra. Algumas dessas expressões do DD são muito comuns em tíquetes (e.g. “bom dia”, “aguardo retorno”) ou tem um impacto grande na classificação de sentimentos (e.g. “impossível de usar”), então desconsiderá-las poderia prejudicar a análise. Portanto, na busca por emoções de um *token* no NRC e no ANEW, algumas expressões são tratadas de forma especial: a busca nos léxicos é realizada considerando como chave apenas o adjetivo presente em cada expressão. Por exemplo: para os *tokens* “bom dia” e “impossível de usar”, o método busca as emoções das palavras “bom” e “impossível”, respectivamente.

4.5.2 Agregação de Emoções

Para cada uma das seis emoções de Ekman, a agregação é o resultado da disjunção dos valores dessa emoção de todos os *tokens* do texto, já que esses valores são binários, ou seja, o *token* apresenta ou não essa emoção. Assim o resultado também é binário, e o texto apresenta ou não cada uma das seis emoções, independente de quantos vezes ela se repita.

Para cada uma das três dimensões de VAD, calcula-se a média aritmética dos valores dos *tokens* que apresentam esse modelo de emoção, determinando o valor final do texto. O resultado é classificado conforme os seguintes intervalos, permitindo rotular de forma discreta o texto em cada uma dessas dimensões: [0,4) como baixo, [4,6] como neutro e (6,10] como alto. Esses limites foram estabelecidos a partir do seguinte levantamento:

- a) foram pesquisados quais os termos que estão no léxico ANEW e que têm os valores nas três dimensões VAD no intervalo [4,6];
- b) dentre esses termos, verificou-se quais eram os dez mais presentes no corpus. O conjunto resultante corresponde a 45% de todos os termos retornados em a);
- c) estes termos foram analisados subjetivamente, e constatou-se que todos são neutros em tíquetes. Exemplos destes termos são: “requisição”, “máquina”, “rede”, “tela” e “aplicação”.

A partir desse levantamento, estabelecemos os limites [4,6] como neutros para as três dimensões VAD, a fim de evitar distorcerem que poderiam ser provocadas por termos neutros no contexto de tíquetes.

4.6 Prova de Conceitos

Como prova de conceitos (POC), foi desenvolvida uma aplicação que provê funcionalidades de suporte a todo processo de análise de sentimentos descrito neste trabalho. O propósito dessa POC é demonstrar que sentimentos também podem ser um critério para monitoração de tíquetes, junto com informações mais clássicas como criticidade, urgência ou tempo (SAHIBUDIN; SHARIFI; AYAT, 2008). Assim, é possível obter um *feedback* antecipado a respeito da qualidade de serviços de TI prestados já na própria demanda, o que pode ser útil para avaliar mudanças recentes em sistemas, infraestrutura ou na equipe/área de TI. O sentimento também pode ser um novo critério para priorização de atendimento de tíquetes. Por exemplo, tíquetes

muito negativos e/ou que apresentam a emoção de raiva podem ser priorizados, dado que é possível reconhecer o quanto o problema afeta o usuário.

As funcionalidades da POC podem ser agrupadas em seis grandes grupos:

- Importação e cadastro de tíquetes;
- manipulação dos dicionários (e.g. visualização e inserção de sementes, quantificações de *tokens*, etc).
- Escolha do método de classificação de sentimento, de acordo com um dos métodos propostos;
- Filtros para seleção de tíquetes com características específicas (e.g. projeto, data, polaridade, etc);
- Sumarizações e recursos de visualização para os sentimentos dos tíquetes (e.g. quantidade de tíquetes de acordo com filtros, evolução do sentimento no tempo, distribuições dos sentimentos, etc);
- Visualização de gráficos contendo informações consolidadas de sentimentos identificados nos tíquetes processados.

A Figura 4.10 apresenta a tela principal da POC. A interface é dividida em painéis, representando as diferentes funcionalidades, as quais serão discutidas em mais detalhes no restante desta seção. Na figura, cada painel foi identificado com uma letra (em vermelho), referenciada nas explicações abaixo.

Quanto à arquitetura, a POC foi desenvolvida utilizando a linguagem de programação Java⁶ (versão 1.6) e o banco de dados relacional⁷ (versão 5.6.17). É um sistema web desenvolvido em três camadas: uma camada de apresentação, uma de negócios e uma de acesso/repositório de dados. Os seguintes frameworks para Java foram usados: Primefaces⁸ (versão 6.0) para a camada de apresentação, Java Server Faces⁹ (versão 2.2) para as camadas de apresentação e de negócios, e Java Database Connectivity¹⁰ (versão 5.1.25) para acesso ao banco de dados. As demais ferramentas já informadas e necessárias para o desenvolvimento das técnicas deste trabalho também foram importadas.

⁶<https://www.java.com>

⁷<https://www.mysql.com/>

⁸<https://www.primefaces.org/>

⁹<http://www.oracle.com/technetwork/java/javaee/javaserverfaces-139869.html>

¹⁰<http://www.oracle.com/technetwork/java/javase/jdbc/index.html>

Figura 4.10 – Tela Principal da POC



4.6.1 Funcionalidades de Cadastro e Importação de Tíquetes

Este conjunto de funcionalidades visa a importação dos tíquetes para o sistema de análise de sentimentos. Em uma situação real, ele seria alimentado continuamente através da integração com o sistema de tíquetes da organização. No contexto restrito desta POC, ele permite o cadastro de fontes de tíquetes e a inserção manual ou importação de uma base de tíquetes. Essas funcionalidades correspondem aos painéis A e B da Figura 4.10, também exibidos de forma aumentada na Figura 4.11.

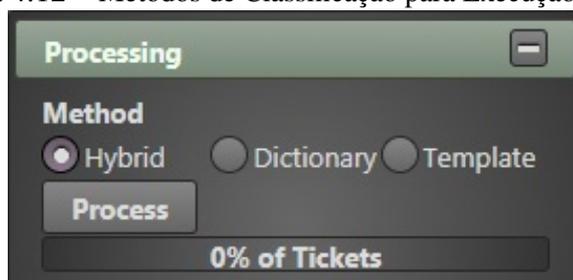
O painel A (*Sources*) permite o cadastro de fontes de tíquetes. O painel B (*Tickets*) permite o cadastro de tíquetes. Devem ser informados a data de abertura do tíquete, o texto descrevendo a demanda, e a qual fonte o tíquete pertence. Pode ser importada uma base contendo inúmeros tíquetes, via arquivo no formato CSV (*Comma-Separated Values*). Os tíquetes cadastrados na POC são utilizados na execução dos métodos de classificação, representado pelo painel D (*Processing*) da Figura 4.10 e pela Figura 4.12. A POC foi previamente carregada com os tíquetes presentes no *gold standard*.

Figura 4.11 – Cadastro de Fontes e Tíquetes na POC

The image shows a software interface with two main panels. The top panel, titled 'Sources', has a 'Name:' input field with an 'Insert' button next to it. To the right is a list of five sources, each with a trash icon: 'Insurance - Internal', 'Research - External', 'Software House - External', 'Software House - Internal', and 'Telecom - Internal'. The bottom panel, titled 'Tickets', has a 'Date:' input field with a calendar icon, a 'Text:' input field, and a radio button selection for the source. The radio buttons are: 'Insurance - Internal', 'Telecom - Internal', 'Software House - External', 'Research - External', and 'Software House - Internal'. At the bottom of the 'Tickets' panel are three buttons: 'Submit', '+ Upload File', and 'List Tickets'.

Fonte: O Autor

Figura 4.12 – Métodos de Classificação para Execução na POC



Fonte: O Autor

4.6.2 Manipulação dos Dicionários

As funcionalidades do painel E (*Domain Dictionary*) da Figura 4.10, também ilustrado na Figura 4.13, permitem a manipulação dos dicionários. O DD da POC é utilizado para a execução dos métodos de classificação na POC, representado pelo painel D (*Processing*).

O DD da POC foi importado com os registros gerados nos experimentos realizados no Capítulo 5. Os dicionários SentiWordNet, NRC e ANEW também foram importados para a POC, nas suas respectivas últimas versões existentes durante o desenvolvimento deste trabalho.

Esse painel permite a inserção de *tokens* semente no formato <lema>#<POS> no DD da POC. Podem ser inseridos *tokens* compostos, i.e. com mais de uma palavra. Quando um *token* semente é cadastrado, é realizado o processo seu processo expansão (limite de iterações=1) e poda, conforme detalhado na Seção 4.3. Junto a cada *token* devem ser informados a orientação de polaridade (positiva/negativa) e se é um delimitador (final/inicial), informações presentes em todos os *tokens* do DD.

Um *token* cadastrado nesse painel também pode ser inserido nos dicionários SentiWordNet, NRC e ANEW da POC. Essa ação tem impacto direto nos resultados de métodos de classificação de sentimentos executados na POC, portanto deve ser realizada com cautela. Quando são informados valores de polaridade e de emoção dos modelos Ekman e VAD no formulário de cadastro do *token* no DD, esses valores são inseridos nos dicionários SentiWordNet (polaridade), NRC (emoções Ekman) e ANEW (emoções VAD) importados na POC, além da inserção do próprio *token* no DD. Para polaridade, o valor corresponde à pontuação de polaridade. Para Ekman, pode ser informado se o *token* possui ou não as seis emoções básicas. Para VAD, valores contínuos para as três dimensões podem ser informados.

Figura 4.13 – Manipulação do Dicionário de Domínio na POC

The screenshot shows a 'Domain Dictionary' window with the following configuration options:

- Token:** A text input field containing 'e.g. 'hate#v', 'hard#a to use#v''.
- Sentiment:**
 - Null Values
 - Score:**
 - Default
 - Value: from -1.000 to
 - Polarity:**
 - Positive
 - Negative
 - Delimiter:**
 - No
 - Initial
 - Final
- NRC:**
 - Null Values
 - Anger:** No Yes
 - Disgust:** No Yes
 - Fear:** No Yes
 - Joy:** No Yes
 - Sadness:** No Yes
 - Surprise:** No Yes
- VAD:**
 - Null Values
 - VAD:**
 - Valence:** 5.00
 - Arousal:** 5.00
 - Dominance:** 5.00

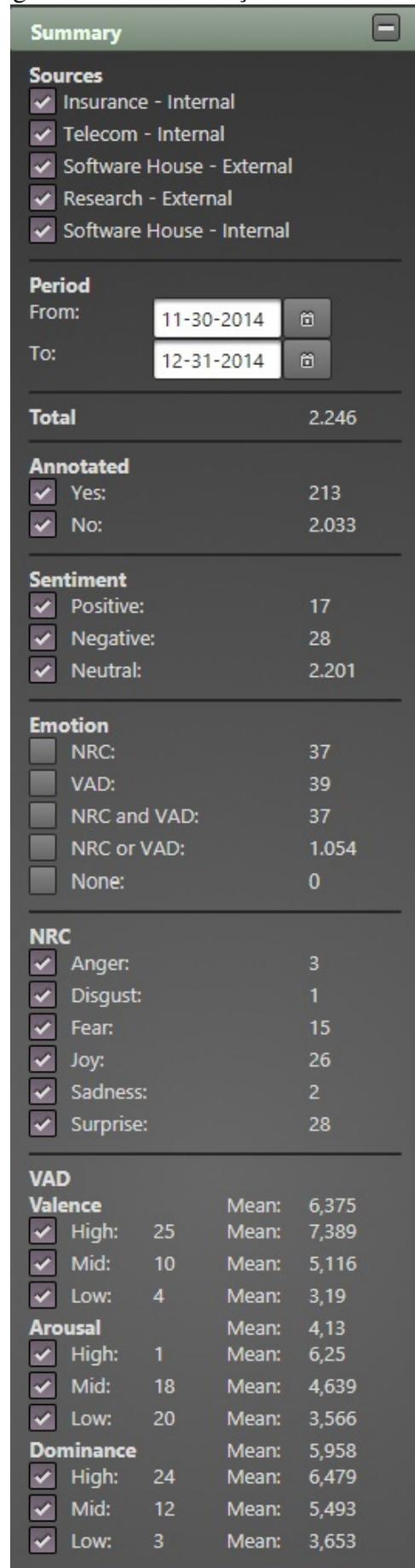
At the bottom, there are two buttons: 'Submit' and 'List Tokens'.

Fonte: O Autor

4.6.3 Método de Classificação

É possível classificar o sentimento segundo os três métodos de classificação desenvolvidos neste trabalho (MD, ME e MH). A seleção do método e o início de sua execução são possíveis pelo painel D (*Processing*) da Figura 4.10. Cada vez que um método é selecionado, todos tóquetes são (re)classificados, e todos os painéis que mostram informações sobre os tóquetes são atualizadas com estas novas classificações.

Figura 4.14 – Sumarização e Filtros na POC



4.6.4 Sumarizações e Filtros

Após a execução de um dos métodos de classificação, é possível visualizar uma sumarização das principais informações de sentimentos dos tíquetes. Também é possível filtrar os tíquetes conforme a seleção de algumas características/valores que esses tíquetes processados apresentam. Essa sumarização e filtros são apresentados no painel C da Figura 4.10 (*Summary*) e na Figura 4.14 de forma aumentada.

São exibidos o número total de tíquetes, quantos desses são anotados, quantos têm as polaridades positiva/negativa/neutra, quantos têm as emoções do modelo Ekman e/ou VAD, quantos possuem as seis emoções de Ekman e quantos possuem cada uma das três componentes VAD (categorizadas em alta/neutra/baixa).

Para filtrar os tíquetes considerados na exibição da POC, é possível selecionar as fontes dos tíquetes, um período de cadastro (entre duas datas), somente os anotados, polaridades específicas (positiva, negativa e/ou neutra), o modelo de emoções (Ekman e/ou VAD), e emoções específicas (as seis emoções de Ekman e cada uma das três componentes VAD). A alteração desses filtros impacta nos valores exibidos na sumarização desse neste mesmo painel, bem como nos gráficos exibidos nos painéis relativos à visualização (F, G, H, I, J, K, L e M).

Pode-se visualizar informações de tíquetes específicos, clicando no botão *List Tickets*, presente no painel B (*Tickets*). Um exemplo de listagem de tíquetes está presente na Figura 4.15. Para cada tíquete da lista, são apresentadas sua fonte, data, se é anotado, polaridade, pontuação de polaridade, a presença/ausência das seis emoções de Ekman e as pontuações das três dimensões VAD.

É possível inspecionar os detalhes de um tíquete específico. O texto de um tíquete desejado pode ser visualizado conforme exemplo da Figura 4.16. Também é possível visualizar as informações de cada um de seus *tokens* de acordo com o processo de classificação, exibido na Figura 4.17 para o tíquete selecionado na Figura 4.16. Para cada *token* do tíquete, são exibidas sua POS, frequência, polaridade, pontuação de polaridade (atribuída conforme o processo de classificação executado), a presença/ausência das seis emoções de Ekman (presentes no NRC) e as pontuações das três dimensões VAD (presentes no ANEW).

Figura 4.15 – Visualização de Tíquetes Filtrados

Text	Tokens	Source	Date	Annotated	Polarity	Score	Anger	Disgust	Fear	Joy	Sadness	Surprise	Valence	Arousal	Dominance
	ρ	Software House - External	12-19-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-19-2014	Yes	Positive	0.23	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-19-2014	Yes	Positive	0.07	No	No	No	Yes	No	Yes	7.64	4.29	6.63
	ρ	Telecom - Internal	12-18-2014	Yes	Positive	0.25	No	No	No	Yes	No	Yes	7.96	3.95	6.70
	ρ	Telecom - Internal	12-16-2014	Yes	Positive	0.03	No	No	No	Yes	No	Yes	7.83	3.50	6.65
	ρ	Software House - External	12-15-2014	Yes	Positive	0.39	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-13-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-12-2014	Yes	Positive	0.12	No	No	No	No	No	No	7.44	4.60	6.93
	ρ	Software House - External	12-11-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Telecom - Internal	12-10-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41

Fonte: O Autor

Figura 4.16 – Visualização do Texto de Tíquete Selecionado

Text	Tokens	Source	Date	Annotated	Polarity	Score	Anger	Disgust	Fear	Joy	Sadness	Surprise	Valence	Arousal	Dominance
	ρ	Software House - External	12-19-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-19-2014	Yes	Positive	0.23	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-19-2014	Yes	Positive	0.07	No	No	No	Yes	No	Yes	7.64	4.29	6.63
	ρ	Telecom - Internal	12-18-2014	Yes	Positive	0.25	No	No	No	Yes	No	Yes	7.96	3.95	6.70
	ρ	Telecom - Internal	12-16-2014	Yes	Positive	0.03	No	No	No	Yes	No	Yes	7.83	3.50	6.65
	ρ	Software House - External	12-15-2014	Yes	Positive	0.39	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-13-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Software House - External	12-12-2014	Yes	Positive	0.12	No	No	No	No	No	No	7.44	4.60	6.93
	ρ	Software House - External	12-11-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41
	ρ	Telecom - Internal	12-10-2014	Yes	Positive	0.33	No	No	No	Yes	No	Yes	7.89	3.66	6.41

Fonte: O Autor

4.6.5 Visualizações

A POC também exibe informações consolidadas a respeito dos tíquetes processados, bem como sobre os *tokens* encontrados nesses tíquetes. Sobre os *tokens*, são geradas diferentes nuvens de *tokens* a partir de diferentes características dos *tokens*. Quanto aos tíquetes, há diferentes gráficos consolidando informações sobre polaridade e emoções de Ekman e VAD.

As nuvens de *tokens* representam diferentes informações a respeito dos *tokens* presentes nos tíquetes processados. As nuvens de *tokens* estão representadas na Figura 4.18 e também nos painéis F, G, H, I e J da Figura 4.10. Apresentam, respectivamente, *tokens* do DD, *tokens* com polaridade, *tokens* com emoções de Ekman, *tokens* com emoções VAD e todos os *tokens* de

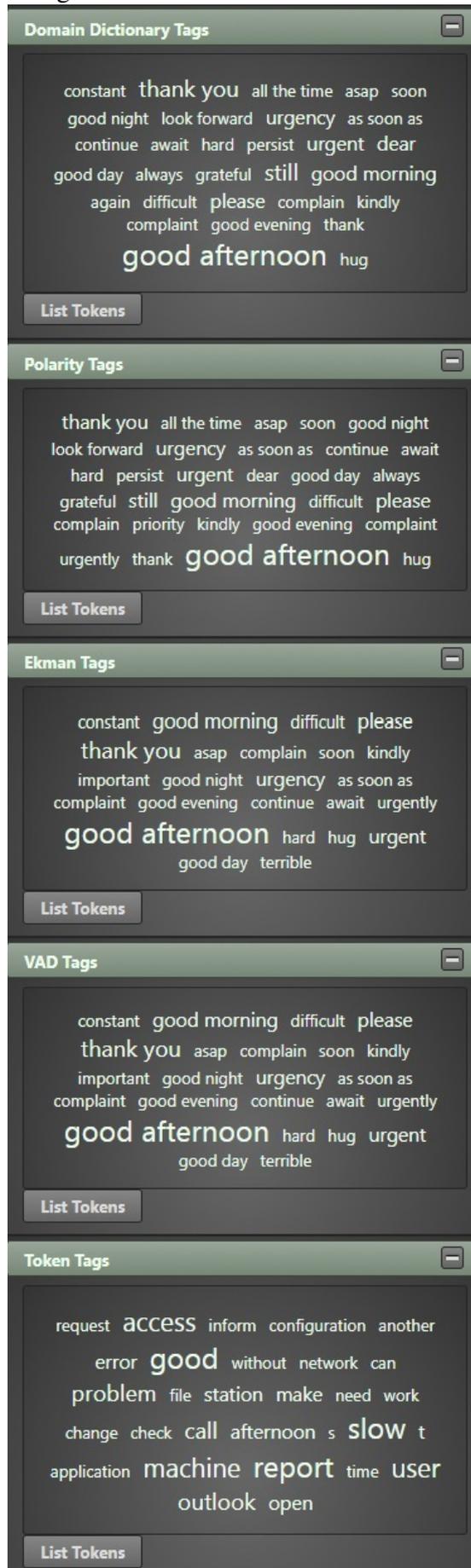
Figura 4.17 – Visualização dos Tokens de um Tíquete

Token	POS	Frequency	Polarity	Score	Anger	Disgust	Fear	Joy	Sadness	Surprise	Valence	Arousal	Dominance
good day	Noun	1	Positive	0.329	No	No	No	Yes	No	Yes	7.89	3.66	6.41
asap	Adverb	2	Negative	-0.459	No	No	No	No	No	No	NULL	NULL	NULL
please	Adverb	1	Positive	0.125	No	No	No	Yes	No	No	NULL	NULL	NULL
thank you	Noun	2	Positive	0.329	No	No	No	Yes	No	No	7.77	3.33	6.89

Fonte: O Autor

uma palavra encontrados nesses tíquetes (mesmo *tokens* que não presentes no DD). Cada nuvem apresenta os *tokens* mais frequentes. Quanto maior o tamanho da fonte do *token* exibido, mais frequente ele é conforme a característica que a nuvem representa. A Figura 4.19 e os painéis K, L e M da Figura 4.10 contêm gráficos que apresentam estatísticas sobre os tíquetes processados. O painel K (*Polarity Charts*) exibe gráficos contendo estatísticas de polaridade desses tíquetes. Um dos gráficos apresenta dados gerais, de todo o período filtrado, e outro exibe os dados por data. O painel L (*Ekman Charts*) exibe gráficos contendo estatísticas das emoções de Ekman desses tíquetes. Um dos gráficos apresenta dados gerais, de todo o período filtrado, e outro exibe os dados por data. Finalmente, o painel M (*VAD Charts*) exibe gráficos contendo estatísticas das emoções VAD desses tíquetes. Um dos gráficos apresenta dados gerais, de todo o período filtrado, outro demonstra histogramas com os valores contínuos de cada dimensão VAD em todo o período, e um terceiro exibe os dados por data.

Figura 4.18 – Nuvens de Tokens na POC



Fonte: O Autor

Figura 4.19 – Gráficos com Estatísticas de Sentimentos na POC



Fonte: O Autor

4.7 Considerações Finais

O principal objetivo deste trabalho, que é identificar sentimentos presentes em tíquetes para TI, foi atingido com as técnicas propostas na abordagem proposta neste trabalho.

Para separar o relato de problemas, característica inerente de tíquetes, de sentimentos usados ao descrevê-lo, um dicionário de domínio foi gerado especificamente para o domínio de tíquetes. Com isto, geramos melhores resultados de classificação, como será evidenciado em nossos experimentos, relatados no próximo capítulo. Os métodos propostos para classificação de polaridade foram desenvolvidos com sucesso, utilizando a abordagem de dicionários, popular na análise de sentimentos, e combinados com intuições sobre o papel da estrutura do tíquete. De forma complementar, a classificação de emoções também foi implementada. Desenvolvemos uma POC, mostrando funcionalidades que poderiam ser incorporadas em sistemas de tíquetes que contemplem sentimentos como critério de monitoração e priorização.

5 EXPERIMENTOS

Neste capítulo, serão descritos os experimentos realizados neste trabalho. O processo de anotação do corpus será detalhado, o qual serviu de base para os experimentos que compararam nossos métodos entre si e com as ferramentas do baseline. Diferentes limites de iterações no processo de expansão do DD também foram testados. O desempenho entre diferentes estratégias de cálculos para polaridade padrão e para polaridades de categorias são descritas. Por fim, os experimentos para identificar possíveis correlações entre polaridade e os dois modelos de emoção estudados neste trabalho são descritos.

5.1 Gold Standard para Polaridade

A fim de avaliar os resultados de nossas técnicas, criamos um *gold standard* a partir da anotação manual de uma amostra aleatoriamente selecionada do corpus descrito na Seção 4.1. A amostra consiste de 2.333 tíquetes (cerca de 6% do corpus).

A anotação foi realizada por três anotadores distintos: a) um membro de uma equipe de TI, graduado em computação (anotador A); b) uma pessoa que trabalha em uma empresa de TI, graduada em outra área (anotador B); e c) um usuário não técnico (C). A anotação envolveu exclusivamente a polaridade, dada a complexidade de anotação de emoções (conforme descrito na Seção 4.1).

Todos os anotadores anotaram cada tíquete da amostra como positivo, negativo ou neutro, em um processo semelhante ao relatado em Wiebe, Wilson e Cardie (2005), Tumitan e Becker (2014). Após a compilação dos resultados, a polaridade final de um tíquete foi determinada com o valor que havia sido estabelecido pela maioria (MOHAMMAD, 2016). Quando todos os anotadores discordaram, eles discutiram para chegar a um consenso.

Como proposto por Wiebe, Wilson e Cardie (2005), elaboramos um treinamento contendo três etapas: instruções de leitura da documentação, visualização de exemplos de tíquetes anotados e anotação supervisionada. Na primeira etapa, as principais instruções foram:

1. Não há regras fixas sobre como um tíquete deve ser anotado com base em seus *tokens*. Alguns exemplos serão discutidos, mas isso não significa que *tokens* específicos determinem uma polaridade;
2. Os tíquetes devem ser interpretados considerando o contexto de TI e com base no que está explicitamente escrito. Os anotadores não devem tirar textos fora do contexto e assumir

o que eles poderiam implicar;

3. Os anotadores devem ser consistentes em suas próprias anotações e com os exemplos que lhes são fornecidos para treinamento.

Na segunda etapa, foram discutidos dezoito exemplos selecionados do corpus, seis para cada polaridade (positivo, negativo e neutro). Estes tíquetes também foram classificados de acordo com a complexidade de determinação da polaridade: fácil, médio e difícil. Exemplos de tíquetes positivos estão ilustrados na Tabela 5.1.

Tabela 5.1 – Exemplos de Tíquetes Positivos para o Tutorial de Anotação

Dificuldade	Tíquete
Fácil	<i>“O problema de falta de energia foi resolvido e a performance do cluster é excelente!”</i>
Médio	<i>“Bom dia! Após completar o formulário de cadastro, estou enfrentando problemas de redirecionamento. Em anexo, seguem prints do erro. Obrigado.”</i>
Difícil	<i>“Prezados, quando eu tento baixar o backup, ocorre um erro. Ver arquivos em anexo. Favor priorizar.”</i>

Fonte: O Autor

Na última etapa do tutorial, os anotadores realizaram uma anotação supervisionada, ou seja, anotamos alguns tíquetes junto com os anotadores. Escolhemos dezoito tíquetes seguindo os mesmos critérios utilizados para os exemplos.

Após o treino, todos os anotadores receberam os tíquetes da amostra e fizeram suas anotações individualmente. As divergências entre os anotadores são apresentadas na Tabela 5.2. O nível de desacordo é comparável aos trabalhos relacionados (WIEBE; WILSON; CARDIE, 2005; TUMITAN; BECKER, 2014). A última linha (*Total*) mostra o número final de tíquetes positivos, negativos e neutros após a compilação de todas as polaridades dos tíquetes estabelecidas pelas três anotações.

Tabela 5.2 – Divergências entre as Anotações dos Anotadores

Anotador	Positivo	Negativo	Neutro	Disagreement
A	273	154	1.942	B: 12%
				C: 10%
				BC: 7%
B	358	172	1.803	A: 12%
				C: 21%
				AC: 7%
C	187	67	2.079	A: 10%
				B: 21%
				AB: 7%
Total	233	105	1.995	

Fonte: O Autor

5.2 Configuração dos Experimentos para Polaridade

5.2.1 Objetivos

Para análise de polaridade, os objetivos dos experimentos deste trabalho são:

- a) Avaliar o valor agregado por uma solução personalizada para análise de polaridade de tíquetes de TI. Comparamos nossos métodos com duas soluções populares no campo de engenharia de software, SentiWordNet e SentiStrength (Capítulo 2);
- b) Avaliar o papel desempenhado pela estrutura de tíquetes na análise dos polaridade. Comparamos os resultados produzidos pelo método baseado em dicionário (MD), com as duas alternativas que consideram a posição dos *tokens* nos tíquetes (ME e MH);
- c) Identificar uma estratégia adequada para atribuir automaticamente pontuações de polaridade aos *tokens* no domínio de TI. Experimentamos estratégias diferentes para determinar uma polaridade padrão para palavras estabelecidas como sementes no MD, bem como determinar polaridades de categorias quando consideramos a posição de *tokens* nos tíquetes (ME e MH).

Esses objetivos permitem avaliar o desempenho dos métodos criados especificamente para classificação de polaridade de tíquetes, comparando com ferramentas gerais para classificação de sentimentos disponíveis para pesquisa.

5.2.2 Limite de Iterações no Processo de Expansão do Dicionário de Domínio

Usando o *gold standard* e o corpus de TI, avaliamos o método de expansão proposto na Seção 4.3 experimentando diferentes limites de iterações. Utilizamos 181 sementes, que foram extraídas de duas fontes: a) uma lista de expressões de saudações e de conclusões, e b) análise manual dos 25% termos mais frequentes encontrados no *gold standard* para identificar aqueles com sentimento. Essas sementes são listadas na Tabela 5.3

A expansão resultou em 208, 224 e 240 *tokens* considerando os limites 1, 2 e 3, respectivamente. Contudo, o número de tíquetes presentes no *gold standard* que contêm os *tokens* gerados não aumentou significativamente nos níveis 2 e 3. Usando apenas as sementes, 959 tíquetes apresentam *tokens* presentes no DD. Considerando apenas os *tokens* gerados com a execução com o limite=1, o número foi de 232, o que é considerável. Já o limite=2 e o limite=3 geraram apenas 14 e 3, respectivamente. Assim, o DD utilizado nos experimentos descritos em seguida foi criado usando o limite=1.

5.2.3 Baselines para Polaridade

Os resultados considerados como baseline para polaridade foram produzidos usando uma variação de MD que utiliza apenas SentiWordNet e a ferramenta SentiStrength disponível para fins acadêmicos. Não foram adotados limites para a agregação das pontuações de polaridade positiva e negativa (i.e. um tíquete é considerado neutro apenas se o cálculo de sua polaridade resulta em 0), pois esse valor produziu os melhores resultados globais. A implementação que usa o SentiWordNet não emprega o DD, ou seja, todos os *tokens* são diretamente pesquisados no SentiWordNet e suas informações, quando disponíveis, são extraídas. A tokenização, o manuseio de pontos de exclamação e agregação são realizados conforme descrito no Capítulo 4. A ferramenta SentiStrength foi executada com a configuração padrão, que retorna uma pontuação de polaridade negativa e positiva para cada texto. Consideramos a polaridade geral do tíquete como a diferença entre as pontuações positiva e negativa.

Tabela 5.3 – Sementes do Dicionário de Domínio

absurd	excessively bad	horrific	thank you
access anything	excessively important	hug	thank you very much
again	excessively long	ill	thankful
all the best	excessively old	important	thankfully
all the time	excessively slow	impossible	tkx
all times	excessively slowly	impossible to use	tkx
always	excessively slowness	impossible to work	tolerate
amaze	extreme bad	inquire to request	too bad
appreciate	extreme important	keep	too important
as soon as	extreme long	kindly	too long
asap	extreme old	like to request	too old
ask to request	extreme slow	look forward	too slow
await	extreme slowly	love	too slowly
aware	extreme slowness	monstrous	too slowness
awful	extremely bad	nice	trick
befuddle	extremely important	nonsense	truly bad
beloved	extremely long	painful	truly important
best regards	extremely old	painful	truly long
best wishes	extremely slow	persist	truly old
bye bye	extremely slowly	please	truly slow
can you help	extremely slowness	prioritize	truly slowly
care to request	fearful	priority	truly slowness
cheers	fearsome	really bad	ugly
come back	felicitation	really important	unfortunately
complain	frequently	really long	unluckily
complaint	good afternoon	really old	unsafe
confound	good bye	really slow	urgency
confuse	good day	really slowly	urgent
congratulation	good evening	really slowness	urgently
constant	good morning	recur	very bad
constantly	good night	redo	very important
continue	goodbye	remain	very long
cordially	grateful	repulsive	very old
could you help	gratefully	rgds	very slow
dangerous	hard	risk	very slowly
dead	hazard	serious	very slowness
dear	heartily	several	wait
deplorable	hello	severe	warm regards
despairing	highly bad	sincerely	warmly
desperate	highly important	soon	will you help
detestable	highly long	standby	woeful
difficult	highly old	standing by	would you help
difficulty	highly slow	still	yet
dire	highly slowly	suffer	yours sincerely
direful	highly slowness	terrible	
excellent	horrible	thank	

5.2.4 Implementação dos Métodos de Classificação de Polaridade

Como discutido na Seção 4.4.2, foi proposto o uso da média e mediana para o cálculo de pontuação de polaridade e das pontuações de categoria. Também foi proposta a estratégia de espelhamento, a qual pode ser adotada ou não. A combinação das duas estratégias gera quatro combinações possíveis. Portanto, implementamos quatro variações de cada um dos algoritmos propostos (MD, ME e MH), de modo a comparar a melhor fórmula de cálculo.

Os três métodos foram executados em suas quatro variações existentes. Usando o *gold standard* como base, comparamos os resultados da classificação de polaridade utilizando as seguintes métricas: acurácia, precisão, revocação e medida-f. Um teste T bicaudal foi aplicado nessas métricas geradas para comparar os métodos e suas respectivas variações de implementação, com um nível de significância de 0,05.

5.3 Polaridade: Resultados e Discussões

5.3.1 Comparação com os Baselines

A Tabela 5.4 exibe as métricas para os métodos considerados como baseline, bem como para todas as quatro variações de MD, ME e MH. Os melhores resultados são destacados em negrito. Todas as métricas foram geradas considerando 0 como limite para distinguir os tíquetes neutros dos positivos/negativos.

É possível observar que qualquer implementação dos métodos propostos superam os baselines. Existe uma única exceção, que é a precisão dos tíquetes neutros. O SentiWordNet conseguiu uma precisão de 100% para tíquetes neutros, mas à custa de uma revocação muito baixa (16%), o que significa que quando um tíquete é classificado como neutro, é sempre correto (precisão), mas são reconhecidos apenas alguns tíquetes neutros como tal (revocação). Como consequência, a medida-f correspondente também é baixa (27%). A precisão alcançada por qualquer implementação de nossos métodos para tíquetes neutros é comparável ao SentiWordNet (98-99%), mas eles apresentam uma excelente revocação (96%). Assim, nossa solução sempre apresenta uma medida-f maior (97%). Considerando a medida-f, que combina revocação e precisão, SentiStrength produziu o melhor resultado, quando comparado ao SentiWordNet.

Tabela 5.4 – Comparação entre os Baselines e os Métodos Propostos

	Baseline		MD						ME						MH					
			Sem Espelhamento		Com Espelhamento		Sem Espelhamento		Com Espelhamento		Sem Espelhamento		Com Espelhamento		Sem Espelhamento		Com Espelhamento			
	Senti WordNet	Senti Strength	Média	Mediana																
Acurácia	25%	71%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%		
Precisão Positiva	17%	53%	84%	83%	84%	84%	83%	83%	80%	82%	83%	83%	83%	84%	84%	85%	84%	84%		
Precisão Negativa	07%	08%	67%	67%	68%	68%	62%	67%	67%	69%	64%	64%	65%	60%	65%	66%	61%	61%		
Precisão Neutra	100%	94%	99%	99%	98%	98%	99%	99%	99%	99%	98%	98%	99%	99%	99%	99%	99%	99%		
Revocação Positiva	91%	78%	95%	97%	94%	94%	91%	97%	97%	97%	94%	94%	90%	90%	94%	94%	90%	90%		
Revocação Negativa	50%	37%	78%	76%	70%	71%	78%	69%	69%	70%	70%	70%	80%	80%	78%	78%	78%	78%		
Revocação Neutra	16%	71%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%	96%		
Medida-F Positiva	29%	63%	89%	89%	88%	88%	87%	87%	87%	89%	89%	89%	87%	87%	89%	89%	87%	87%		
Medida-F Negativa	12%	14%	72%	71%	69%	70%	69%	68%	68%	70%	67%	67%	69%	69%	71%	71%	69%	69%		
Medida-F Neutra	27%	81%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%		

Fonte: O Autor

O ponto fraco de todas as soluções do baseline é a classificação de tíquetes negativos, apresentando desempenho inferior para as três métricas. A precisão é quase igual no SentiWordNet e no SentiStrength (7% e 8%, respectivamente), com um impacto correspondente na medida-f correspondente. Esse resultado mostra que nossos métodos lidam eficientemente com a distinção entre os termos usados para relatar um problema e os utilizados para descrever o sentimento.

As métricas para tíquetes positivos revelam resultados discrepantes para o baseline. O SentiWordNet reconhece a maioria dos tíquetes positivos (91% do revocação), mas classifica os tíquetes negativos/neutros como positivos (17% na precisão). Em comparação, o SentiStrength reconhece tíquetes menos positivos (78% de revocação), mas mais tíquetes são corretamente classificados como positivos (53% de precisão). Nossa medida-f para tíquetes positivos varia de 87% a 89%, superando as soluções do baseline, tanto na precisão quanto na revocação.

5.3.2 Comparação entre Métodos

Ao considerar apenas as implementações de MD, ME e MH, observamos na Tabela 5.4 que os resultados são muito próximos. A mesma precisão é alcançada em todos os métodos e suas variações: 95%. O mesmo acontece com as métricas para os tíquetes neutros, onde a revocação e a medida-f são iguais (96% e 97%) e a diferença de precisão é irrelevante (98-99%). Os resultados para as outras classes também são muito bons, mostrando uma diferença muito pequena. Por exemplo, a medida-f para a classe negativa varia entre 69-72%, enquanto para a classe positiva, a variação é de 87-89%.

Portanto, realizamos duas análises adicionais:

1. comparação dos três métodos propostos utilizando o desempenho médio de suas variações, a fim de avaliar a contribuição da análise da estrutura do tíquete em uma simples abordagem baseada em dicionário;
2. comparação dos métodos de cálculo de polaridade padrão e de polaridade de categoria de acordo com o a escolha entre média/mediana e uso/não uso de espelhamento. Realizamos uma análise das diferenças usando um teste bicaudal T, com nível de significância 0,05 usando o ambiente Weka.

Os resultados da comparação dos métodos são exibidos na Tabela 5.5, na qual o valor médio para cada métrica com o desvio padrão relevante é mostrado. O baseline é os resultados do MD, e os métodos comparados são aqueles que se valhem de informação estrutural. A

tabela utiliza as convenções do Experimenter Weka, onde o símbolo * denota que o baseline é estatisticamente superior a um dado método, e ausência de símbolos denota que não há diferença estatisticamente significativa. Além dos resultados de precisão, revocação e medida-f por classe, adicionamos uma linha com os valores médios de cada métrica.

Os resultados são muito semelhantes, sem diferença estatística na maioria dos casos. Os resultados da ME provaram ser, em média, inferiores aos de MD e MH na maioria dos casos. Comparados com MD, essas diferenças são estatisticamente significantes apenas em dois casos: medida-f média e medida-f para a classe positiva. Isso significa que a combinação ponderada de precisão/revocação por classe prova a inferioridade dos resultados. Os resultados médios para os métodos MH são ligeiramente inferiores quando comparados à MD, mas essa diferença não é estatisticamente significativa. Por isso, a influência da estrutura do tíquete na determinação do sentimento não pôde ser confirmada.

Em seguida, comparamos todas as implementações de acordo com a estratégia da medida (média e mediana). Independentemente do método ou adoção do espelhamento, os resultados foram iguais ou similares, e não foi observada diferença estatística. Isso significa que tanto a mediana como a média são adequadas para calcular a polaridade padrão e a polaridade de categoria. A Tabela 5.6 exibe esses valores.

Finalmente, a Tabela 5.7 compara o desempenho das implementações sem espelhamento com as que usam espelhamento, independentemente do método e da estratégia da medida (média e mediana). Os resultados são ligeiramente inferiores na maioria dos casos, quando o espelhamento é adotado, mas estatisticamente significativa apenas em três casos: revocação e medida-f para tíquetes negativos e precisão para tíquetes neutros. Assim, a influência positiva/negativa do espelhamento no cálculo automático da polaridade padrão e da polaridade de categoria não pôde ser confirmada.

Tabela 5.5 – Comparação entre o Método de Dicionário, o Método de Estrutura e o Método Híbrido

Métrica	DM	TM	HM
Precisão Positiva	83,00(0,82)	82,50(1,73)	83,75(0,50)
Precisão Negativa	66,75(2,06)	66,25(2,87)	65,00(3,56)
Precisão Neutra	98,75(0,50)	98,50(0,58)	98,50(0,58)
Revocação Positiva	95,75(1,50)	94,00(2,45)	92,75(1,89)
Revocação Negativa	73,50(4,12)	72,00(4,08)	75,00(4,69)
Revocação Neutra	96,00(0,00)	96,00(0,00)	96,00(0,00)
Medida-F Positiva	89,00(0,00)	87,50(0,58)*	88,00(0,82)
Medida-F Negativa	70,00(2,16)	69,00(0,82)	69,75(0,96)
Medida-F Neutra	97,00(0,00)	97,00(0,00)	97,00(0,00)
Precisão Geral	82,83(13,70)	82,42(13,87)	82,42(14,44)
Revocação Geral	88,42(11,25)	87,33(11,63)	87,92(10,00)
Medida-F Geral	85,33(11,88)	84,50(12,15)*	84,92(11,86)

Fonte: O Autor

Tabela 5.6 – Comparação entre as Medidas Média e Mediana

Métrica	Mediana	Média
Precisão Positiva	83,33(0,82)	82,83(1,47)
Precisão Negativa	65,67(3,72)	66,33(1,51)
Precisão Neutra	98,67(0,52)	98,50(0,55)
Revocação Positiva	93,50(2,59)	94,83(1,72)
Revocação Negativa	74,50(4,64)	72,50(3,62)
Revocação Neutra	96,00(0,00)	96,00(0,00)
Medida-F Positiva	88,00(0,89)	88,33(0,82)
Medida-F Negativa	69,83(1,17)	69,33(1,63)
Medida-F Neutra	97,00(0,00)	97,00(0,00)
Precisão Geral	82,56(14,03)	82,56(13,57)
Revocação Geral	88,00(10,29)	87,78(11,34)
Medida-F Geral	84,94(11,65)	84,89(11,93)

Fonte: O Autor

Métrica	Sem Espelhamento	Com Espelhamento
Precisão Positiva	83,00(1,55)	83,17(0,75)
Precisão Negativa	64,67(3,01)	67,33(1,75)
Precisão Neutra	99,00(0,00)	98,17(0,41)*
Revocação Positiva	94,00(2,97)	94,33(1,37)
Revocação Negativa	76,50(3,89)	70,50(0,55)*
Revocação Neutra	96,00(0,00)	96,00(0,00)
Medida-F Positiva	88,00(1,10)	88,33(0,52)
Medida-F Negativa	70,00(1,55)	69,17(1,17)*
Medida-F Neutra	97,00(0,00)	97,00(0,00)
Precisão Geral	88,83(9,39)	86,94(12,01)
Revocação Geral	88,83(9,39)	86,94(12,01)
Medida-F Geral	85,00(11,60)	84,83(11,99)

Fonte: O Autor

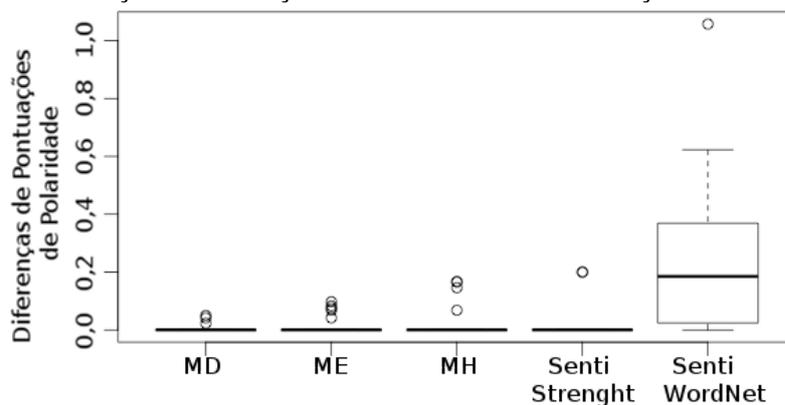
Tabela 5.7 – Comparação entre a Aplicação e a Não Aplicação do Espelhamento

5.3.3 Influência da Tradução

Para aumentar a confiança em nossa premissa de que a tradução automática é madura para a análise do sentimento [3, 18, 25], decidimos comparar os efeitos das traduções automáticas e manuais, considerando um pequeno conjunto de tíquetes. Seleccionamos aleatoriamente trinta (30) tíquetes, distribuídos em igual número por polaridade. Então, comparamos os resultados produzidos considerando as traduções manuais e automáticas de um mesmo tíquete. A Figura 5.1 exibe as diferenças nas pontuações de sentimento quantitativos usando um boxplot. Adotamos a implementação que considera a mediana e não utiliza espelhamento para cálculo da polaridade padrão e das polaridades de categorias no MD, ME e MH.

Houve um único caso de mudança de polaridade, observado para SentiWordNet (representado como o outlier). Em todos os outros casos, observamos apenas pontos de sentimento diferentes dentro da mesma polaridade. Isso significa que a tradução automática tende a preservar a valência de termos, eventualmente com uma intensidade diferente. Em particular, observamos termos menos negativos e mais positivos. Nossos métodos são menos sensíveis a tal alteração, pois calculam a pontuação como um valor agregado (mediano ou médio). O MH foi o método que apresentou a maioria das diferenças (5 tíquetes), e MD, o mínimo (2). Resultados para SentiStrength foram normalizados, porque mede o sentimento em uma escala de [-5..5]. Observamos apenas duas diferenças na pontuação. O método mais sensível para a tradução foi SentiWordNet (50% dos tíquetes). O intervalo de diferenças é maior para SentiWordNet, ao contrário dos outros métodos, onde são percebidos como outliers.

Figura 5.1 – Diferenças de Pontuações de Polaridade entre Traduções Manual e Automática



Fonte: O Autor

5.4 Configuração dos Experimentos para Emoção

5.4.1 Objetivos

Neste trabalho, o estudo de emoções é complementar ao de polaridade. Como não dispúnhamos de um *gold standard* para emoções, limitamos nosso trabalho à caracterização das emoções presentes e sua relação com a polaridade.

Assim, o principal objetivo relacionado a emoções é aplicar regras de associação para verificar possíveis correlações entre polaridade e os dois modelos de emoções no domínio de tíquetes de TI.

5.4.2 Geração de Regras de Associação entre Polaridade e Emoções

Para avaliar as regras de associação geradas, foram usadas as métricas suporte, confiança e lift.

Para identificar possíveis correlações entre polaridade e os dois modelos de emoções em tíquetes, foi desenvolvido um algoritmo na linguagem R para utilizar regras de associação. Esse algoritmo apresenta os seguintes passos:

1. Importação dos tíquetes e suas as informações de polaridades e emoções geradas na execução do MD. Foram importados apenas os tíquetes pertencentes ao *gold standard*, corretamente classificados com polaridade positiva ou negativa e que contêm as emoções VAD e NRC. Assim garante-se que estão sendo analisados apenas os tíquetes classificados com a polaridade correta. O total de tíquetes importados é de 868;

2. Geração das regras de associação, via algoritmo Apriori. Foram considerados os parâmetros padrão deste algoritmo, que são 0,1 para o suporte, 0,8 de confiança. Este parâmetros resultaram em 804 regras de associação;
3. Remoção das regras geradas que são redundantes (ver Seção 2.8), via poda, para facilitar a análise dos resultados. Restaram 23 regras após a poda. A Tabela 5.8 exibe alguns exemplos de regras de associação redundantes, que foram desconsideradas, e as regras às quais essas redundantes repetiam. As emoções VAD e Ekman representam o antecedente de cada regra, separados em duas colunas apenas para facilitar a compreensão, enquanto que a coluna Polaridade (abreviada para *Pol.*) representa o consequente;
4. Exibição dos resultados.

Tabela 5.8 – Exemplos de Regras Redundantes

Regras Mantidas						Regras Desconsideradas/Redundantes					
Pol.	VAD	Ekman	Suporte	Confiança	Lift	Pol.	VAD	Ekman	Suporte	Confiança	Lift
Neg.	Val.=Média	Alegria=Não	0,1	1	4,1	Neg.	Val.=Média	Alegria=Não Nojo=Não Tristeza=Não Raiva=Não	0,1	1	4,1
						Neg.		Alegria=Não Nojo=Não Tristeza=Não			
Neg.		Alegria=Não	0,2	1	3,9	Neg.		Alegria=Não Nojo=Não Tristeza=Não Raiva=Não	0,1	1	3,9
						Neg.		Alegria=Não Nojo=Não Tristeza=Não			
Neg.	Dom.=Neutra		0,1	0,8	3,2	Neg.	Dom.=Neutra	Nojo=Não Tristeza=Não	0,1	0,8	3,1
Pos.	Val.=Alta Exc.=Baixa Dom.=Alta		0,6	1	1,3	Pos.	Val.=Alta Exc.=Baixa Dom.=Alta	Nojo=0 Medo=0	0,6	1	1,3
						Pos.		Raiva=0 Tristeza=0			

Fonte: O Autor

Tabela 5.9 – Correlações entre Polaridade e Emoções em Tíquetes

Grupo	VAD	Ekman	Polaridade	Suporte	Confiança	Lift
N1	Valência=Neutra	Alegria=Não	Negativa	0,1	1	4,1
		Alegria=Não	Negativa	0,2	1	3,9
	Valência=Neutra		Negativa	0,1	1	3,9
N2	Dominância=Neutra		Negativa	0,1	0,8	3,2
N3	Dominância=Neutra	Medo=Sim	Negativa	0,1	0,9	3,8
		Medo=Sim	Negativa	0,1	0,9	3,6
N4	Excitação=Neutra		Negativa	0,1	0,8	2,3
P1	Excitação=Baixa	Surpresa=Sim	Positiva	0,4	0,9	1,3
		Medo=Não Surpresa=Sim	Positiva	0,4	0,9	1,3
	Valência=Alta Dominância=Alta	Surpresa=Sim	Positiva	0,4	1	1,3
	Valência=Alta	Alegria=Sim Surpresa=Sim	Positiva	0,4	0,9	1,3
		Surpresa=Sim	Positiva	0,4	0,8	1,1
P2	Valência=Alta Excitação=Baixa	Medo=Não Alegria=Sim	Positiva	0,7	0,9	1,3
		Medo=Não Alegria=Sim	Positiva	0,7	0,9	1,3
	Valência=Alta Excitação=Baixa	Alegria=Sim Tristeza=Não	Positiva	0,7	0,9	1,3
		Alegria=Sim	Positiva	0,6	0,9	1,3
	Valência=Alta Excitação=Baixa	Medo=Não Tristeza=Não	Positiva	0,7	0,9	1,3
		Alegria=Sim	Positiva	0,7	0,9	1,2
	Valência=Alta		Positiva	0,7	0,9	1,2
P3	Valência=Alta Excitação=Baixa Dominância=Alta		Positiva	0,6	1	1,3
			Positiva	0,7	0,9	1,2
	Dominância=Alta Excitação=Baixa		Positiva	0,7	0,9	1,1
P4		Medo=Não	Positiva	0,7	0,8	1,1

Fonte: O Autor

5.5 Emoção: Resultados e Discussões

5.5.1 Correlação entre Polaridade e Emoções

A Tabela 5.9 exibe as regras de associação geradas para analisar as correlações entre polaridade e os dois modelos de emoção estudados neste trabalho. Cada linha da tabela representa uma regra. As colunas VAD e Ekman representam o antecedente de cada regra, separados em duas colunas apenas para facilitar a compreensão, enquanto que a coluna Polaridade representa o conseqüente. As regras geradas foram agrupadas subjetivamente para melhor compreensão, levando em consideração a Polaridade e as emoções que possuem. Identificamos quatro grupos negativos (N1, N2, N3 e N4) e quatro positivos (P1, P2, P3 e P4). Dentro de cada grupo, listamos as regras em ordem decrescente de lift.

Chega-se às seguintes conclusões em cada grupo, a partir da análise das colunas suporte, confiança e lift:

- N1: as regras desse grupo apresentam lifts altos (3,9-4,1) e o limite máximo de confiança (1), apesar do suporte baixo (0,1-0,2). São indicadores de que a *valência neutra* e a *ausência de alegria* indicam polaridade negativa;
- N2: a regra do grupo também apresenta lift alto (3,2). Apesar do suporte baixo, é um indicador de que *dominância neutra* está correlacionada à polaridade negativa;
- N3: é composto por duas regras, que apresentam medo, e uma delas tem dominância neutra. Possuem suporte baixo (0,1), confiança alta (0,9) e lift alto (3,6-3,8). Conclui-se que a presença de *medo* implica em polaridade negativa. Ainda, reforça a constatação do grupo N2 sobre *dominância neutra*;
- N4: esta regra possui suporte baixo (0,1), mas um lift relativamente alto (2,3), assim indica que *excitação neutra* implica em tíquetes negativos;
- P1: é composto por cinco regras, sendo que todas possuem a presença de *surpresa*. Duas regras também apresentam *ausência de medo* e presença de *alegria*. *Valência e dominância altas e excitação baixa* estão presentes em três regras separadas. Apesar do lift não ser tão alto (1,1-1,3), a confiança chega ao limite (1) e o suporte é razoável (0,4). Pode-se concluir que a presença de *surpresa* correlaciona com polaridade positiva;
- P2: é composto por sete regras. Cinco regras apresentam valência alta, quatro excitação baixa e duas dominância alta. A presença de alegria aparece em cinco regras, enquanto que a ausência de medo em três. O suporte e a confiança são altos (0,6-0,7 e 0,9, respectivamente), embora o lift seja relativamente baixo (1,2-1,3). Essas regras indicam que a presença de *alegria* implica em tíquetes positivos. Considerando esse e o grupo P1, também pode-se concluir que *valência alta e excitação baixa* estão relacionados à polaridade positiva;
- P3: apresenta apenas emoções VAD em seus antecedentes: valência alta em uma regra, e excitação baixa e dominância alta em duas regras. O suporte e a confiança são altos (0,6-0,7 e 0,9-1, respectivamente) e o lift é baixo (1,1-1,3). Reforça as constatações sobre VAD do grupo P2, e, associado a P2 e P1, conclui-se que *dominância alta* e polaridade positiva estão correlacionados;
- P4: possui apenas a ausência de medo em sua regra. Apresenta suporte e confiança altos (0,7 e 0,8, respectivamente). Apesar do lift baixo (1,1), reforça a consideração do P2 a respeito da *ausência de medo*.

De forma geral, as regras que têm a polaridade negativa como consequente apresentam lift alto, mas um suporte baixo em quase todas as regras, no valor considerado como ponto de corte na geração das regras (0,1). Também nota-se que as emoções VAD têm resultados pouco marcantes, já que correlaciona valores neutros nas três dimensões. Já as que têm polaridade positiva, apresentam suporte maior e lift menor. A confiança é similar em todos, até porque o ponto de corte para a geração das regras é um valor alto (0,8). A quantidade de regras geradas é maior, e algumas das emoções VAD e NRC se repetem nelas e entre os grupos formados.

Em resumo, analisando as regras e os grupos, pode-se considerar as seguintes correlações entre polaridade e emoções:

- **Positivo:** presença de surpresa, presença de alegria, ausência de medo, valência alta, excitação baixa e dominância alta;
- **Negativa:** ausência de alegria, presença de medo, valência neutra, excitação neutra e dominância neutra.

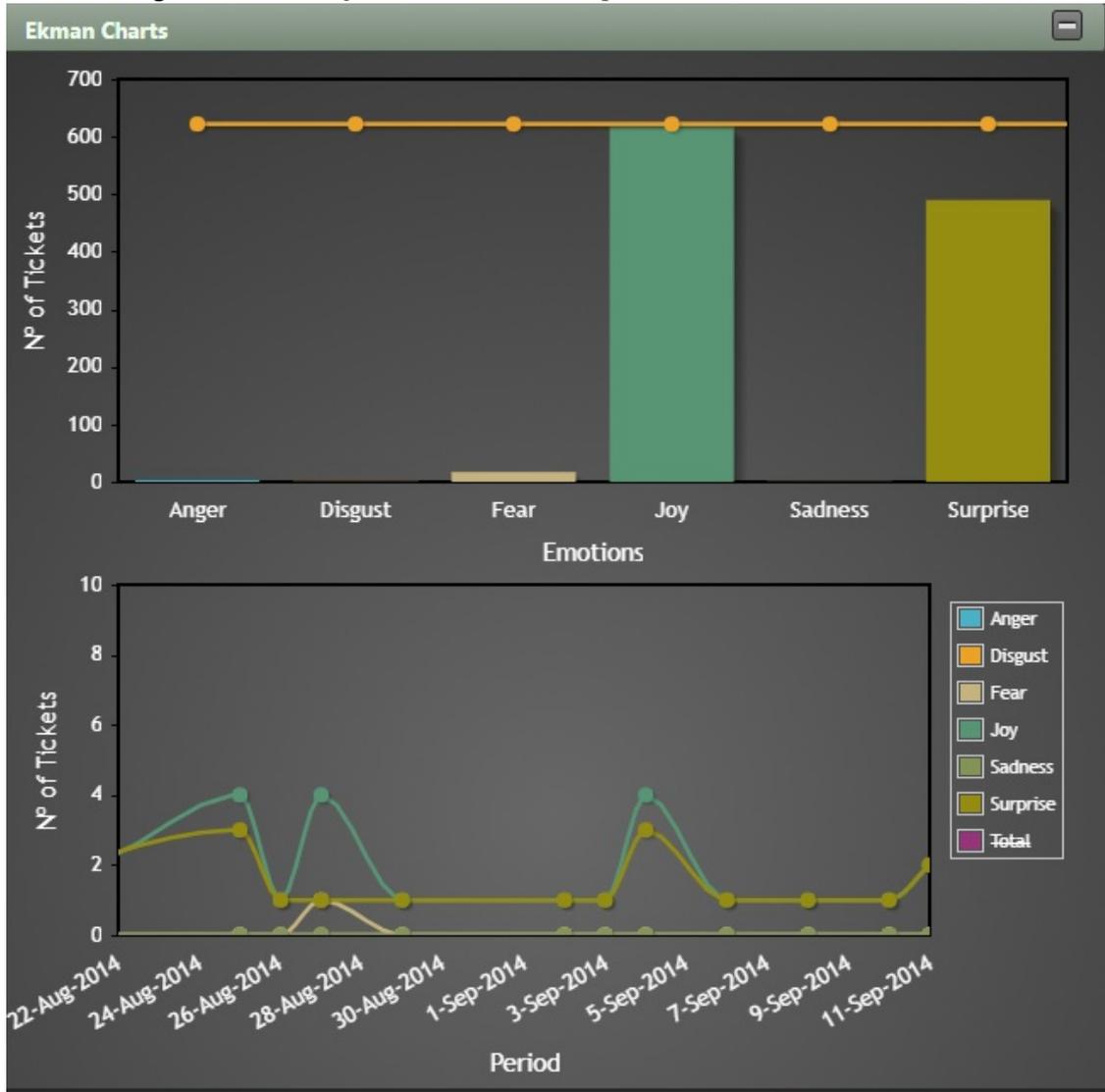
5.5.2 Análise das Relações entre Polaridade e Emoções na POC

Finalmente, comparamos os resultados da análise das correlações entre polaridade e os dois modelos de emoções com as informações apresentadas na POC. Analisamos 1191 tíquetes da POC, que correspondem a todos os tíquetes que possuem emoções de Ekman e/ou VAD na execução do MD usando o *gold standard* como base. Desse total de 1191 tíquetes, 664 são positivos e 527 são negativos.

As figuras 5.2 e 5.3 apresentam informações das emoções de Ekman e VAD, respectivamente, desses tíquetes positivos. No primeiro gráfico da Figura 5.2, o total desses tíquetes que apresentam emoções de Ekman é representado pela linha laranja. Nota-se que quase todos os tíquetes apresentam alegria e grande parte apresentam surpresa. O segundo gráfico mostra uma predominância dessas duas emoções ao longo do tempo. Esse segundo gráfico foi ampliado e não exibe os valores de todos os tíquetes positivos, apenas para facilitar a visualização. Foram filtrados tíquetes positivos cuja data de cadastro está entre duas datas próximas escolhidas aleatoriamente.

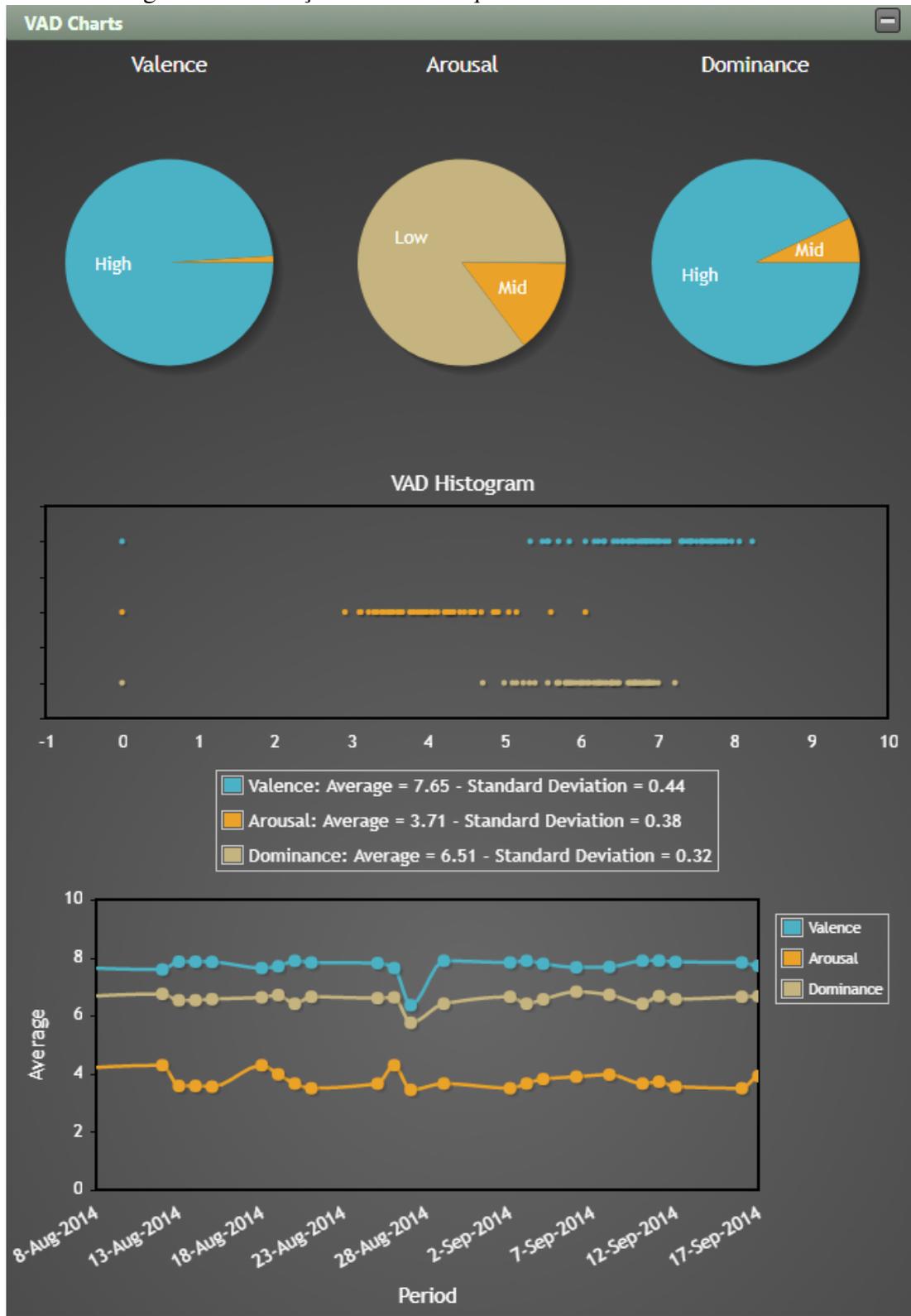
Na Figura 5.3, no primeiro gráfico, nota-se que quase todos os tíquetes têm valência e dominância altas, e que grande parte tem excitação baixa. Os outros dois gráficos dessa figura também apresentam um maior acúmulo de tíquetes com valência e dominância altas e com

Figura 5.2 – Emoções de Ekman nos Tíquetes Positivos do Gold Standard



Fonte: O Autor

Figura 5.3 – Emoções VAD nos Tíquetes Positivos do Gold Standard



Fonte: O Autor

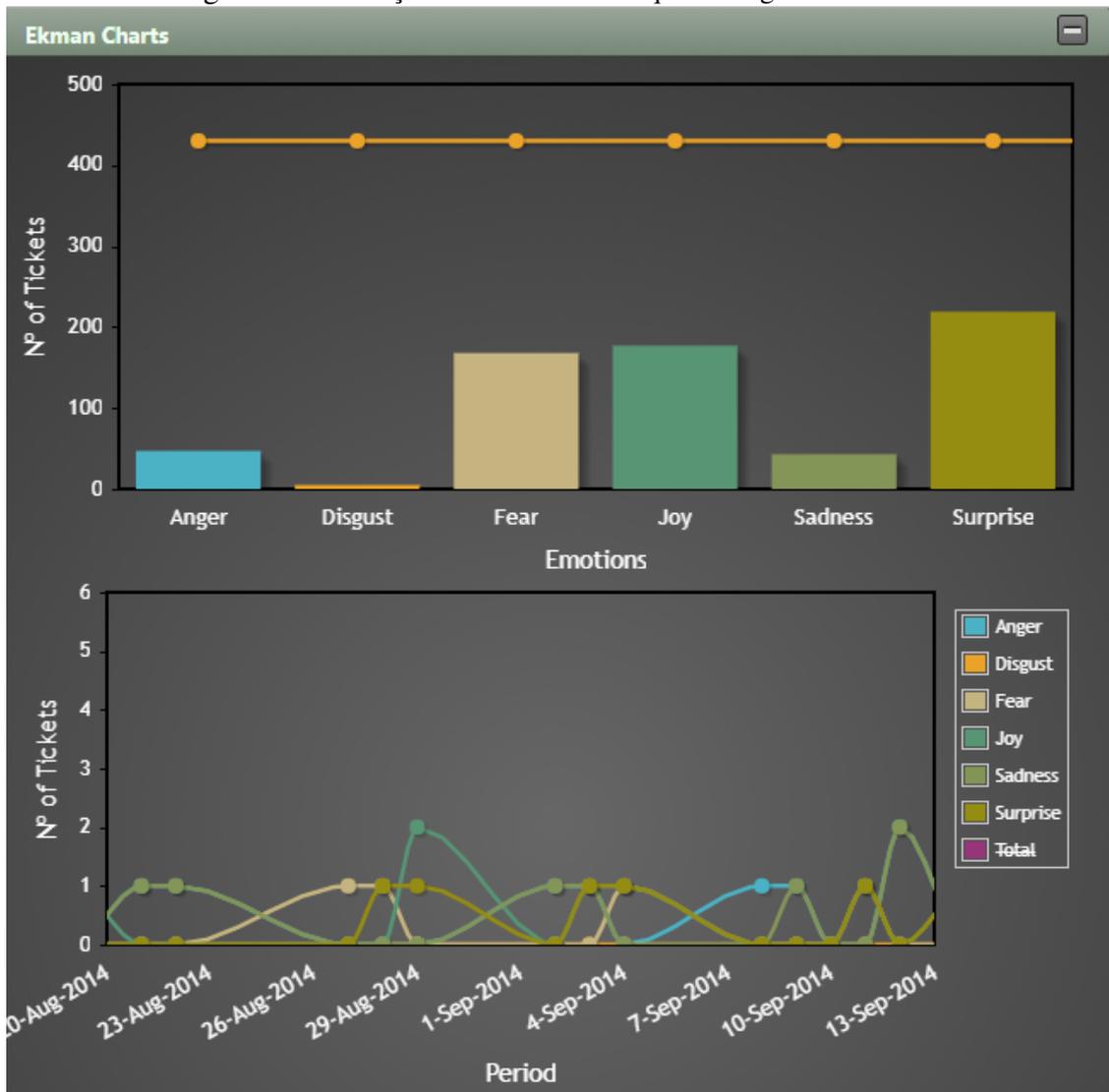
excitação baixa. O terceiro gráfico dessa figura também foi ampliado apenas para facilitar a visualização, utilizando os mesmos critérios de filtragem acima.

Essas informações sobre tíquetes positivos e suas respectivas emoções de Ekman e VAD consolidam os resultados das regras de associação geradas para tíquetes positivos.

Os tíquetes negativos estão representados pelas figuras 5.4 e 5.5, as quais exibem informações das emoções de Ekman e VAD, respectivamente. No primeiro gráfico da Figura 5.4, a quantidade das emoções surpresa, alegria e medo são parecidas e representam as emoções mais encontradas. Há um menor número de tíquetes com as emoções tristeza e raiva. As emoções representadas no segundo gráfico dessa figura variam bastante e não é possível chegar a alguma conclusão mais contundente. Esse segundo gráfico foi ampliado para facilitar a visualização, filtrando tíquetes negativos cuja data de cadastro está entre duas datas próximas escolhidas aleatoriamente.

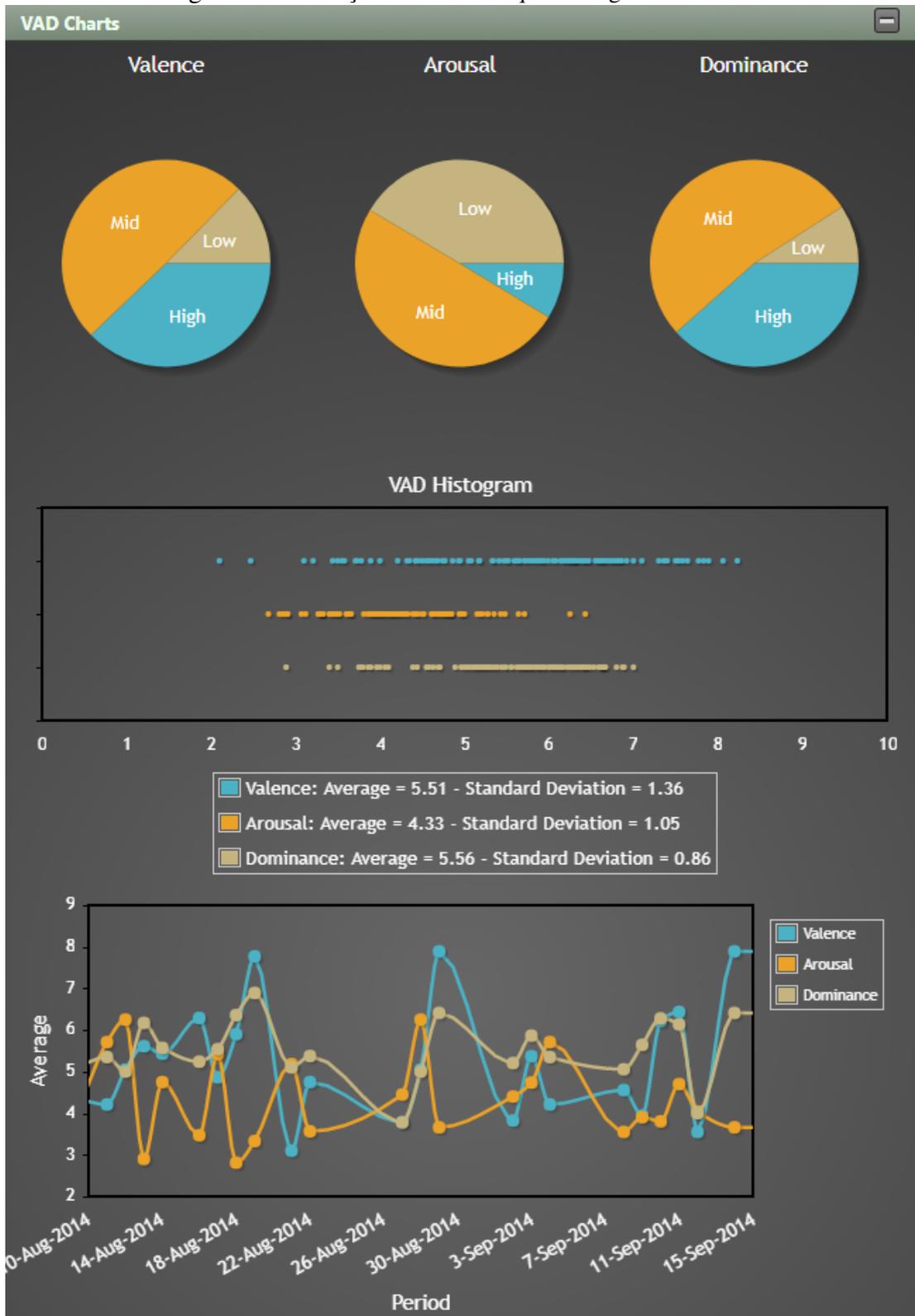
O primeiro gráfico da Figura 5.5 apresenta uma predominância de tíquetes com emoções mais neutras para os três componentes do modelo VAD, mas em torno da metade dos tíquetes que apresentam esses valores. No histograma, a valência é bem dispersa, a excitação apresenta valores mais neutros/baixos, e a dominância, neutros/altos. O gráfico por período também é confuso nesse modelo, mesmo sendo com um conjunto filtrado de tíquetes. Dessa forma, percebe-se que os tíquetes negativos não têm um padrão muito claro, apesar de tenderem a serem neutros no modelo VAD. Quanto às emoções de Ekman, há divergência com os resultados das regras de associação geradas, principalmente porque os tíquetes negativos apresentam as emoções alegria e surpresa na POC. Em parte isto é justificado pela observação de que tíquetes negativos apresentam saudações (e.g. bom dia, prezado) e conclusões cordiais (e.g. obrigado, abraços), as quais são associadas a essas duas emoções mais positivas. Contudo, esses tipos de emoções são mais amenas, enquanto que o relato é o trecho mais importante e apresenta as emoções expressando um relato mais apreensivo ou agressivo.

Figura 5.4 – Emoções de Ekman nos Tíquetes Negativos da POC



Fonte: O Autor

Figura 5.5 – Emoções VAD dos Tíquetes Negativos da POC



Fonte: O Autor

5.5.3 Discussão

Conforme mostrado na Tabela 5.2, apenas 105 (4,5%) tíquetes do *gold standard* foram anotados como negativos, então qualquer erro relacionado à classe negativa tem um efeito maior nas métricas, em comparação com os que se referem aos tíquetes positivos e neutros. Isso explica parcialmente porque todos os resultados para tíquetes negativos são inferiores quando comparados aos relacionados a tíquetes positivos/neutros. Em compensação, os tíquetes neutros representam 85,5% do *gold standard*. Portanto, cada erro gerado para essa classe não é tão expressivo nas métricas. Isso também afetou o poder estatístico dos testes realizados.

Uma análise manual em parte das regras revelou que a principal razão para classificação errada é a ausência de algumas expressões de sentimento do DD. Por exemplo, “O sistema mobile não está funcionando. Que briga!” foi classificado erroneamente como neutro porque o *token* “briga” não era suficientemente frequente no corpus e não foi adicionado no DD.

Uma possível solução para a questão acima mencionada seria ajustar o DD para incluir mais sementes. No entanto, esta solução precisa ser cuidadosamente avaliada, pois pode afetar a interpretação de tíquetes neutros. Por exemplo, a palavra “já” é usada para expressar idéias objetivas e subjetivas. O texto “eu já pedi para corrigi-lo várias vezes...” certamente é negativo, mas no tíquete “Nós já compramos as licenças, proceda com a instalação do software” é neutro.

Quanto à correlação entre polaridade e os dois modelos de emoção estudados, conforme apresentado na Seção 5.5.1, é possível correlacionar as polaridades negativa e positiva com algumas emoções. Contudo, para a polaridade negativa alguns resultados não são expressivos, enquanto que para a positiva são mais contundentes. Analisando os tíquetes, nota-se que as pessoas são controladas ao expressar suas emoções nos tíquetes. São corteses, mas diretas em seus textos. Também tendem a escrever de uma forma controlada e formal para expressar o positivo/negativo, não usando ironias ou palavras de baixo calão, por exemplo. Uma pesquisa focada em apenas uma ou algumas emoções, como é realizado por Gachechiladze et al. (2017), e buscando suas correlações com polaridade pode trazer resultados mais detalhados.

6 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho, descrevemos e avaliamos uma abordagem para a análise do sentimento em tíquetes de TI. Essa abordagem baseia-se na construção de um dicionário de domínio para filtrar termos candidatos a sentimentos no domínio de TI e um processo para análise de sentimento que atribui automaticamente pontuações de sentimento com base em um léxico de sentimento e/ou estrutura de tíquete. Foram propostos três métodos para classificação de polaridade (MD, ME, e MH), e quatro variações de cálculo para pontuação automática de polaridade. De forma complementar, analisamos a existência de emoções em tíquetes, considerando os modelos de Ekman e VAD, utilizando o método puramente baseado em dicionário (MD). Analisamos também a relação entre polaridade e as diferentes formas de mensurar emoção usando regras de associação. Estas contribuições foram reportadas em um artigo publicado no 13th International Conference on Mining Software Repositories (BLAZ; BECKER, 2016). Os resultados alcançados nesse artigo possibilitaram o desenvolvimento desta dissertação.

Para a polaridade, os métodos propostos superam em muito o SentiStrength, uma ferramenta genérica de análise de sentimento popular no campo da engenharia de software. Os três métodos criados apresentaram desempenho semelhante entre si, com uma pequena vantagem para o MD e o MH. Contudo, os resultados dos testes estatísticos não nos permitiu determinar o papel desempenhado pela estrutura do tíquetes na identificação do sentimento. Também não foram observadas diferenças significativas entre o uso de média/mediana e aplicação da estratégia nomeada espelhamento para derivar automaticamente a pontuação padrão e as pontuações de categorias, embora o desempenho do espelhamento tenha sido um pouco pior. No geral, alcançamos resultados muito bons na classificação da polaridade dos tíquetes, particularmente para tíquetes neutros e positivos. Já a identificação de tíquetes negativos requer melhorias. Aplicamos nossos métodos em todo o corpus de TI e encontramos 5% de tíquetes negativos e 10% de tíquetes positivos, uma proporção muito semelhante à distribuição encontrada em nosso *gold standard*.

Já a classificação da emoção não pôde ser avaliada se forma mais precisa pela falta de um *gold standard*. Contudo, através das regras de associação, identificamos correlações bem estabelecidas entre emoções e a polaridade positivo. Já para a polaridade negativa, os resultados não foram tão conclusivos. Portanto o estudo de emoções requer aprofundamento., principalmente em relação às emoções consideradas negativas.

A solução proposta emprega recursos direcionados à língua Inglesa, e foi aplicada ao nosso corpus em Português do Brasil com a ajuda da tradução automática. Assim, além do

Inglês, ela pode ser adaptada a qualquer idioma para o qual os tradutores automáticos maduros estejam disponíveis. A análise de uma pequena amostra revelou que nossos métodos são pouco sensíveis aos problemas de tradução.

O DD é criado através de um processo clássico de expansão e poda. Apesar da criação do DD necessitar de um conjunto de sementes, estas foram selecionadas e preparadas através de um processo bastante fácil. Adotamos listas existentes de saudações e expressões de conclusão, bem como *tokens* frequentes extraídos do *gold standard*. As pontuações de polaridade foram atribuídas automaticamente através de diferentes estratégias, com base nas sementes e nos *tokens* expandidos. O DD resultante é, portanto, adequado aos tíquetes de TI em geral, mas o processo pode ser replicado para outros artefatos com características semelhantes sem grande esforço.

A POC demonstrou que sentimentos também podem ser um critério para atendimento de tíquetes. É possível cadastrar e importar tíquetes. O DD usado para o processamento de métodos de classificação pode ser manipulado conforme a necessidade. A partir da filtragem de tíquetes, os principais *tokens* escritos pelos usuários podem ser visualizados. Gráficos e valores consolidados das polaridades e de emoções dos tíquetes também podem ser analisados, permitindo um prévio *feedback* dos usuários.

Trabalhos futuros podem investigar sentimentos em outros tipos de fontes de tíquetes/textos de usuários à área de TI. Meios mais formais (e.g. atas, emails) e até mesmo trocas de mensagens em aplicativos recentes (e.g. WhatsApp, Facebook) provêm textos que podem apresentar sentimentos.

Novas formas de pré-processamentos (e.g. tratamento de maiúsculas, negação de termos) podem alterar a intensidade de sentimentos presentes em tíquetes similar ao ponto de exclamação, conforme estudado neste trabalho. Portanto podem ser tratadas para melhorar o desempenho de métodos de classificação.

Os algoritmos MD, ME e MH podem ser melhorados em diferentes aspectos. Além de elaborar novas formas de explorar a estrutura dos tíquetes, podem ser experimentadas outras estratégias para a definição da polaridade padrão e das polaridades das categorias. As limitações do SentiWordNet no que diz respeito ao domínio de TI podem ser superadas com diferentes alternativas, como a combinação de soluções genéricas de sentimento (por exemplo, SentiStrength), ou a atribuição de pontuações diretamente no DD, por exemplo, através de colaboração coletiva (*crowdsourcing*).

A comparação das abordagens léxica, de aprendizado de máquina e de outras estratégias focando no domínio de TI permitira identificar qual apresenta melhor desempenho, principal-

mente a partir da construção/desenvolvimento de métodos/técnicas específicas a esse domínio.

O *gold standard* foi gerado aleatoriamente a partir do corpus, portanto as categorias maiores são mais representadas. No entanto, tíquetes abordam diferentes tipos de questões, sistemas e departamentos. No futuro, melhores formas de compor as amostras permitiram generalizar nossas descobertas em relação a todos os tipos de tíquetes e corporações de TI. Outros trabalhos também podem aprofundar o estudo de traduções e comparar/explorar ferramentas de análise de sentimentos que consideram textos em Português do Brasil.

Finalmente, um estudo mais profundo sobre emoções em tíquetes forneceria novos conhecimentos à análise de sentimentos em tíquetes de TI. Métodos específicos para a classificação de emoções podem ser desenvolvidos. Contudo, a maior dificuldade é avaliar os resultados obtidos, já que a anotação de corpus para os modelos de emoções estudados previamente é complexa. Uma alternativa para resolver essa questão é estudar com uma ou algumas emoções específicas desses modelos, o que restringiria o contexto da pesquisa.

REFERÊNCIAS

- AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **SIGMOD Rec.**, ACM, New York, NY, USA, v. 22, n. 2, p. 207–216, jun. 1993. ISSN 0163-5808. Disponível em: <<http://doi.acm.org/10.1145/170036.170072>>.
- BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: **Proc. of the International Conference on Language Resources and Evaluation (LREC), Valletta, Malta**. [s.n.], 2010. Disponível em: <<http://www.lrec-conf.org/proceedings/lrec2010/summaries/769.html>>.
- BALAHUR, A.; PEREA-ORTEGA, J. M. Sentiment analysis system adaptation for multilingual processing: The case of tweets. **Information Processing & Management**, 2014.
- BALAHUR, A.; TURCHI, M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. **Computer Speech & Language**, v. 28, n. 1, p. 56–75, 2014.
- BECKER, K.; MOREIRA, V. P.; SANTOS, A. G. dos. Multilingual emotion classification using supervised learning. **Inf. Process. Manage.**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 53, n. 3, p. 684–704, maio 2017. ISSN 0306-4573. Disponível em: <<https://doi.org/10.1016/j.ipm.2016.12.008>>.
- BECKER, K.; TUMITAN, D. Introdução à mineração de opiniões: Conceitos, aplicações e desafios. In: **Lectures of the 28th Brazilian Symposium on Databases**. [S.l.: s.n.], 2013.
- BLAZ, C. C. A.; BECKER, K. Sentiment analysis in tickets for it support. In: **Proceedings of the 13th International Conference on Mining Software Repositories**. New York, NY, USA: ACM, 2016. (MSR '16), p. 235–246. ISBN 978-1-4503-4186-8. Disponível em: <<http://doi.acm.org/10.1145/2901739.2901781>>.
- BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **J. Mach. Learn. Res.**, JMLR.org, v. 3, p. 993–1022, mar. 2003. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=944919.944937>>.
- BOLLEGALA, D.; MU, T.; GOULERMAS, J. Y. Cross-domain sentiment classification using sentiment sensitive embeddings. **IEEE Trans. on Knowl. and Data Eng.**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 28, n. 2, p. 398–410, fev. 2016. ISSN 1041-4347. Disponível em: <<http://dx.doi.org/10.1109/TKDE.2015.2475761>>.
- BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, Elsevier, v. 2, n. 1, p. 1–8, 2011.
- BRADLEY, M. M.; LANG, P. J. **Affective norms for English words (ANEW): Instruction manual and affective ratings**. [S.l.], 1999.
- EKMAN, P. An argument for basic emotions. **Cognition & emotion**, v. 6, n. 3-4, p. 169–200, 1992.
- EL-HALEES, A. M. Software Usability Evaluation Using Opinion Mining. **Journal of Software**, v. 9, n. 2, p. 343–349, fev 2014. ISSN 1796-217X. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-84894430241{&}partnerID=tZOtx>>.

- FELLBAUM, C. **WordNet: An Electronic Lexical Database**. [S.l.]: Bradford Books, 1998.
- GACHECHILADZE, D. et al. Anger and its direction in collaborative software development. In: **Proceedings of the 39th International Conference on Software Engineering: New Ideas and Emerging Results Track**. Piscataway, NJ, USA: IEEE Press, 2017. (ICSE-NIER '17), p. 11–14. ISBN 978-1-5386-2675-7. Disponível em: <<https://doi.org/10.1109/ICSE-NIER.2017.18>>.
- GODBOLE, N.; SRINIVASIAIAH, M.; SKIENA, S. Large-scale sentiment analysis for news and blogs. In: **Proc. of the First International AAAI Conference on Weblogs and Social Media (ICWSM)**. [S.l.: s.n.], 2007. v. 2.
- GUZMAN, E.; AZÓCAR, D.; LI, Y. Sentiment analysis of commit comments in github: an empirical study. In: **Proc. of the 11th Working Conference on Mining Software Repositories, MSR 2014, Hyderabad, India**. [s.n.], 2014. p. 352–355. Disponível em: <<http://doi.acm.org/10.1145/2597073.2597118>>.
- GUZMAN, E.; BRUEGGE, B. Towards Emotional Awareness in Software Development Teams. In: **Proc. of the 9th Joint Meeting on Foundations of Software Engineering**. New York, New York, USA: ACM Press, 2013. p. 671–674. ISBN 978-1-4503-2237-9. Disponível em: <<http://doi.acm.org/10.1145/2491411.2494578>>.
- HAN, J. **Data Mining: Concepts and Techniques**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 1558609016.
- HU, M.; LIU, B. Mining and summarizing customer reviews. In: ACM. **Proc. of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)**. [S.l.], 2004. p. 168–177.
- ISLAM, M. R.; ZIBRAN, M. F. Leveraging automated sentiment analysis in software engineering. In: **Proceedings of the 14th International Conference on Mining Software Repositories**. Piscataway, NJ, USA: IEEE Press, 2017. (MSR '17), p. 203–214. ISBN 978-1-5386-1544-7. Disponível em: <<https://doi.org/10.1109/MSR.2017.9>>.
- JONGELING, R.; SEREBRENIK, A. Choosing Your Weapons : On Sentiment Analysis Tools for Software Engineering Research. In: **Proc. of the IEEE International Conference on Software Maintenance and Evolution**. Bremen: [s.n.], 2015. p. 531–535. ISBN 9781467375313.
- JURADO, F.; RODRIGUEZ, P. Sentiment Analysis in monitoring software development processes: An exploratory case study on GitHub's project issues. **Journal of Systems and Software**, Elsevier Ltd., v. 104, p. 82–89, 2015. ISSN 01641212. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0164121215000485>>.
- JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 1st. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000. ISBN 0130950696.
- KOBAYASHI, N.; INUI, K.; MATSUMOTO, Y. Extracting aspect-evaluation and aspect-of relations in opinion mining. In: **EMNLP-CoNLL 2007, Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, June 28-30, 2007, Prague, Czech Republic**. [s.n.], 2007. p. 1065–1074. Disponível em: <<http://www.aclweb.org/anthology/D07-1114>>.

LIU, B. **Sentiment analysis and opinion mining**. [S.l.]: Morgan & Claypool Publishers, 2012.

LIU, B. **Opinions, Sentiment, and Emotion in Text**. Cambridge University Press, 2015. (Sentiment Analysis: Mining Opinions, Sentiments, and Emotions). ISBN 9781107017894. Disponível em: <<https://books.google.com.br/books?id=6IdsCQAAQBAJ>>.

MAKHOUL, J. et al. Performance measures for information extraction. In: **In Proceedings of DARPA Broadcast News Workshop**. [S.l.: s.n.], 1999. p. 249–252.

MÄNTYLÄ, M. V. et al. Bootstrapping a lexicon for emotional arousal in software engineering. In: **Proceedings of the 14th International Conference on Mining Software Repositories**. Piscataway, NJ, USA: IEEE Press, 2017. (MSR '17), p. 198–202. ISBN 978-1-5386-1544-7. Disponível em: <<https://doi.org/10.1109/MSR.2017.47>>.

MOHAMMAD, S. M. From once upon a time to happily ever after: Tracking emotions in mail and books. **Decis. Support Syst.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 53, n. 4, p. 730–741, nov. 2012. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2012.05.030>>.

MOHAMMAD, S. M. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In: MEISELMAN, H. (Ed.). **Emotion Measurement**. [S.l.]: Elsevier, 2016.

MOHAMMAD, S. M.; TURNEY, P. D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: **Proc. of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text**. [S.l.: s.n.], 2010. (CAAGET '10), p. 26–34.

MOLINA-GONZALEZ, M. D. et al. Semantic orientation for polarity classification in spanish reviews. **Expert Systems with Applications**, v. 40, n. 18, p. 7250 – 7257, 2013. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417413004752>>.

MURGIA, A. et al. Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In: **Proc. of the 11th Working Conference on Mining Software Repositories, MSR 2014, Hyderabad, India**. [S.l.: s.n.], 2014.

ORTU, M. et al. Are Bullies more Productive ? Empirical Study of Affectiveness vs . Issue Fixing Time. In: **Proc. of the IEEE/ACM Working Conference on Mining Software Repositories**. Florence: [s.n.], 2015.

ORTU, M. et al. The emotional side of software developers in jira. In: **Proceedings of the 13th International Conference on Mining Software Repositories**. New York, NY, USA: ACM, 2016. (MSR '16), p. 480–483. ISBN 978-1-4503-4186-8. Disponível em: <<http://doi.acm.org/10.1145/2901739.2903505>>.

PARROTT, W. **Emotions in Social Psychology: Essential Readings**. Psychology Press, 2001. (Key readings in social psychology). ISBN 9780863776823. Disponível em: <<https://books.google.com.br/books?id=jV5QVgM6Me8C>>.

PHILLIPS, S. J.; DUDÍK, M.; SCHAPIRE, R. E. A maximum entropy approach to species distribution modeling. In: **Proceedings of the Twenty-first International Conference**

on **Machine Learning**. New York, NY, USA: ACM, 2004. (ICML '04), p. 83–. ISBN 1-58113-838-5. Disponível em: <<http://doi.acm.org/10.1145/1015330.1015412>>.

PLETEA, D.; VASILESCU, B.; SEREBRENIK, A. Security and Emotion: Sentiment Analysis of Security Discussions on GitHub. In: **Proc. of the IEEE/ACM Working Conference on Mining Software Repositories**. [S.l.: s.n.], 2014. p. 348–351.

PLUTCHIK, R. **The Emotions**. University Press of America, 1991. ISBN 9780819182869. Disponível em: <<https://books.google.com.br/books?id=OWF9AAAAMAAJ>>.

PRABOWO, R.; THELWALL, M. Sentiment analysis: A combined approach. **Journal of Informetrics**, v. 3, n. 2, p. 143–157, 2009. Disponível em: <<http://EconPapers.repec.org/RePEc:eee:infome:v:3:y:2009:i:2:p:143-157>>.

QIU, G. et al. Expanding domain sentiment lexicon through double propagation. In: **Proceedings of the 21st International Joint Conference on Artificial Intelligence**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009. (IJCAI'09), p. 1199–1204. Disponível em: <<http://dl.acm.org/citation.cfm?id=1661445.1661637>>.

SAHIBUDIN, S.; SHARIFI, M.; AYAT, M. Combining itil, cobit and iso/iec 27002 in order to design a comprehensive it framework in organizations. In: **Proc. of the Second Asia International Conference on Modeling Simulation, 2008**. [S.l.: s.n.], 2008. p. 749–753.

SANTOS A. G. L. DOS; BECKER, K.; MOREIRA, V. Um estudo de caso de mineração de emoções em textos multilíngues. In: **Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM'13)**. [S.l.: s.n.], 2014.

SANTOS A. G. L. DOS; BECKER, K.; MOREIRA, V. **Experimentos Comparativos Combinando Aprendizado Supervisionado e Tradução Automática para Mineração de Emoções em Textos Multilíngues**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2017.

SILVA, M.; CARVALHO, P.; SARMENTO, L. Building a sentiment lexicon for social judgement mining. **Computational Processing of the Portuguese Language**, Springer, p. 218–228, 2012.

SINHA, V.; LAZAR, A.; SHARIF, B. Analyzing developer sentiment in commit logs. In: **Proceedings of the 13th International Conference on Mining Software Repositories**. New York, NY, USA: ACM, 2016. (MSR '16), p. 520–523. ISBN 978-1-4503-4186-8. Disponível em: <<http://doi.acm.org/10.1145/2901739.2903501>>.

STEINBERGER, J. et al. Creating sentiment dictionaries via triangulation. In: **Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. (WASSA '11), p. 28–36. ISBN 9781937284060. Disponível em: <<http://dl.acm.org/citation.cfm?id=2107653.2107657>>.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining, (First Edition)**. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005. ISBN 0321321367.

THELWALL, M. et al. Sentiment in short strength detection informal text. **J. Am. Soc. Inf. Sci. Technol.**, John Wiley & Sons, Inc., New York, NY, USA, v. 61, n. 12, p. 2544–2558, dez. 2010. ISSN 1532-2882. Disponível em: <<http://dx.doi.org/10.1002/asi.v61:12>>.

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

TUMITAN, D.; BECKER, K. Sentiment-based features for predicting election polls: A case study on the brazilian scenario. In: **Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 02**. Washington, DC, USA: IEEE Computer Society, 2014. (WI-IAT '14), p. 126–133. ISBN 978-1-4799-4143-8. Disponível em: <<http://dx.doi.org/10.1109/WI-IAT.2014.89>>.

WARRINER, A. B.; KUPERMAN, V.; BRYSSBAERT, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. **Behavior research methods**, v. 45, n. 4, p. 1191–207, dez. 2013.

WIEBE, J.; WILSON, T.; CARDIE, C. Annotating expressions of opinions and emotions in language. **Language resources and evaluation**, Springer, v. 39, n. 2-3, p. 165–210, 2005.

WU, C.-H.; CHUANG, Z.-J.; LIN, Y.-C. Emotion recognition from text using semantic labels and separable mixture models. ACM, New York, NY, USA, v. 5, n. 2, p. 165–183, jun. 2006. ISSN 1530-0226. Disponível em: <<http://doi.acm.org/10.1145/1165255.1165259>>.

ZAKI, M. J. Generating non-redundant association rules. In: **Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2000. (KDD '00), p. 34–43. ISBN 1-58113-233-6. Disponível em: <<http://doi.acm.org/10.1145/347090.347101>>.