



UNIVERSIDADE
ESTADUAL DE LONDRINA

ALEX MARINO GONÇALVES DE ALMEIDA

SENTIC LEVELS COMO DESCRITORES NA
CLASSIFICAÇÃO MULTI-CLASSES EM ANÁLISE DE
SENTIMENTOS

Londrina
2016

ALEX MARINO GONÇALVES DE ALMEIDA

SENTIC LEVELS COMO DESCRITORES NA
CLASSIFICAÇÃO MULTI-CLASSES EM ANÁLISE DE
SENTIMENTOS

Dissertação apresentada ao Programa de Mestrado em Ciência da Computação da Universidade Estadual de Londrina como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Sylvio Barbon Jr

Londrina
2016

Alex Marino Gonçalves de Almeida

Sentic Levels como descritores na classificação multi-classes em Análise de Sentimentos/ Alex Marino Gonçalves de Almeida. – Londrina–PR, 2016-

79 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Sylvio Barbon Jr

– Universidade Estadual de Londrina, 2016.

1. Análise de Sentimentos - Teses. 2. Mineração de Opinião - Teses. 3. Classificação Multi-classes - Teses. I. Barbon Jr, Sylvio. II. Universidade Estadual de Londrina. III. Centro de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação. IV. Título

CDU 02:141:005.7

ALEX MARINO GONÇALVES DE ALMEIDA

**SENTIC LEVELS COMO DESCRITORES NA CLASSIFICAÇÃO
MULTI-CLASSES EM ANÁLISE DE SENTIMENTOS**

Dissertação apresentada ao Programa de Mestrado em Ciência da Computação da Universidade Estadual de Londrina como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

BANCA EXAMINADORA

Orientador: Prof.Dr. Sylvio Barbon Jr.
Universidade Estadual de Londrina –
UEL

Prof. Dr. Bruno Bogaz Zarpelão
Universidade Estadual de Londrina –
UEL

Prof. Dr. Alan Salvany Felinto
Universidade Estadual de Londrina –
UEL

Prof. Dr. Elieser Botelho Manhas Jr.
Universidade Estadual de Londrina –
UEL

Londrina, 22 de dezembro de 2016.

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

AGRADECIMENTOS

Agradeço a Deus por absolutamente tudo, em especial por tornar possível os agradecimentos subsequentes.

Em escala humana não posso eximir de agradecimento e amor, minha família pelo sacrifício que lhes fora imposto para que este fim fosse atingido.

Ao meu orientador é necessário agradecer a paciência, atenção e dedicação que tornou possível este trabalho.

*“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2)*

ALMEIDA, A. M. G.. *Sentic Levels como descritores na classificação multi-classes em Análise de Sentimentos*. 79 p. Dissertação de Mestrado (Mestrado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina–PR, 2016.

RESUMO

A Análise de Sentimentos (AS) como um problema de classificação em mídias textuais eletrônicas é uma tarefa cada vez mais presente no cotidiano das pessoas bem como a relevância destes temas no ambiente acadêmico. Para auxiliar a classificação de emoções em conteúdo textuais, usou-se neste trabalho uma base de dados obtida de um site brasileiro de notícias escritas em língua portuguesa. Para extração de características textuais fez-se uso de recursos do *Hourglass of Emotions* capazes de aferir as dimensões e intensidades afetivas presentes no texto e compor características baseadas nos dois níveis de sentimento (*Sentic Level*) extraídas em função do *TF-IDF*, *Strength* e na quantidades de palavras com peso afetivo. O principal objetivo da classificação é distinguir as emoções presentes no conteúdo textual onde este trabalho propõe um modelo de classificação multi-classes apoiado em técnicas de decomposição (Um contra Todos e Um contra Um) de problemas multi-classes em problemas binários e subsequentes métodos de agregação. A etapa de classificação ocorre após a preparação do *dataset* em função dos métodos Um contra Um e Um contra Todos, possibilitando assim submetê-los aos algoritmos binários (*Support Vector Machine*, *Árvore de Decisão*, *Naïve Bayes*, *Random Forest*). As emoções resultaram da saída dos classificadores modelados encaminhadas para os métodos de agregação predizendo assim a emoção final do documento. Para avaliação dos resultados utilizou-se como métrica, adequada para este tipo de experimento, a Acurácia. Os resultados obtidos dos experimentos mostram que o modelo proposto produziu uma acurácia compatível com trabalhos no Estado da Arte.

Palavras-chave: Aprendizado de Máquina, Multi-classe, SenticNet, Mineração de Opinião, Análise de Sentimentos

ALMEIDA, A. M. G.. **Sentic Levels as descriptors for Multi-class classification Problem in Sentiment Analysis** . 79 p. Master's Thesis (Master in Science in Computer Science) – State University of Londrina, Londrina–PR, 2016.

ABSTRACT

Nowadays the Sentiment Analysis as a classification problem in electronic textual media is increasingly present in the way of life people as well as the relevance of these issues in the research environment. To help Emotions rating in textual content in this work used a dataset extracted from a Brazilian news site written in Portuguese. The extraction of textual features was made use of Hourglass of Emotions resources which are capable of measuring the affective dimensions and intensities in the text. The features are composed of both Sentic Levels extracted from the TF-IDF, Strength and amounts of Word with emotional weight. The primary purpose of classification is to distinguish the emotions present in the textual content where this paper proposes a multi-class classification model based on decomposition techniques (One against All and One on One) of multi-class problems in binary problems and following methods aggregation. The classification task occurs after the dataset preparation depending on the methods One on One and One against All, thus enabling submit them to binary classifiers as Support Vector Machine, Decision Tree, Naïve Bayes, Random Forest. The results obtained from the output of binary classifiers forwarded to the aggregation methods to predict the document final emotion. To evaluate the results was used the accuracy as a metric suitable for this type of experiment. The tests results show that the proposed model produced a consistent accuracy with work in the State of the Art.

Keywords: Machine Learning, Multi-class, SenticNet, Opinion Mining, Sentiment Analysis

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo Circumplexo de Russel é um dos modelos mais recentes de representação dimensional de emoção.	31
Figura 2 – Roda das emoções de Plutchik	32
Figura 3 – Hourglass Model	33
Figura 4 – A passagem de um <i>sentic level</i> para outro regulado por uma curva Gaussiana.	34
Figura 5 – Computação Tradicional e Aprendizado de Máquina	36
Figura 6 – Hierarquia do Aprendizado	37
Figura 7 – Processo de Aprendizado	38
Figura 8 – Esquema geral da abordagem de Decomposição e Agregação adotada neste trabalho.	40
Figura 9 – Decomposição OVO	41
Figura 10 – Decomposição OVA	42
Figura 11 – Margem e equações dos hiperplanos para o hiperplano canônico de maior margem. Os hiperplanos marginais são representados pelas linhas tracejadas.	45
Figura 12 – Espaço bi-dimensional demonstrando as possíveis partições.	46
Figura 13 – Esquema de uma Floresta Aleatória.	47
Figura 14 – Sumário dos Experimentos	53
Figura 15 – Estratégia de preparação dos dados - <i>dataset</i> para Matriz Documento	55
Figura 16 – Esquema de extração das características derivadas relativas ao TF-IDF, <i>Word</i> e <i>Strength</i> em função dos SL1 e SL2.	57
Figura 17 – Preparação OVA para emoção Raiva	59
Figura 18 – Esquema geral das tarefas de decomposição e Agregação.	60
Figura 19 – Aplicação do método de agregação Máxima Confidência (OVA) para sentença preparada com Características SL1/ <i>Strength</i>	61
Figura 20 – Comparando influência das sentenças com presença do neutro.	65
Figura 21 – Acurácia para algoritmos e métodos de agregação para classificar o sentimento da sentença	67
Figura 22 – <i>Ranking</i> médio para algoritmos e métodos de agregação para classificar o sentimento da sentença	68
Figura 23 – Comparação entre as acurácias médias por configuração (algoritmo e agregação) de acordo o teste Nemenyi. Grupos de classificadores que não apresentam diferença significativa ($\alpha = 0,05$) estão conectados pelas linhas horizontais	69

LISTA DE TABELAS

Tabela 1 – Algumas definições existentes de emoções básicas	30
Tabela 2 – Sentic Level 1	35
Tabela 3 – Sentic Level 2	35
Tabela 4 – Amostra de 5 notícias extraídas do site G1	54
Tabela 5 – Rótulos das seis emoções básicas para o <i>dataset</i>	55
Tabela 6 – Lista de caracteres removidos na operação de limpeza	56
Tabela 7 – Lista de <i>stopwords</i> idioma Português	56
Tabela 8 – Sentic Level 1 - <i>strength</i> extraídos de 5 sentenças	58
Tabela 9 – Classificadores SVM treinados com dados de treinamento com características SL1/ <i>strength</i> e respectivos graus de confidências	62
Tabela 10 – Matriz Resultado e cômputo para predição final pelo método de agregação de Votação	62
Tabela 11 – Resultado das 10 melhores configurações segundo acurácia média em 30 repetições	66
Tabela 12 – Complexidade computacional dos Algoritmos, onde n é número de amostras e m número de características.	66
Tabela 13 – Resultado SL x Descritores	66
Tabela 14 – Relação da menor, média e maior acurácia por Método de Agregação.	67
Tabela 15 – Relação da menor, média e maior acurácia por Algoritmo.	67

LISTA DE ABREVIATURAS E SIGLAS

AD	Árvore de Decisão
AM	Aprendizado de Máquina
AS	Análise de Sentimentos
FDP	Função Densidade de Probabilidade
HOE	<i>Hourglass of Emotions</i>
IHM	Interação Homem Máquina
KULE	Método de agregação por distância minimizada de Kullback-Leibler
LVPC	Método de agregação <i>Learning valued preference for classification</i>
MCON	Método de agregação por máxima confiança
MO	Mineração de Opinião
NB	Naïve Bayes
OVA	<i>One versus All</i>
OVO	<i>One versus One</i>
PCB	Problema de Classificação Binário
PCMC	Problema de Classificação Multi-classes
RF	<i>Random Forest</i>
SL1	<i>Sentic Level 1</i>
SL2	<i>Sentic Level 2</i>
SVM	<i>Support Vector Machine</i>
VOTE	Método de agregação por votação
WENG	Método de Agregação Weng
WVOT	Método de agregação por votação por peso

SUMÁRIO

1	INTRODUÇÃO	23
2	REFERENCIAL TEÓRICO	27
2.1	Análise de Sentimentos e <i>Sentic Computing</i> (Computação Sentimental) .	27
2.1.1	Categorização das Emoções	29
2.1.2	Abordagens Dimensionais <i>versus</i> Categóricas	30
2.1.3	<i>Hourglass of Emotions</i>	32
2.2	Problemas de Classificação com Aprendizado de Máquina	35
2.2.1	Hierarquia de Aprendizado	36
2.2.2	Processo de Aprendizado	37
2.2.3	Paradigmas do aprendizado	38
2.3	Problema de Classificação Multi-Classe	39
2.3.1	Técnicas de Decomposição	40
2.3.2	Métodos de Agregação	42
2.3.2.1	Agregações OVO	43
2.3.2.2	Agregações OVA	44
2.4	Classificadores básicos	44
2.4.1	<i>Support Vector Machine</i>	44
2.4.2	Árvore de Decisão	46
2.4.3	<i>Random Forests</i>	47
2.4.4	Naïve Bayes	48
2.5	Medida de Performance	48
2.6	Teste Friedman e <i>post hoc</i> Nemenyi	49
2.7	Trabalhos Relacionados	50
3	METODOLOGIA	53
3.1	Experimentos	53
3.2	<i>Dataset</i>	54
3.3	Preparação dos dados	55
3.4	Extração de Características	57
3.5	Preparação OVA e OVO	59
3.6	Decomposição e Agregação	60
4	RESULTADOS	65
5	CONCLUSÃO	71

REFERÊNCIAS	73
Trabalhos Publicados pelo Autor	79

1 INTRODUÇÃO

Análise de Sentimentos (AS), também conhecida como Mineração de Opinião (MO), é o campo de estudo que analisa opiniões, sentimentos, avaliações, atitudes e emoções. Trata-se de um extenso campo de estudo que abriga muitos nomes e tarefas levemente diferentes, como por exemplo a AS, MO, Extração de Opinião, Mineração de Sentimento, Análise de Subjetividade, Análise Afetiva, Análise de Emoção, etc. Entretanto estão todas contidas na área de AS ou MO que representam o mesmo campo de estudo [1].

Dentro do contexto AS o presente estudo volta a atenção para as emoções, que para Darwin [2] se envolvem em funções cognitivas peculiares, uma vez que as emoções comunicam informações de um animal para outro. As emoções são cruciais no cotidiano e modo de vida das pessoas, deste modo as aplicações de AS e MO têm se tornado cada vez mais presentes na vida real e igualmente popular causando grande impacto social. A AS e Análise de Emoções são amplamente adotadas em promoções de vendas [3], predição de mercado de ações [4], detecção de polaridade de tópicos políticos emergentes [5] entre outras aplicações.

Este tipo de impacto encoraja diversas áreas de pesquisa a dedicar atenção ao assunto, tais como: Processamento de Linguagem Natural, Recuperação de Informação, Mineração de Texto, Mineração de Dados. A AS, em particular, é a área de pesquisa que estuda a identificação de sentimentos e emoções na forma textual. Neste sentido a Web tem gerado continuamente avaliação de produtos, textos emocionais e sentimentais sobre percepção de marcas e questões políticas [6], que podem ser utilizados para capturar as opiniões dos consumidores, opinião pública geral quanto a preferência de produtos, estratégias corporativas, movimentações políticas [7]. AS refere-se à inferência do ponto de vista do usuário, posição e atitudes em seus documentos escritos [8].

AS é uma área de pesquisa importante e atuante, que se dedica exaustivamente a estudar classificação de sentimentos em fontes de dados textuais denominadas documento. A tarefa de classificação de sentimento é tipicamente um Problema de Classificação Binário (PCB) [1] uma vez que é um problema bipolar (sentimentos positivos ou negativos), ou seja, pode assumir apenas duas classes. Contudo, ao levar em conta que na concepção de Ekman [9], as pessoas têm seis emoções de ordem primária que podem ainda desdobrar-se em conjuntos de ordem secundária e terciária, assim diferentemente de um típico PCB, as emoções podem ser melhor descritas como um Problema de Classificação Multi-classes (PCMC) [10, 11, 12]. Da união da necessidade de se classificar as emoções oriundas de mídias textuais sob uma abordagem Multi-classes decorre naturalmente no uso de recursos de Computação dos Sentimentos de modo a reconhecer de forma mais apropriada, interpretar e processar sentimentos e emoções [13].

Para a psicologia as emoções representam reações variáveis em duração e intensidade

que ocorrem no corpo e mente humana, usualmente disparadas por um conteúdo mental [14]. A Ampulheta de Emoções ou *Hourglass of Emotions* (HOE) é um modelo de categorização afetiva que reinterpreta o modelo de Plutchik rearranjando as emoções básicas em quatro dimensões independentes e concomitantes. Os diferentes níveis de ativação destas dimensões medem o estado emocional mental. O modelo da HOE é capaz de sintetizar as experiências emocionais em termos das dimensões **Amenidade, Atenção, Sensibilidade e Atitude** do inglês *Pleasantness, Attention, Sensitivity e Aptitude*. A principal motivação do uso do modelo HOE é a possibilidade de inferência entre informações cognitivas e afetivas contidas no texto [15].

O principal artefato empregado neste trabalho para classificação das emoções é o Aprendizado de Máquina (AM). Muitas das abordagens de AS são baseadas em técnicas de AM, tais como, *Support Vector Machine* (SVM) [16], Naïve Bayes (NB) [17], Árvore de Decisão (AD) [18], Máxima Entropia [19] têm sido amplamente utilizadas na construção de modelos de classificação em AS. Outros classificadores tais como K vizinhos próximos [20] e Adaboost [21] são figuras de estudos em diferentes domínios de AS.

A resolução de um PCMC pode ser solucionada pela decomposição do problema em diversos PCB e com as saídas submetê-las aos métodos de agregação. Para os métodos de decomposição temos os métodos Um contra Um ou *One vs One* (OVO) [22] e Um contra Todos ou *One vs All* (OVA) [23]. O problema de classificação abordado neste trabalho é um PCMC de simples rótulo, ou seja, onde um dado exemplo pode assumir apenas uma classe. Entretanto não é incomum problemas onde os exemplos possam assumir mais de um rótulo como classe e para este tipo de problema dá-se o nome de Problema de Classificação Multirrótulo que no passado eram motivados principalmente para tarefas de classificação de textos [24].

Neste trabalho se aplicam diferentes abordagens de resolução de PCMC de simples rótulo conjuminadas com AS para identificar emoções de um conjunto de dados de notícias escritas em língua portuguesa. A extração de características é, em paralelo à abordagem PCMC, ponto crucial deste trabalho. Baseando-se no Nível de Sentimento ou *Sentic level 1* (SL1) e Nível de Sentimento ou *Sentic level 2* (SL2), provenientes da HOE [15] extraídos do *lexicon SenticNet* com auxílio da *Sentic API* para cada sentença contida no documento. Para realização deste trabalho utilizou-se um conjunto de dados composto por 2000 sentenças extraídas de um site de notícias e previamente rotuladas com as seis emoções básicas de Ekman [9]: **raiva, desgosto, medo, alegria, tristeza e surpresa** ou originalmente em inglês *angry, disgust, fear, joy, sadness, surprise*. Na implementação da resolução do PCMC adotou-se duas abordagens de decomposição: OVO e OVA; sucedendo um aparato elaborado por um conjunto de classificadores binários básicos induzidos pelos algoritmos SVM, NB, AD e *Random Forests* (RF) e métodos de agregação.

A principal contribuição deste trabalho é a adoção de uma abordagem PCMC com objetivo de classificar emoções presentes em textos por meio de um novo conjunto de características inspiradas no HOE, mais especificamente os SL1 e SL2. Ao término deste trabalho propor-se-á

responder às seguintes questões: A) As sentenças neutras podem influenciar na classificação?; B) É possível descrever as sentenças textuais por meio dos SL e obter resultados aceitáveis?; C) Existe um algoritmo mais apropriado para identificação de emoções?; D) Existe método de agregação mais adequado?

O restante deste trabalho é assim organizado:

- No Capítulo 2 serão descritos os principais elementos deste trabalho como AS e Computação dos Sentimentos, PCMC, Classificadores Básicos, Medidas de Performance, Teste estatístico de Friedman / *post hoc* Nemenyi e Trabalhos Relacionados.
- O Capítulo 3 descreve a metodologia empregada na execução dos experimentos deste trabalho.
- O Capítulo 4 apresenta os resultados obtidos nos experimentos acompanhados de discussão dos mesmos.
- O Capítulo 5 apresenta a conclusão. São feitas as considerações finais e direção dos futuros trabalhos.

2 REFERENCIAL TEÓRICO

Neste capítulo será apresentada a fundamentação dos assuntos envolvidos neste trabalho. Na Seção 2.1 serão apresentados os conceitos de AS e como se entrelaça com a Computação Sentimental ou *Sentic Computing* abordando principalmente a fundamentação de emoção e afeto que possibilitaram a categorização das emoções humanas dentro do HOE e por conseguinte os SL1 e SL2. Na seção 2.2 serão descritos os principais conceitos e abordagens referentes ao AM. Na Seção 2.3 apresentaremos os princípios da abordagem PCMC e os diferentes métodos de decomposição e agregação. De forma complementar ao PCMC apresentaremos também os algoritmos de AM utilizados como classificadores básicos na Seção 2.4. Na sequência destaca-se na Seção 2.5 a medida de performance utilizada neste trabalho, seguida pelos testes estatísticos na Seção 2.6 e com a Seção 2.7 descrevendo os trabalhos relacionados finaliza-se este capítulo.

2.1 Análise de Sentimentos e *Sentic Computing* (Computação Sentimental)

Apesar de o histórico da pesquisa científica na área de emoções decorrer desde o século XIX, quando Charles Darwin e Willian James propuseram teorias das emoções que continuam a influenciar o pensamento nos dias de hoje [25, 26], a inclusão de afeto e emoções no âmbito computacional é recente. Durante a maior parte do século XX, a pesquisa em emoção foi capitaneada, principalmente, pelos campos da filosofia e psicologia cujos trabalhos baseavam-se num pequeno conjunto de teorias emocionais que continuam a sustentar estas áreas. Na década de 80 do século XX, pesquisadores como Turkle [27] começaram a especular sobre como o computador poderia ser utilizado para estudar as emoções. Desde então programas de pesquisa emergiram e na década de 90, por exemplo, Scherer [28] implementou um modelo computacional de emoções como um *Expert System*. Pouco tempo depois Picard lançou o livro sobre computação afetiva [29] gerando uma onda de interesse de pesquisadores na busca de uma maneira de melhorar as interfaces homem-máquina coordenando emoções e cognições. Picard descreveu três tipos de aplicações para computação afetiva:

1. Sistemas que detectam emoções do usuário
2. Sistemas que expressam o que o ser humano pode perceber como emoção
3. Sistemas que sentem emoções

A comunidade acadêmica tem empenhado parte de sua energia no estudo da AS, que é o estudo computacional das opiniões, sentimentos, emoções e atitudes expressas em textos

[30]. AS compreende as tarefas de detectar, extrair e classificar sentimentos e emoções. Com o advento da Web 2.0, as pessoas passaram a se expressar e compartilhar suas opiniões na web com maior avidez, passando a retratar e relatar seus costumes, cotidianos e opiniões de forma frequente. Grande parte dessa mudança de costumes pode ser atribuída à evolução das redes de mídias sociais. Este novo fenômeno despertou na indústria e academia o interesse nas áreas AS e MO com intuito de extrair e analisar o humor, sentimentos ou opiniões expressas pelo público em geral [7].

Neste sentido podemos enquadrar a *Sentic Computing* dentro da AS como uma abordagem multi-disciplinar uma vez que se ocupa de interpretar sentimentos e emoções na forma textual [13]. Multi-disciplinar porque explora tanto a ciência social quanto computação de maneira que possa reconhecer, interpretar e processar os sentimentos e emoções. Para *Sentic Computing*, cujos termos derivam do latim *sentire* e *sensus*, a análise da linguagem natural é baseada em ontologias afetivas que possibilitam a análise de textos não somente em nível de documento, página, parágrafo mas também em nível de sentenças e cláusulas. *Sentic Computing* envolve o uso de Inteligência Artificial e técnicas de Web Semântica para inferência e representação do conhecimento, Matemática para redução de dimensionalidade, Linguística para análise de discurso; Psicologia para modelagem afetiva e cognitiva; sociologia para compreensão da dinâmica das redes sociais e, finalmente, a Ética para entender as características denotadas pela natureza do pensamento e criação de máquinas emocionais [13].

Como se pode notar, de fato a AS, e mais especificamente a *Sentic Computing*, é uma abordagem que traz consigo lições da computação afetiva em conjunto com computação de senso comum. No âmbito da MO não apenas o conhecimento de senso comum é importante mas também o conhecimento emocional, de modo a ajustar as informações cognitivas e afetivas associadas às opiniões e sentimentos em linguagem natural.

Efetivamente a *Sentic Computing* permeia a área Interação Homem Máquina (IHM) e modelagem de afetos e emoções, cujo foco principal é a detecção de emoções de mídias textuais. A complexidade da detecção de afeto se apresenta devido ao fato de que uma interface afetiva pode não responder ao estado afetivo do usuário caso ela não sinta seu estado afetivo. A detecção de afeto é portanto um problema muito desafiador, isto porque as emoções são construções com limites nebulosos e variam de indivíduo para indivíduo. Para superar tal obstáculo, a *Sentic Computing* constrói uma categorização afetiva inspirada na Biologia e na Psicologia conjuntamente. Tal categorização é potencialmente apta para descrever toda a gama de experiências emocionais considerando quatro dimensões independentes e concomitantes, cujos diferentes níveis de ativação compõem o estado emocional mental.

Para melhor compreender os artefatos que sustentam o estudo das emoções dentro da *Sentic Computing* foi necessário dividir esta Seção. Na Sub-seção 2.1.1 são apresentadas as formas de categorização das emoções. Na Sub-seção 2.1.2 são apresentadas as diferentes abordagens e por último na Sub-seção 2.1.3 apresenta-se o medelo HOE.

2.1.1 Categorização das Emoções

Opiniões e sentimentos são transmitidos, com muita frequência, de forma implícita através do contexto, domínio ou até mesmo de conceitos em vez de formas explícitas através de palavras com significados afetivos. Assim, com objetivo de analisar semântica e afetivamente textos em linguagem natural para AS ou MO é preciso contar com uma base de conhecimento que contenha informações cognitivas e afetivas associadas aos conceitos. Para este fim, uma rede semântica de conhecimento de senso comum foi mesclada com recursos linguísticos para uma representação léxica de conhecimento afetivo, a fim de se obter uma nova base de conhecimento onde os conceitos e características emocionais estão interligadas. Essa base de conhecimento possibilita explorar o desenvolvimento de uma nova visualização de conhecimento afetivo de senso comum [13].

O estudo das emoções é um dos capítulos mais confusos da história da Psicologia. Isto deve-se grandemente ao fato da ambiguidade da linguagem natural, que não facilita a descrição de misturas de emoções de modo inequívoco. Tomando a palavra emocional *amor* e outras palavras como *raiva* e *medo*, de fato, são palavras “malas” (que carregam em seu bojo mais de um significado), não definidas com clareza porque seu significado pode variar de pessoa para pessoa [31]. Neste último século mais de noventa teorias sobre emoções foram elaboradas em distintos trabalhos, nenhuma delas mencionando ou descrevendo a complexa cadeia de palavras onde cada qual exerça seu sentido emocional sobre outra ou outras palavras. Algumas categorizações incluem emoções cognitivas contra emoções não cognitivas, instintiva (originária da amígdala) contra cognitiva (originária do córtex pré-frontal) e também categorização baseada em duração, como por exemplo, a surpresa que pode ocorrer por segundos ou frações de segundos ou o amor que pode durar anos.

A teoria de James-Lange define que a experiência emocional é abrangente devido às mudanças corporais [25]. Sua principal contribuição é a ênfase na personificação das emoções, especialmente o argumento de que as mudanças corporais podem alterar a experiência emocional em intensidade. Mais recentemente neurocientistas encamparam e modificaram o ponto de vista de James-Lange considerando que as mudanças corporais modulam a experiência emocional [32]. Neste sentido as emoções são relativas a certas atividades em áreas do cérebro que direcionam sua atenção, motivando seu comportamento e determinando a significância do que acontece no seu entorno. Outros trabalhos como Broca [33], Papez [34] e MacLean [35] sugeriram que as emoções estão relacionadas a um grupo de estruturas na região central do cérebro chamada sistema límbico que inclui o hipotálamo, giro do cíngulo, hipocampo e outras estruturas. Pesquisas mais recentes, entretanto, mostram que algumas das estruturas límbicas não estão diretamente relacionadas às emoções como as outras estruturas, enquanto algumas estruturas não límbicas têm apresentado maior relevância emocional [36].

2.1.2 Abordagens Dimensionais *versus* Categóricas

Estudos filosóficos de emoções datam da antiguidade greco-romana, Cícero por exemplo organizou as emoções em quatro categorias básicas:

1. *metus* (medo)
2. *aegritudo* (dor)
3. *libido* (luxúria)
4. *laetitia* (amenidade, alegria)

Mais recentemente no século XIX Darwin [2], com a teoria evolucionária, abordou a emoção com a seleção natural e conseqüentemente com contrapartidas universais e culturalmente cruzadas. Ekman [9] nos anos de 1970 encontrou evidências que os seres humanos compartilham seis emoções básicas: alegria, tristeza, desgosto, medo, raiva e surpresa. Recentemente outras tentativas para detectar estados afetivos avançados, tais como fadiga, ansiedade, satisfação, confusão, dentre outras, têm sido empenhadas e podem ser visualizadas na Tabela 1.

Tabela 1 – Algumas definições existentes de emoções básicas

Autor	# Emoções	Emoções básicas
Ekman	6	alegria, tristeza, desgosto, medo, raiva e surpresa
Parrot	6	medo, alegria, amor, tristeza, surpresa
Frijda	6	desejo, felicidade, interesse, surpresa, maravilha, tristeza
Plutchik	8	aceitação, raiva, antecipação, aversão, alegria, medo, tristeza, surpresa
Tomkins	9	desejo, felicidade, interesse, surpresa, maravilha, tristeza
Matsumoto	22	alegria, antecipação, raiva, desgosto, tristeza, surpresa, medo, aceitação, timidez, orgulho, apreciação, calma, admiração, desprezo, amor, felicidade, excitação, tristeza, facilidade, desconforto, respeito

Em 1980, Averill avançou com a ideia de que as emoções não poderiam ser explicadas exclusivamente por fatores psicológicos e cognitivos. Em vez disso, ele afirmara que as emoções são construções sociais; assim uma análise social é necessária para compreender realmente a natureza das emoções [37]. A relação entre emoção e linguagem (considerando que a linguagem de emoções é considerada parte vital da experiência emocional) tem sido utilizada por antropólogos e cientistas sociais construtivistas para questionar a universalidade dos estudos de Ekman, isto porque os rótulos linguísticos usados para codificar emoções são restritos ao inglês americano. Somado a isso, outras culturas podem possuir rótulos que não podem ser traduzidos literalmente para o inglês, como por exemplo o termo “saudade” na língua portuguesa.

De modo a estabelecer uma conexão profunda entre linguagem e os rótulos (limitados) utilizados, a abordagem categórica constantemente falha em descrever a complexa gama de

emoções que ocorrem na comunicação do dia a dia. A abordagem dimensional [38] por sua vez representa as emoções como coordenadas num espaço multi-dimensional. Tanto em teoria como na prática, muitos pesquisadores preferem definir as emoções em duas ou mais dimensões. Um bom exemplo é o modelo circunplexo de Russel [39] descrito na Figura 1, que utiliza dimensões de estímulo e valência para plotar 150 rótulos afetivos.



Figura 1 – Modelo Circunplexo de Russel é um dos modelos mais recentes de representação dimensional de emoção. Adaptado de [15].

De forma similar, Wissell considera as emoções como espaço bi-dimensional contínuo onde as dimensões são avaliação e ativação [40]. A dimensão avaliação mede os sentimentos humanos, do positivo para o negativo. Já a dimensão ativação mede se os humanos estão mais ou menos susceptíveis a tomar uma ação sob um determinado estado emocional, variando do ativo para o passivo. Whissel anotou os pares (ativação, avaliação) para 9000 palavras com conotações afetivas possibilitando a construção de seu Dicionário de Afetos na Linguagem.

No modelo bi-dimensional de Plutchik elaborado sob um círculo de emoções, que propõe uma teoria baseada em princípios evolucionários [14]. Seguindo o pensamento darwiniano, a abordagem funcional para emoções sustenta que emoções são envolvidas em funções singulares, tal como manter-se seguro [41].

Segundo Plutchik, as emoções são adaptativas e têm um surgimento complexo, oriundo de uma longa história evolutiva. No entanto, Plutchik, concebendo as emoções como estados de sentimentos afirma que o estado de sentimento é parte do processo cognitivo e comportamental. Em 1980, Plutchik criou a roda das emoções, visualizada na Figura 2 que consiste de 8 emoções



Figura 2 – Roda das emoções de Plutchik [15].

básicas e 8 emoções avançadas cada uma composta de 2 básicas. Em seu modelo, a dimensão vertical representa a intensidade enquanto a dimensão radial representa o grau de similaridade entre as emoções.

As abordagens dimensionais, no entanto, possuem limitações. Contudo o espaço dimensional permite comparar palavras afetivas em função de suas distâncias. A maioria das representações dimensionais não permite que seus modelos contemplem o fato de que mais de uma emoção possa ser experimentada ao mesmo tempo. Eventualmente, estas abordagens trabalham em função de palavras, impossibilitando assim compreender a valência afetiva do conceito de multi-palavras.

2.1.3 *Hourglass of Emotions*

A HOE [15] é um modelo de categorização afetiva inspirado no estudo de emoções de Plutchik [14], organizando as emoções primárias em torno de quatro dimensões independentes e concomitantes, com diferentes níveis de ativação possibilitando aferir um estado emocional totalizado. Esta reinterpretação é baseada na teoria da mente de Minsk [31], segundo a qual o modelo consiste em diferentes recursos independentes e que os estados emocionais

resultam da ligação de um conjunto de recursos e do desligamento de outro conjunto de recursos. Desta forma, o modelo pode sintetizar uma larga variedade de experiências emocionais em termos de *pleasantness*, *aptitude*, *sensitivity* e *attention*, assim como a combinação de diferentes valores das quatro dimensões afetivas podem modelar estados afetivos para os quais não temos definições específicas, devido a ambiguidade da linguagem natural da natureza elusiva das emoções. A principal motivação para elaboração deste modelo é a inferência em nível de

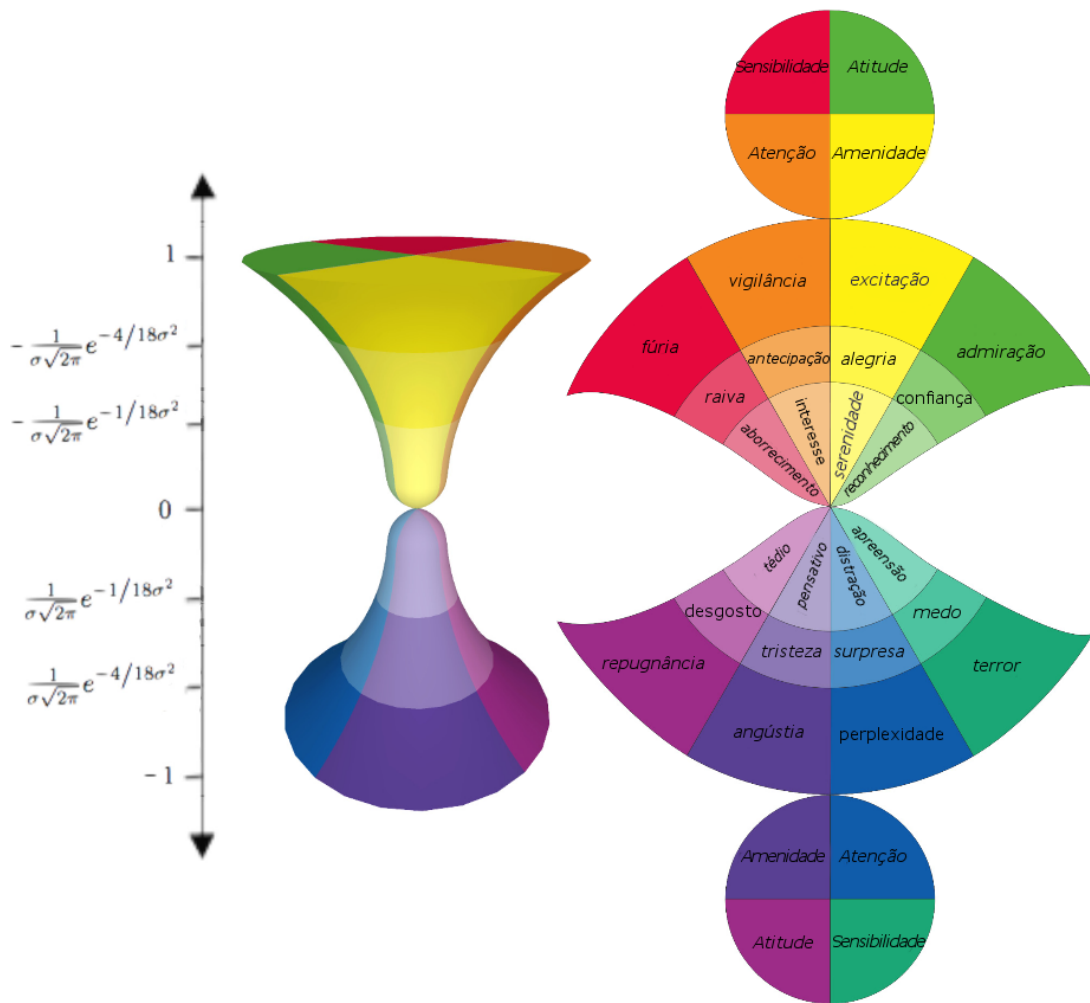


Figura 3 – Hourglass Model [15]

conceito da associação das informações cognitivas e afetivas com o texto. Esta informação é importante, dentro da *Sentic Computing*, para a AS baseada em características. No entanto o modelo da HOE exclui as emoções auto-conscientes ou emoções morais, como por exemplo orgulho, culpa, vergonha ou humilhação [42].

O modelo de fato é biologicamente inspirado e psicologicamente motivado na ideia de que os estados emocionais resultam do ligamento e desligamento seletivo de diferentes recursos do cérebro. Cada conjunto de seleções de ligamento ou desligamento dos recursos cerebrais modifica o modo como pensamos pela simples mudança de nossa atividade cerebral. Por exemplo, o estado de raiva seleciona um conjunto de recursos que nos auxilia a reagir de forma rápida e

contudente enquanto suprime outros recursos que nos faria agir de forma mais prudente. A sobrevivência depende de um padrão comportamental que é repetido e reiterado em situações de vida e morte. Com a evolução, este processo foi incorporado geneticamente nos seres humanos e em todas espécies animais no que é conhecido como instinto de “luta ou fuga”[43].

O primeiro fator que podemos aferir sobre as emoções que podemos sentir é chamado de “*Strength*”. Assim, quando sentimos uma forte emoção, é porque sentimos uma emoção muito específica. Por outro lado, não podemos sentir uma emoção muito específica como medo ou surpresa sem que esta emoção seja razoavelmente forte. Por esta razão, a transição entre diferentes estados emocionais produz um gráfico na forma de sino invertido como pode ser visto na Figura 4.

A justificativa para assumir que a função Gaussiana (em vez de uma função linear) é mais apropriada para modelar a variação de intensidade emocional é baseada em pesquisas da correlação neural e comportamental das emoções. De fato, ninguém genuinamente conhece que tipo de função subjetiva de intensidade emocional segue, porque nunca foi efetivamente medida [44].

A HOE pode ser explorada no contexto IHM de modo a medir:

1. o nível de atenção do usuário numa interação
2. o quão confortável está o usuário numa dinâmica de interação
3. o nível de confiança do usuário em benefício da interação

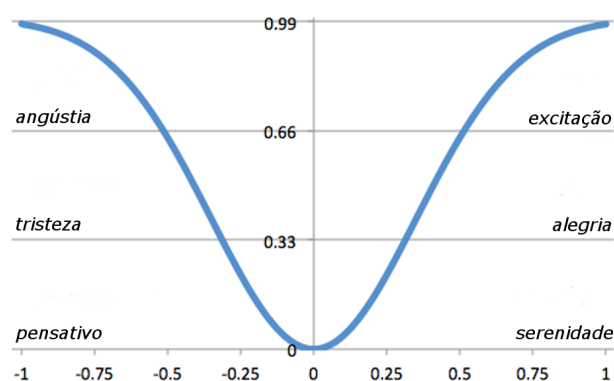


Figura 4 – A passagem de um *sentic level* para outro regulado por uma curva Gaussiana [15].

Cada dimensão afetiva é caracterizada em seis níveis de ativação (medida de intensidade de uma emoção), definida como *sentic level*, que representa as intensidades limítrofes da emoção percebida ou expressa. Estes níveis são rotulados num conjunto de 24 emoções básicas, 6 para dimensão afetiva, e possibilitando ao modelo especificar a informação afetiva associada ao texto de forma dimensional e discreta conforme demonstrado na Tabela 2.

Tabela 2 – Sentic Level 1

Intervalo	Amenidade	Atenção	Sensibilidade	Atitude
[G(1), G(2/3))	excitação	vigilância	fúria	admiração
[G(2/3), G(1/3))	alegria	antecipação	raiva	confiança
[G(1/3), G(0))	serenidade	interesse	aborrecimento	reconhecimento
(G(0), -G(1/3)]	pensativo	distração	apreensão	tédio
(-G(1/3), -G(2/3)]	tristeza	surpresa	medo	desgosto
(-G(2/3), -G(1)]	angústia	perplexidade	terror	repugnância

A forma dimensional é definida como “vetor de sentimentos” e é um vetor em ponto flutuante de quatro dimensões que pode sintetizar toda gama de experiências emocionais em termos de Amenidade, Sensibilidade, Atenção e Atitude. No modelo, a dimensão vertical representa a intensidade das diferentes dimensões afetivas e seus níveis de ativação, enquanto a dimensão radial representa as linhas K [31] que podem ativar as configurações mentais, que podem precisar de segundos ou até mesmo anos. O modelo segue o padrão utilizado na teoria das cores permitindo combinações entre as mesmas, por exemplo a emoção resultante da combinação de duas ou mais emoções assim como a combinação das cores vermelha e azul resultam no roxo que pode ser visualizado na Figura 3.

Ainda referente aos vetores de sentimentos, alguns têm nomes especiais, tratam-se de composições emocionais descritas na Tabela 3. Por exemplo, o vetor de sentimento compreendido com Atitude e Amenidade positivas compreende o sentimento Amor.

Tabela 3 – Sentic Level 2

	Atenção>0	Atenção<0	Atitude>0	Atitude<0
Amenidade>0	otimismo	frivolidade	amor	regozijo
Amenidade<0	frustração	desaprovação	inveja	remorso
Sensibilidade>0	agressivo	rejeição	rivalidade	desprezo
Sensibilidade<0	ansiedade	temor	submissão	coerção

Emoções mais complexas podem ser sintetizadas com a combinação de 3 ou até mesmo 4 *sentic levels*, por exemplo a combinação de Alegria, Raiva e Confiança resulta em Ciúmes. Entretanto, de forma análoga ao modo como ocorre a combinação das cores primárias na geração de cores secundárias, as emoções primárias da HOE podem ser igualmente misturadas e resultar numa vasta gama de experiências humanas emocionais.

2.2 Problemas de Classificação com Aprendizado de Máquina

Aprendizado de Máquina é uma área de Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. De modo sintético tem-se dentre os principais conceitos de AM:

- Representação - Armazenar o conhecimento
- Mecanismo de inferência - Aplicar o conhecimento para resolver um problema
- Aprendizagem - Adquirir novos conhecimentos

Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas por meio de solução bem sucedida de problemas anteriores. É uma ferramenta poderosa para aquisição automática de conhecimento, entretanto, não existe um único algoritmo que apresente melhor desempenho para todos problemas.

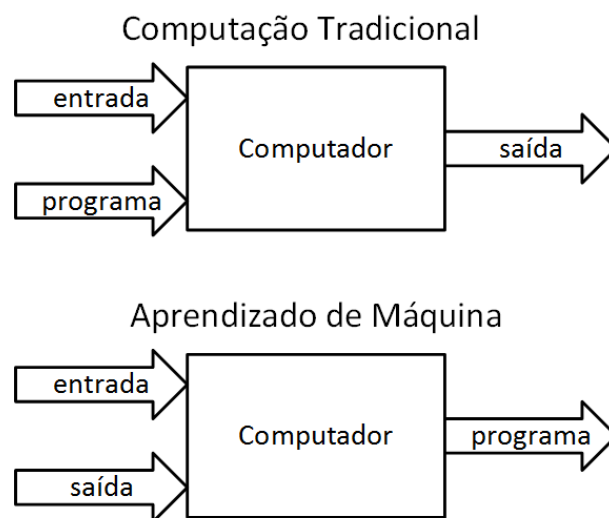


Figura 5 – Computação Tradicional e Aprendizado de Máquina

Podemos diferenciar a computação tradicional do AM que para a computação tradicional temos os dados de entrada e um programa submetidos a um computador que gera uma saída. Em AM para uma entrada e uma saída submetidas a um computador geramos um programa, conforme visualizado na Figura 5.

2.2.1 Hierarquia de Aprendizado

Enquanto os humanos usam raciocínio dedutivo para deduzir nova informação a partir de informações relacionadas logicamente, em AM o conceito é aprendido efetuando-se inferência indutiva sobre os exemplos apresentados, cuja indução é a forma de inferência lógica que permite obter conclusões a partir de um conjunto de exemplos.

Na Figura 6 é exibida a hierarquia de aprendizado encabeçada pelo Aprendizado Indutivo que tem por objetivo encontrar uma hipótese que concorde com o conjunto de exemplos. Uma hipótese é considerada satisfatória quando possui capacidade de generalização. O Aprendizado Indutivo divide-se em dois grupos onde o primeiro compreende os mecanismos de Aprendizado Supervisionado que recebe esse nome porque a supervisão se dá por meio das

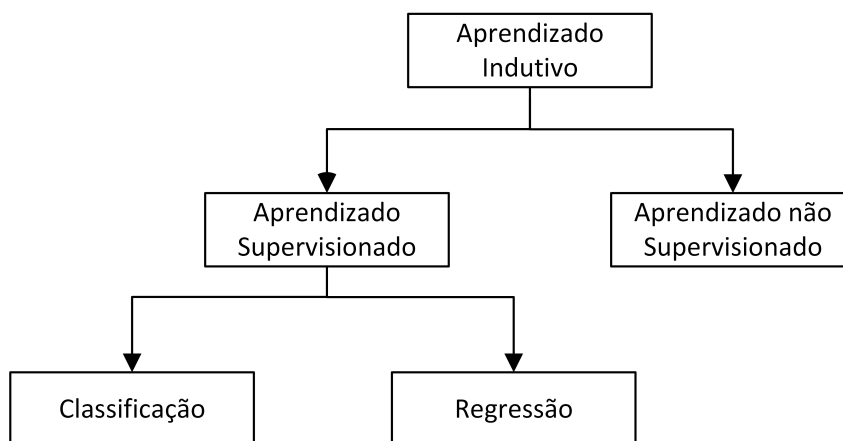


Figura 6 – Hierarquia do Aprendizado

observações do conjunto de treinamento por rótulos indicando as classes às quais cada exemplo pertence. No Aprendizado não Supervisionado não existe classe pré-definida para nenhum dos atributos. Um conjunto de observações é dado com o propósito de se estabelecer a existência das classes ou agrupamentos.

Relativos ainda ao Aprendizado Supervisionado temos os métodos de Classificação e Regressão, no qual o primeiro tem por objetivo prever a classe do exemplo dados seus valores de atributos preditores, e cada exemplo pertence a uma classe pré-definida. Na Classificação os rótulos assumem valores discretos, na regressão os rótulos assumem valores contínuos.

2.2.2 Processo de Aprendizado

O processo de aprendizado tem como saída um classificador/preditor que é gerado pela entrada de dados conforme ilustrado na Figura 7. A primeira etapa se dá com a aquisição de dados seguida por uma etapa de Preparação dos Dados onde ocorre a eliminação ou diminuição de ruídos e redução de dimensionalidade dos conjuntos de dados. Na preparação são eliminados exemplos incorretos, transformação do formato dos exemplos para que sejam adequados ao indutor. A redução de dimensionalidade se dá também pela seleção de atributos relevantes para o problema em questão.

O conhecimento do domínio do especialista é um ente importante dentro do processo de aprendizado pois o conhecimento a respeito do problema influencia a qualidade do classificador gerado pelo indutor. Os itens de conhecimento do domínio podem ser avaliados quanto:

- Informação sobre valores válidos de um atributo
- Critérios para escolher atributos
- Critérios para escolher hipóteses
- Restrições no relacionamento dos atributos

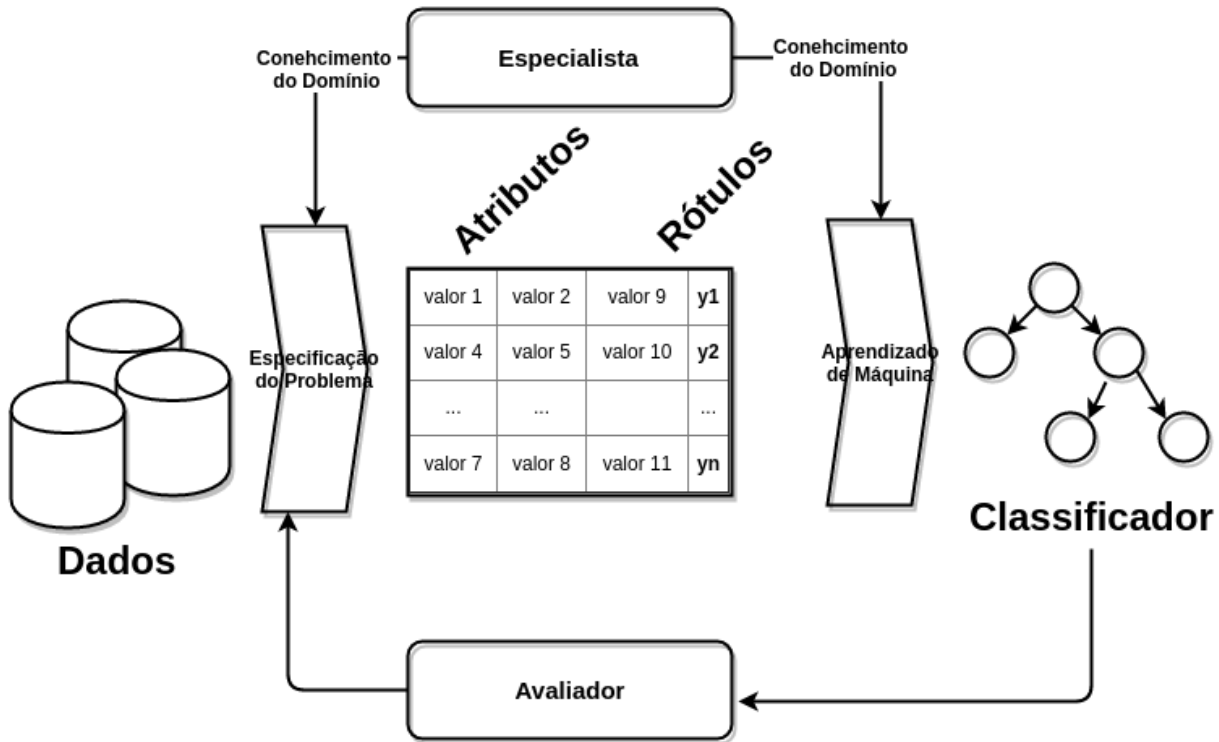


Figura 7 – Processo de Aprendizado

- Regras para geração de conceitos de nível mais alto
- Construção de novos atributos derivados dos atributos originais

2.2.3 Paradigmas do aprendizado

Quanto aos paradigmas do aprendizado temos:

- **Simbólico** - Busca aprender construindo representações simbólicas (expressão lógica, árvores de decisão regras).
- **Estatístico** - Busca métodos estatísticos (Aprendizado Bayesiano)
- **Baseados em Exemplos** - Sistemas lazy (RBC, *K Nearest Neighbors*).
- **Conexionista** - Modelos inspirados no modelo biológico do sistema nervoso (Redes Neurais).
- **Evolutivo** - Teoria de Darwin (Algoritmos Genéticos).

Os sistemas de aprendizado simbólico buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contraexemplos desse conceito. As representações simbólicas estão tipicamente na forma de alguma expressão lógica como por exemplo, as árvores de decisão, regras ou redes semânticas.

Nos sistemas de aprendizado estatísticos a ideia geral consiste em utilizar modelos estatísticos para encontrar uma boa aproximação do conceito induzido. Vários desses métodos são paramétricos, assumindo alguma forma de modelo, e então encontrando valores apropriados para os parâmetros do modelo a partir dos exemplos. Dentre os métodos estatísticos, destacam-se os de Aprendizado Bayesiano, que utilizam um modelo probabilístico baseado no conhecimento prévio do problema, o qual é combinado com os exemplos de treinamento para determinar a probabilidade final de uma hipótese.

Para os sistemas baseados em exemplos um modo de classificar um exemplo é lembrar de outro similar cuja classe é conhecida e assumir que o novo exemplo terá a mesma classe. Essa filosofia exemplifica os sistemas baseados em exemplos, que classificam exemplos nunca vistos através de exemplos similares conhecidos. Os algoritmos mais conhecidos neste paradigma são os de Vizinhos mais Próximos (*Nearest Neighbours*) e Raciocínio Baseado em Casos (*Case Based Reasoning*).

Dentro da abordagem conexionista, a representação de uma Rede Neural envolve unidades altamente interconectadas e, por esse motivo, o nome conexionismo é utilizado para descrever a área de estudo. As analogias com a biologia têm levado muitos pesquisadores a acreditar que as Redes Neurais possuem um grande potencial na resolução de problemas que requerem intenso processamento sensorial humano, tais como visão e reconhecimento de voz.

O paradigma evolutivo é derivado do modelo evolucionário de aprendizado onde um classificador genético consiste de uma população de elementos de classificação que competem para fazer a predição. Este paradigma possui uma analogia direta com a teoria de Darwin, na qual sobrevivem os mais bem adaptados ao ambiente. Alguns operadores genéticos básicos que aplicados a população geram novos indivíduos são: Reprodução, Cruzamento, Mutação e Inversão.

2.3 Problema de Classificação Multi-Classe

AM Supervisionado consiste em extrair conhecimento de um dado conjunto com n exemplos de entrada x_1, x_2, \dots, x_n caracterizado por um número i de características/descriptores $a_1, a_2, \dots, a_i \in A$, que podem ser valores numéricos ou nominais, onde cada instância está associada a uma determinada saída y_j cujo objetivo é elaborar um sistema capaz de predizer a saída de novos exemplos com saídas desconhecidas com razoável nível de acerto. Esta saída pode ser um valor contínuo $y_j \in R$ ou um rótulo de classe $y_j \in C$ (considerando um problema de m classes como $C = c_1, c_2, \dots, c_m$). O primeiro caso é um problema de regressão enquanto o segundo é um problema de classificação [45]. Na classificação, o sistema produzido por um algoritmo de aprendizado é uma função de mapeamento sobre padrões definido $A^i \rightarrow C$ e conhecida como classificador.

A tarefa de classificação é muito usada em aplicações do mundo real. Muitas delas di-

zem respeito a problemas de classificação que envolvem mais de duas classes, razão pela qual são chamadas de problemas multi-classes. Este domínio de aplicações é muito diverso e podemos encontrar problemas multi-classes dentro da bioinformática, visão computacional, na medicina, dentre outras áreas. A tarefa de construir um classificador capaz de distinguir duas classes é mais branda em comparação à construção de um classificador com mais de duas classes. Por este motivo as técnicas de binarização de problemas multi-classes têm se confirmado como boas soluções através da decomposição do problema de multi-classe em vários problemas binários. Em consequência das estratégias de decomposição, temos a geração de múltiplos classificadores binários que são conhecidos, neste caso, como classificadores básicos [46].

Na Figura 8 é possível visualizar as etapas da abordagem de PCMC adotada neste trabalho contendo as tarefas de Decomposição e Agregação, onde OVA e OVO são as estratégias de decomposição de um problema inicialmente multi-classes em vários problemas binários, enquanto abaixo dos quadros OVO e OVA identificam-se os respectivos métodos de agregação que serão detalhados na Sub-seção 2.3.2.



Figura 8 – Esquema geral da abordagem de Decomposição e Agregação adotada neste trabalho.

Para responder à necessidade de melhor compreensão o restante desta Seção foi dividida em duas sub-seções. A Sub-seção 2.3.1 apresenta as principais técnicas de decomposição. Por fim a Sub-seção 2.3.2 apresenta os métodos de agregação pertinentes aos métodos OVO e OVA.

2.3.1 Técnicas de Decomposição

Diferentes estratégias de decomposição podem ser encontradas na literatura, entretanto no bojo deste trabalho serão abordadas as estratégias OVA [47] e OVO [48].

One vs One

(OVO) Esta estratégia consiste em dividir o problema em vários problemas binários com todas possíveis combinações entre pares de classes. Assim um classificador é treinado para discriminar entre o par de classes envolvido no problema, então as saídas geradas pelos classificadores básicos são combinadas de maneira a prever a classe de saída. Especificamente a estratégia OVO divide um problema de K classes em $K(K - 1)/2$ problemas binários que é responsável por distinguir entre um par de classes. A fase de aprendizado é realizada utilizando somente os dados de treinamento, que contêm apenas instâncias rotuladas com o par de classes envolvidas no problema binário e as demais instâncias são ignoradas assim como mostrado na Figura 9.

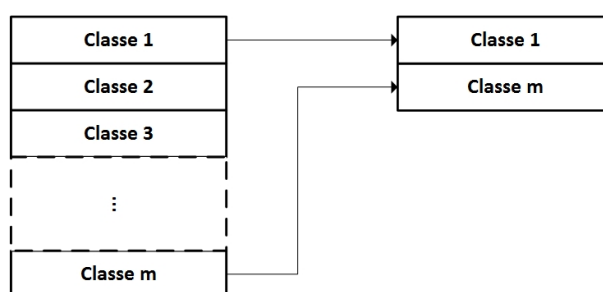


Figura 9 – Decomposição OVO

Na fase de validação, um padrão é apresentado para cada um dos classificadores binários. A saída do classificador é dada pela confiança, $r_{ij} \in [0, 1]$, do classificador binário distinguindo as classes i e j em favor da primeira classe. A confiança do classificador para classes posteriores é computada por $r_{ji} = 1 - r_{ij}$ caso o classificador não gere diretamente (a classe com maior confiança é a classe de saída do classificador). Estas saídas são representadas por uma matriz R tal qual demonstrada na Equação 2.1:

$$R = \begin{pmatrix} - & r_{12} & \dots & r_{1m} \\ r_{21} & - & \dots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \dots & - \end{pmatrix} \quad (2.1)$$

O resultado final do sistema é derivado de diferentes métodos de agregação aplicados sobre a matriz R . Mesmo com o número de classificadores sendo da ordem m^2 , cada classificador é treinado apenas com os exemplos das classes envolvidas no problema que consequentemente faz com que o tempo de classificação não seja elevado. Esta é uma vantagem em usar apenas exemplos envolvidos no problema.

One vs All

Esta abordagem treina os classificadores para cada uma das classes envolvidas no problema, e a classe é discriminada perante as demais classes envolvidas no problema. Assim o classificador básico diante de uma resposta positiva indica a classe de saída. O método de decomposição OVA divide um problema multi-classe de ordem m em m classificadores binários. Cada problema é transformado num classificador binário que é responsável por distinguir uma das classes diante das classes restantes. A etapa de aprendizado do classificador é realizada usando todos os dados de treinamento assim como apresentado na Figura 10.

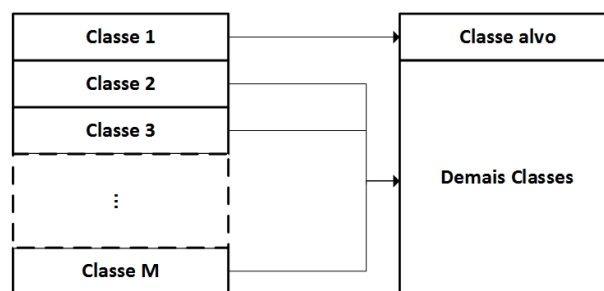


Figura 10 – Decomposição OVA

Na fase de validação, é representado um padrão para cada um dos classificadores binários e só então o classificador que indicar uma saída positiva será a classe predita. Em alguns casos, mais de um classificador pode apresentar uma saída positiva fazendo-se necessário o uso de técnicas de desempate para definir classe predita. A forma mais comum de definir a saída é pelo uso da confiança. Diferentemente da abordagem OVO os classificadores OVA formam um Vetor R em vez de uma Matriz R . conforme pode ser visualizado na Equação 2.2:

$$R = (r_1, r_2, r_3, \dots, r_m), \quad (2.2)$$

onde $r_i \in [0, 1]$ é a confiança da classe m .

2.3.2 Métodos de Agregação

Nesta Sub-seção descreveremos as técnicas de agregação para as técnicas de binarização exploradas neste trabalho. Dividiremos em duas partes: primeiro descreveremos as combinações orientadas às decomposições OVO, cujas agregações são realizadas em função da Matriz R ; segundo descreveremos as combinações para o esquema OVA, cujas agregações são aplicadas em função do Vetor R .

2.3.2.1 Agregações OVO

A Estratégia de **votação (VOTE)** é o método de agregação mais conhecido para o esquema de decomposição OVO. A estratégia de votação recebe como entrada as confidências r_{ij} advindas da Matriz R geradas pelos classificadores binários básicos. Os “votos” em cada classe são totalizados de modo a se obter o resultados final, onde a classe com maior quantidade de votos é considerada classe predita. A estratégia de votação é denotada pela equação 2.3:

$$Classe = \arg \max_{i=1, \dots, m} \sum_{1 \leq j \neq i \leq m} \begin{cases} 1 & r_{ij} > r_{ji} \\ 0 & \text{demais} \end{cases} \quad (2.3)$$

Pouco diferente da estratégia de **votação** é a estratégia de **votação por peso (WVOT)**, onde cada classificador binário fornece o valor da confidência para cada classe r_{ij} (*classes alvo e não alvo*). O exemplo é associado à classe com a somatória da maior confidência como definido na Equação 2.4:

$$Classe = \arg \max_{i=1, \dots, m} \sum_{1 \leq j \neq i \leq m} r_{ij} \quad (2.4)$$

A classificação por acoplamento de pares foi proposta por [22] baseada na probabilidade das classes. Dado $r_{ij} = \text{Prob}(Classe_i | Classe_i \text{ ou } Classe_j)$, o método procura pela melhor aproximação da probabilidade classes posterior dada por $\hat{p} = (\hat{p}_1, \dots, \hat{p}_m)$ em função das saídas dos classificadores binários. A predição da classe resulta da maior probabilidade posterior, assim como mostrada na Equação 2.5:

$$Class = \arg \max_{i=1, \dots, m} \sum_{1 \leq j \neq i \leq m} \hat{p}_{ij} \quad (2.5)$$

Para gerar a probabilidade posterior, utiliza-se a distância minimizada de Kullback-Leibler (**KULE**) entre r_{ij} e μ_{ij} como mostrado na Equação 2.6

$$l(p) = \sum_{1 \leq j \neq i \leq m} n_{ij} r_{ij} \log \frac{r_{ij}}{\mu_{ij}} = \sum_{i < j} (r_{ij} \log \frac{r_{ij}}{\mu_{ij}} + (1 - r_{ij}) \log \frac{1 - r_{ij}}{1 - \mu_{ij}}) \quad (2.6)$$

onde $\mu_{ij} = p_i + j / (p_i + p_j)$, $r_{ji} = 1 - r_{ij}$ e n_{ij} representa o número de instâncias presentes nos dados de treinamento para a i -ésima e j -ésima classes [10, 22].

O método de classificação por preferência de valor aprendido igualmente (**LVPC**) trata a Matriz R , entretanto como uma relação de preferência nebulosa (*Fuzzy*) [49]; a relação primária é separada em três novas relações com 3 medidas: preferência, conflito e ignorância. Deste modo, por acoplamento de pares o exemplo é associado à classe pela resolução da Equação 2.7:

$$Classe = \arg \max_{i=1, \dots, m} \sum_{1 \leq j \neq i \leq m} P_{ij} + \frac{1}{2} C_{ij} + \frac{N_i}{N_i + N_j} I_{ij} \quad (2.7)$$

onde N_i é o número de instâncias da classe i presentes nos dados de treinamento, e C_{ij} denota o grau de conflito dado por $C_{ij} = \min\{r_{ij}, r_{ji}\}$, e I_{ij} é o grau de ignorância dado por $I_{ij} =$

$1 - \max\{r_{ij}, r_{ji}\}$, enquanto P_{ij} e P_{ji} são as preferências restritas, respectivamente para i e j dadas por $P_{ij} = 1 - \min\{r_{ij}, r_{ji}\}$ e $P_{ji} = 1 - \min\{r_{ij}, r_{ji}\}$.

Para o método Weng (**WENG**), de forma similar, a probabilidade é estimada por acoplamento em pares [50]. Weng estima as probabilidades posteriores de cada classe e é predita a classe com maior valor. No entanto a formulação de otimização é dada pela Equação 2.8:

$$\max_p \sum_{i=1}^m \sum_{1 \leq j \neq i \leq m} (r_{ji}p_i - r_{ij}p_j)^2 \text{ sujeito à } \sum p_i = 1, p_i \geq 0, \forall i \quad (2.8)$$

2.3.2.2 Agregações OVA

Nesta Sub-seção será descrito brevemente o método de predição de classes sob a utilização do Vetor R produzido pela estratégia de decomposição OVA. Neste caso agregação é elaborada com o objetivo de tratar os desempates quando ocorrem de mais de um classificador binário produzir uma resposta positiva para sua classe de trabalho.

A estratégia de confiança máxima (**MCON**) é similar a estratégia de votação por peso do esquema OVO, onde a classe é predita pelo classificador com a maior confiança, dada pela Equação 2.9:

$$\text{Classe} = \arg \max_{i=1, \dots, m} r_{ij} \quad (2.9)$$

2.4 Classificadores básicos

Nesta Sub-seção apresentaremos os algoritmos utilizados para indução de classificadores nos experimentos deste trabalho. Para tal serão descritos os Algoritmos usados em AM já consagrados na academia na solução de PCB. A estratégia de dividir para conquistar adequa-se a abordagem PCMC porque os métodos de decomposição dividem o problema multi-classes em diversos problemas binários onde o uso dos algoritmos encaixam-se perfeitamente na solução de PCMC.

A Sub-seção 2.4.1 apresenta o algoritmo SVM. Na Sub-seção 2.4.2 descreve-se as características básicas do algoritmo AD, seguido pela descrição do algoritmo *Random Forest* na Sub-seção 2.4.3 e por fim na Sub-seção 2.4.4 apresentam-se a características do algoritmo NB.

2.4.1 Support Vector Machine

Os mecanismos de aprendizado SVM constituem uma técnica baseada na Teoria do Aprendizado Estatístico [51] que visa estabelecer condições matemáticas que permitam a escolha de um classificador com bom desempenho para os conjuntos de treinamento e teste. Para isto, a minimização do erro de treinamento e a complexidade da função escolhida para separação dos dados em classes. Com o controle da dimensão VC (Vapnik e Chervonenskis) é possível minimizar o erro de treinamento e a complexidade da função para separação dos dados em classes. As principais características das SVMs que tornam seu uso atrativo são [52]:

1. Boa capacidade de generalização
2. Robustez em grandes dimensões
3. Convexidade da função objetivo
4. Base teórica estatística bem definida

No “coração” da arquitetura de um classificador de SVM estão os conceitos de hiperplano e margem. A equação geral de um hiperplano em \mathbb{R}^N é

$$w^T x + w_0 = 0 \quad (2.10)$$

onde $w \in \mathbb{R}^N$ e é vetor normal para o hiperplano e $b \in \mathbb{R}$ é um escalar. Consequentemente, para um hiperplano que não passe por qualquer ponto, podemos escalar w e b apropriadamente tal que $\min_{(x,y) \in S} |w \cdot x + b| = 1$.

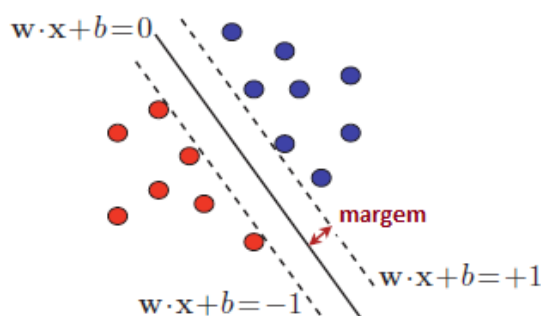


Figura 11 – Margem e equações dos hiperplanos para o hiperplano canônico de maior margem. Os hiperplanos marginais são representados pelas linhas tracejadas. Adaptado de [53]

A Figura 11 ilustra a maior margem para o hiperplano com a representação canônica (w,b) . Ela também mostra os hiperplanos marginais, que são os hiperplanos paralelos para o hiperplano separador que passa pelos pontos mais próximos dos lados positivos e negativos.

A margem de um classificador é definida como a menor distância entre os exemplos do conjunto de treinamento e o hiperplano utilizado na separação destes dados em classes.

Um dos principais problemas das SVMs reside na complexidade computacional na busca de soluções que é na ordem de n^2 ou n^3 , onde n representa o número de exemplos de treinamento. Mesmo que polinomial, onde em se tratando de amostras muito numerosas, as SVMs apresentam tempo de treinamento aceitáveis comparadas a outros algoritmos de AM [53]

2.4.2 Árvore de Decisão

AD é induzida a partir de exemplos de treinamento onde as classes são previamente conhecidas. A estrutura de uma árvore é assim organizada:

1. Cada nó interno é rotulado com o nome de um dos atributos previsores;
2. As arestas saindo de um nó interno são rotuladas com valores do atributo daquele nó;
3. Cada folha é rotulada com uma classe na qual é prevista para exemplos que pertençam àquele nó.

A classificação de um exemplo se dá fazendo com que o mesmo “caminhe” pelos nós, a partir da raiz, percorrendo as arestas que ligam os nós, obedecendo às condições impostas nas mesmas. Uma vez que se atinge o nó folha, a classe que rotula o nó é atribuída ao exemplo.

No espaço bi-dimensional definido pelos atributos os nós são representados por hiper-retângulos (w_1, w_2, w_3) , e a intersecção destes é um conjunto vazio, conforme visualizado na Figura 12.

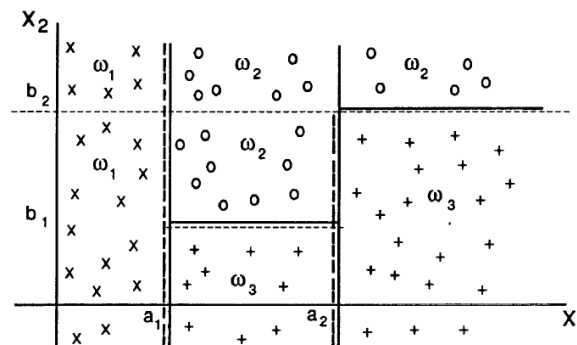


Figura 12 – Espaço bi-dimensional demonstrando as possíveis partições [54].

O ponto crucial de um algoritmo de construção de uma árvore de decisão é a escolha do atributo para rotular o nó atual da árvore. O ideal é que se escolha o atributo com maior poder de discriminação entre as classes para os exemplos do nó atual. Assim é preponderante o uso de uma medida de discriminação das classes. As principais medidas, baseadas na Teoria da Informação [55] utilizadas no Algoritmo C4.5, são:

1. Ganho de Informação
2. Taxa de Ganho
3. Regras de Reprodução

Os classificadores de árvores de decisão demonstram grande potencial em muitos problemas de reconhecimento de padrão. Talvez uma das principais características seja sua flexibilidade, como por exemplo, a capacidade de usar diferentes conjuntos de características e regras de decisão em diferentes estágios da classificação e uma boa relação entre a acurácia na classificação e performance de treinamento quanto ao tempo [54].

2.4.3 *Random Forests*

Floresta Aleatória ou RF é uma classe de métodos de conjuntos desenvolvidos para classificadores de árvore de decisão. Ele combina previsões feitas por múltiplas árvores de decisão, onde cada árvore é gerada a partir dos valores de um vetor aleatório assim como visto na Figura 13. Estes vetores são gerados a partir de uma distribuição de probabilidade fixa [56].



Figura 13 – Esquema de uma Floresta Aleatória. Adaptado de [57]

Esse vetor pode selecionar aleatoriamente F características de entrada dos dados, e baseado nessas F características fazer a divisão dos nós da árvore. Depois das construções das árvores, as previsões são combinadas utilizando a votação da maioria. Essa abordagem é conhecida como Floresta-RI, onde RI faz referência à seleção aleatória de entrada. O processo conhecido como *bagging* também pode ser utilizado para gerar amostras de *bootstrap* para a Floresta-RI adaptado [56, 58].

Esse F escolhido determina a correlação entre as árvores, se F for suficientemente pequeno, então as árvores tendem a se tornar menos correlacionadas. Por outro lado, se for grande a força do classificador tende a melhorar. Tem-se uma outra abordagem conhecida por Floresta-RC (RC faz menção ao procedimento que usa combinações lineares de entradas), onde o número de entradas das características é um número pequeno, dificultando a escolha do F . Dessa forma, é criado um subconjunto através de combinações lineares das características de entrada[59].

Existe uma terceira abordagem, em que o algoritmo escolhe as F melhores divisões de cada nó da árvore. Esse fato pode conduzir a formação de árvores mais correlacionadas, mas pelo fato de ter que fazer essa procura pelos melhores F , essa abordagem acaba gastando um tempo maior de processamento que as demais abordagens. Todas as árvores são geradas em sua integridade, ou seja, não há poda em nenhuma delas [59].

RF é uma ferramenta efetiva de predição e pode-se delinear entre suas vantagens: a acurácia não inferior Adaboost porém é mais rápida e não produz *over-fitting* (quando o modelo é sobre-ajustado ao conjunto de dados/amostra); trabalha melhor com ruído em função da seleção aleatória de características.

2.4.4 Naïve Bayes

O classificador NB é provavelmente o classificador mais utilizado em AM. É um classificador denominado como ingênuo por assumir que os atributos são condicionalmente independentes. Mesmo sendo ingênuo, o classificador reporta o melhor desempenho em várias tarefas de classificação [60]. Na tarefa de classificação, dado um padrão tem-se a tarefa de classificá-lo numa determinada classes c , assumindo ainda que o número de classes é conhecido a priori. Cada padrão é representado por um conjunto de valores representativos de suas características, $x(i), i = 1, 2, \dots, l$, a qual forma um vetor de características de dimensão l dada por $x = [x(1), x(2), \dots, x(l)]^T \in \mathbb{R}$. Assume-se que cada padrão é representado unicamente por um simples vetor de características e que pertence a apenas uma classe [61]. Dado que $x \in R^l$ e um conjunto de classes $c, w_i, i = 1, 2, \dots, c$, a Teoria de Bayes afirma que:

$$P(w_i|x)p(x) = p(x|w_i)P(w_i) \quad (2.11)$$

onde

$$p(x) = \sum_{i=1}^c p(x|w_i)P(w_i) \quad (2.12)$$

onde $P(w_i)$ é a probabilidade anterior da classe $w_i; i = 1, 2, \dots, c$, e $P(w_i|x)$ é a probabilidade posterior da classe w_i dado o valor de x ; $p(x)$ é a Função Densidade de Probabilidade (FDP) de x ; e $p(x|w_i)$ é a FDP da classe condicional de x dada por w_i (também chamada de probabilidade de w_i com relação a x). De acordo com a teoria de decisão de Bayes, x é associado a uma classe w_i se

$$P(w_i|x) > P(w_j|x), \forall j \neq i \quad (2.13)$$

ou, considerando a equação 2.11 e dado que $p(x)$ é positivo e da mesma forma para as demais classes, se

$$P(w_i|x)P(w_i) > P(w_j|x)P(w_j), \forall j \neq i \quad (2.14)$$

O classificador Bayesiano é ótimo no sentido em que minimiza a probabilidade de erro [61].

2.5 Medida de Performance

Dado que este trabalho lida com um *dataset* previamente rotulado com as seis emoções básicas de Ekman, conseqüentemente uma abordagem PCMC faz-se necessária. Ainda que a

maioria das medidas de performance tenham sido elaboradas para trabalhar com problemas binários [62]. Ainda assim, neste trabalho, foi utilizada a acurácia para comparar as taxas de predição porque é a medida largamente utilizada em AM [63]. A medida de acurácia é dada pela Equação 2.15:

$$Acurácia = \frac{\sum_{i=1}^c T_{c_i}}{\sum_{i=1}^c T_{c_i} + \sum_{i=1}^c F_{c_i}} \quad (2.15)$$

onde c é a classe, and T é a quantidade de acertos para uma dada classe c_i enquanto F é a quantidades de erros para uma dada classe c_i .

2.6 Teste Friedman e *post hoc* Nemenyi

O Teste de Friedman é utilizado em casos onde existe a necessidade de se comparar múltiplos domínios e deve ser seguido por um teste *post hoc*, afim de se verificar se existe diferença entre os classificadores. O teste de Friedman é a contrapartida não paramétrica do teste de medidas repetidas *one-way* ANOVA. O teste não paramétrico homólogo do teste t, a análise baseia-se nas fileiras de cada classificador sobre cada conjunto de dados e não sobre as medidas de desempenho explícitos. Novamente, considere n conjuntos de dados e K classificadores a se avaliar. A avaliação se dá da seguinte forma:

Cada algoritmo é ranqueado para cada *dataset* separadamente de acordo sua medida de performance pm , ordenada de forma crescente, do melhor classificador para o pior classificador. Assim, para o *dataset* S_i , o classificador f_i tal que $pm_{ij} > pm_{ji}, \forall j', j, j' \in \{1, 2, \dots, k\}, j \neq j'$, é ranqueado como 1. No caso de desempate para para um dado *rank* r associar o valor extraído de $[(r+1) + (r+2) + \dots + (r+d)]/d$ para cada um dos classificadores empatados. Dado que R seja o *rank* do classificador f_i para o *dataset* S_i .

Computar o *rank* médio do classificador f_i para todos *datasets* pela Equação 2.16:

$$\bar{R}_j = \frac{1}{n} \sum_{i=1}^n R_{ij} \quad (2.16)$$

Computar o *rank* médio para todos classificadores e *datasets* pela Equação 2.17:

$$\bar{R} = \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k R_{ij} \quad (2.17)$$

Computar a soma dos quadrados dos totais denotando a variação em termos de *rank* pela Equação 2.18:

$$SS_{Total} = n \sum_{j=1}^k (\bar{R}_j - \bar{R})^2 \quad (2.18)$$

Computar a soma dos quadrados dos erros denotando a variação de erro pela Equação 2.19:

$$SS_{erro} = \frac{1}{n(k-1)} \sum_{i=1}^n \sum_{j=1}^k (\bar{R}_j - \bar{R})^2 \quad (2.19)$$

Computar a soma dos quadrados dos erros denotando a variação de erro pela Equação 2.20:

$$SS_{erro} = \frac{1}{n(k-1)} \sum_{i=1}^n \sum_{j=1}^k (\bar{R}_j - \bar{R})^2 \quad (2.20)$$

O teste estatístico de Friedman é calculado pela equação 2.21

$$X_F^2 = \frac{SS_{Total}}{SS_{Erro}} \quad (2.21)$$

De acordo com o teste de hipótese nula, os estados de todos os classificadores são equivalentes em seu desempenho e conseqüentemente seus *ranks* R_{ij} devem ser iguais, X_F^2 atende a uma distribuição X^2 com grau de liberdade $k - 1$ para um n usualmente maior que 15 e um k usualmente maior que 5.

O teste Nemenyi computa uma estatística q sobre a diferença entre os *ranks* médios dos classificadores obtidos pelo teste Friedman. Para quaisquer dois classificadores se calcula a estatística q conforme a Equação 2.22:

$$q = \frac{\bar{R}_{j1} - \bar{R}_{j2}}{\sqrt{\frac{k(k+1)}{6n}}} \quad (2.22)$$

A hipótese nula é rejeitada após a comparação do valor q obtido com o o valor q desejado para o nível de significância da tabela de significância para os valores críticos q_α , onde α refere-se ao nível de significância. Rejeita-se a hipótese nula se o valor q obtido excede o valor de q_α .

2.7 Trabalhos Relacionados

Esta Seção apresenta os trabalhos relacionados encontrados na literatura referentes aos assuntos tratados neste trabalho. Serão descritos trabalhos da literatura recente acerca do assunto referente à Análise de Sentimentos, trabalhos que abordam resolução de problemas multi-classes.

A AS, mais especificamente a classificação de sentimentos, é a determinação da orientação dos sentimentos para um dado texto em duas ou mais classes. Esta atividade pode ser realizada levando em consideração diversas abordagens e, dentre essas, podem citar-se a abordagem baseada em AM ou baseada em *lexicon* ou até mesmo de forma mista [30]. Dentro da abordagem de AM distinguem-se as abordagens supervisionadas e não-supervisionadas. Para as abordagens supervisionadas os classificadores mais utilizados são AD, SVM, Redes Neurais, NB e Máxima Entropia. Dentro da abordagem baseada em *lexicon* é possível usar a abordagem baseada em dicionário e baseada em *corpus*. Com grande frequência dentro da literatura recente os trabalhos pertinentes a AS resolvem problemas binários (polaridade com sentimentos negativos ou positivos) e ternários com a inclusão de sentenças neutras.

A classificação de sentimentos pode ser realizada em fontes de dados como avaliação de produtos, *blogs*, artigos de notícias e *microblog*. Entretanto há necessidade de um alto nível de

processamento assim como técnicas avançadas de análise. Os trabalhos de [64, 65, 66, 67] demonstraram que as aplicações de técnicas diversas de classificação de sentimentos são eficazes em fontes de dados de *microblogs*. Pode-se ainda ser agrupada dentro das abordagens baseadas em AM, baseadas em *lexicon* e híbridas.

Quanto aos trabalhos fundamentalmente realizados utilizando a abordagem baseada em AM temos [68] como pioneiro quando aplicou técnicas de AM envolvendo NB, Entropia Máxima e SVM na determinação de polaridade com uma fonte de dados de avaliações de filmes de cinema. Para realização de seus experimentos coletou-se *reviews* do site IMDB¹. No seu conjunto de experimentos, contemplando vários tipos de descritores, obteve o melhor resultado com 82,9% de acurácia com o classificador SVM utilizando unigramas. Argumentaram ainda que a análise de discurso em conjunto com resolução de co-referências poderia melhorar a acurácia.

McDonald et al. [69] apresentaram um modelo estruturado para classificação de sentença sentenças e documentos. Eles representaram os documentos e suas sentenças como pequenos grupos exclusivos assumindo que os rótulos das sentenças e documentos eram interdependentes. Aplicaram o algoritmo de Viterbi [70] para inferência e o algoritmo MIRA foi empregado na classificação [71, 72] e utilizaram como descritores unigramas, bigramas, trigramas em conjunto com *part of speech*. Como fonte de dados utilizaram 600 *reviews* do site Amazon² para 3 domínios distintos. Como melhor resultado obtiveram 62,5% de acurácia no nível de sentenças e 82,8% no nível de documento.

Dang et al. [73] usando SVM e diferentes métodos de seleção de características classificaram sentimentos em termos de polaridade. Como fonte de dados utilizaram dois *datasets* o primeiro *corpus* com 305 *reviews* positivos e 307 *reviews* negativos de uma câmera digital; o segundo *dataset* com múltiplos domínios [74]. Para treinar a SVM usou-se 3 coleções de características: independentes do domínio, dependentes do domínio e descritores de sentimentos. Como forma de redução de dimensionalidade utilizou ganho de informação para os distintos conjuntos indiscriminadamente. Os experimentos com redução de dimensionalidade surtiram melhor efeito para o *dataset* de domínio múltiplo com uma acurácia de 84,15%.

Saleh et al. [75] realizaram vinte e sete experimentos distintos usando SVM com vários métodos de seleção de características. Seus experimentos foram realizados com *corpus* de Pang e Lee [76], com *corpus* de múltiplos domínios de Taboada e Grieve [77] e o *corpus* com *reviews* de câmeras digitais SINAI. Usando validação cruzada de 10 pastas atingiu as melhores acurácias em 83,5% para o *corpus* de Pang e Lee usando ocorrências binárias e trigramas, 73,25% para o *corpus* de Taboada e Grieve usando *Term Frequency–Inverse Document Frequency* (TF-IDF) e trigramas, 91,51% para o *corpus* SINAI usando TF-IDF e bigramas.

Zhang et al. [78] classificaram sentimentos utilizando algoritmos de AM, SVM e NB, para um *corpus* de avaliações de restaurantes escritos em cantonês contendo 1500 positivos e

¹ www.imdb.com

² www.amazon.com

1500 negativos. Eles estudaram os efeitos das representações e tamanhos dos descritores quanto a performance. Eles investigaram várias representações descritores tais como unigramas e unigramas/frequências, bigramas e bigramas/frequência variando o número de descritores de 50 a 1500. A maior acurácia obtida foi de 95,67% para NB com 900 - 1100 descritores.

Wang et al. [79] compararam o desempenho de três métodos *ensemble* clássicos: *bagging*, *boosting*, e um subespaço aleatório baseado em cinco classificadores básicos: NB, Máxima Entropia, AD, *K-Nearest Neighbor* e SVM. Os experimentos utilizaram 10 diferentes *datasets* e reportaram a melhor acurácia e custo computacional na combinação com classificadores básicos.

Moraes et al. [80] compararam a SVM e NB com Redes Neurais Artificiais para classificação de sentimentos. Os experimentos realizados recaíram sobre *datasets* balanceados e desbalanceados. Dentre os 4 *datasets* como balanceado foi utilizado um *dataset* de *reviews* de filmes de cinema e como desbalanceados utilizou três *datasets* de *reviews* de produtos (GPS, Livros e Câmeras digitais). Para os *datasets* desbalanceados tanto as Redes Neurais Artificiais como a SVM obtiveram resultados satisfatórios. Outro resultado relevante foi a constatação de que o método de seleção de características, ganho de informação, não contribuiu em termos de resultado para mais de mil descritores.

Nos últimos anos, têm sido elaborados muitos métodos combinando as saídas geradas pelos classificadores básicos, como por exemplo as novas abordagens com *frameworks* de estimativas probabilísticas [50], estratégias baseadas em árvores binárias [81], esquemas de classificação dinâmicos [82], métodos de relações de preferências [83], e ainda os métodos tradicionais de combinações por acoplamento de pares [22], regras de maiores vencedores [84].

Dentro da literatura especializada, existem alguns trabalhos comparando estas técnicas entre as estratégias OVA e OVO e as diferentes estratégias de decomposição. Hsu et al. [85] estudaram os métodos de decomposição OVO e OVA junto com os métodos correção de erros de saídas, mas somente com algoritmos de SVM com suporte multi-classes enquanto Lorena et al. [86] apresentaram um conjunto de métodos de binarização mas sem compará-los entre si. Fürnkranz [87] comparou a aplicabilidade da estratégia OVO em algoritmos de AD junto com outros métodos *ensemble* tais como *bagging* e *boosting* demonstrando uma melhoria com o uso de valores de confiança na combinação das saídas. Wu et al. [50] realizaram um comparação sistemática dos métodos contidos no *framework* de estimativas de probabilísticas, mas não incrementaram novos métodos de agregação.

3 METODOLOGIA

Este Capítulo apresenta as etapas percorridas para execução e configurações dos experimentos. Na Seção 3.1 é oferecida descrição superficial dos experimentos, ferramentas utilizadas e configurações. Na Seção 3.2 serão descritas as características e origem da Base de Dados. A Seção 3.3 descreve as tarefas responsáveis pela redução de dimensionalidade e formação da Matriz Documento. A Seção 3.4 descreve o método de extração das características textuais baseadas na HOE e SLs e o modo como foram elaborados os cálculos das características derivadas da quantidade de palavras ou *Word*, *Strength* e TF-IDF. A Seção 3.5 descreve a forma como foram construídos os dados de treinamento em consonância com os métodos de decomposição OVA e OVO. Por fim na Seção 3.6 serão descritos como se deram as etapas de treinamento, geração dos classificadores básicos e aplicação dos métodos de agregação de modo a se obter os resultados.

3.1 Experimentos

Considerando os algoritmos de classificação, escolhemos SVM, NB, RF e CART AD porque segundo Fernandez-Delgado et al. [11] estes algoritmos são responsáveis por excelentes resultados. Todos os experimentos foram implementados na Ferramenta R¹. Os pacotes correspondentes usados para implementar cada algoritmo de classificação foram `rpart`, `e1071`, e `randomForest`.

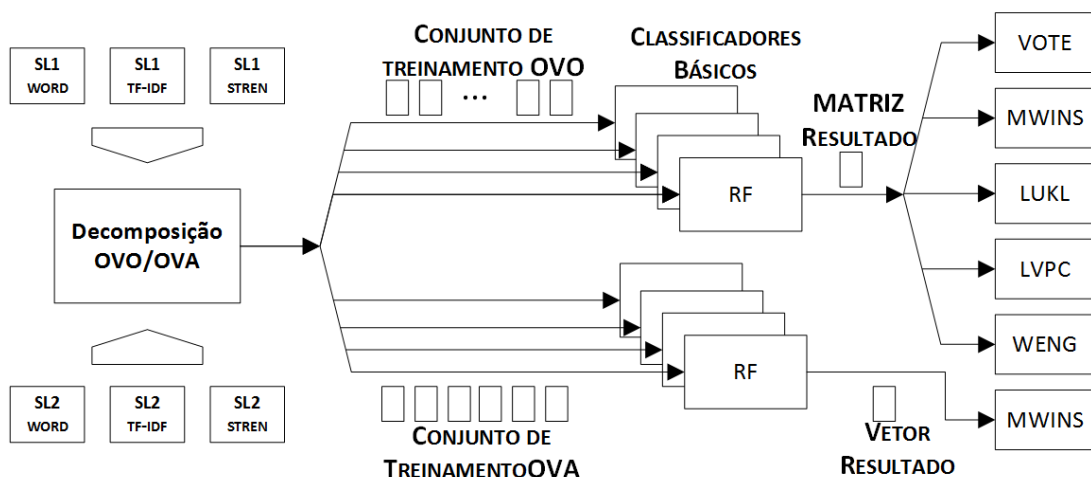


Figura 14 – Sumário dos Experimentos

A Figura 14 mostra de forma resumida os procedimentos dos experimentos. Foram,

¹ <https://www.r-project.org/>

ao todo, realizadas 288 configurações: 4 algoritmos de classificação, 21 sub *datasets* OVO, 7 sub *datasets* OVA, SL1 e SL2 com e sem emoções neutras. Cada configuração de experimento sofreu 30 repetições para se obter uma acurácia média. Uma vez realizadas todas as configurações de experimentos torna-se possível determinar as diferenças críticas entre os métodos de classificação de acordo seus ranqueamentos, para tal foi implementado o teste Friedman [88]. A escolha pelo teste Friedman se deu devido ao fato de que os cenários de experimentos utilizados neste trabalho contemplam vários algoritmos e múltiplos domínios. Neste caso, as diferenças estatisticamente significativas quanto as performances puderam ser detectadas devido a aplicação do teste *post hoc* de Nemeniy.

3.2 Dataset

Para realização deste trabalho, compilou-se uma Base de Dados originária de um conjunto de notícias do site G1². O processo de anotação de um corpus de textos em Português do Brasil. O corpus é composto por textos curtos (título e linha fina de notícias). Utilizou-se esse tipo de textos porque em geral, os trabalhos na área de AS utilizam pequenos textos, como *tweets*, *posts*, avaliações de produtos, etc. O *dataset* resultante contém 2000 sentenças rotuladas conforme o trabalho de [89] com base nas seis emoções básicas de Ekman [9].

Tabela 4 – Amostra de 5 notícias extraídas do site G1

ID	TEXTO	EMOÇÃO
1	Dupla assalta posto de combustíveis no Centro de Pitangui e foge em moto. Suspeitos ameaçaram frentista com uma arma e levaram dinheiro do caixa. Polícia informou que fez rondas, mas ninguém foi preso até o momento.	Desgosto
2	Farc asseguram respeito ao cessar-fogo. As Forças Armadas Revolucionárias da Colômbia (Farc) asseguraram neste domingo seu respeito ao cessar-fogo anunciado pelo comando rebelde há seis dias, mas reforçaram seu direito de se defender caso seja necessário.	Tristeza
3	Tigre branco mata tratador na frente de visitantes em zoo da Nova Zelândia: Ataque ocorreu quando dois tratadores limpavam a jaula do animal. Santuário de felinos foi fechado, e o tigre foi sacrificado.	Tristeza
4	PSOL entra no STF com pedido de explicações contra Paulo Duque: Presidente do Conselho de Ética afirmou que partido 'não existe'. Após interpelação, PSOL pode tomar outras medidas contra peemedebista.	Tristeza
5	Chinês fica 'preso' na Coreia do Sul após filho rabiscar seu passaporte. Menino de 4 anos fez desenhos em todo o documento. Homem foi impedido de viajar pelas autoridades no aeroporto.	Surpresa

Uma pequena fração da Base de Dados original pode ser visualizada na Tabela 4, onde ID representa uma identificação do sentença, TEXTO indica o conteúdo textual da mensagem e emoção indica a emoção unânime resultante da rotulação dos linguistas. Na Tabela 5 é possível visualizar a distribuição dos rótulos emocionais para as sentenças contidas na Base de Dados.

² www.globo.com

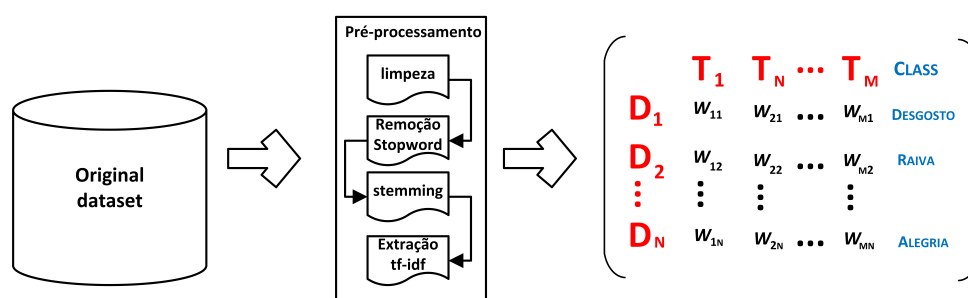
Tabela 5 – Rótulos das seis emoções básicas para o *dataset*

Emoção	# de sentenças	%
Alegria	184	14,00
Desgosto	263	13,15
Medo	222	11,10
Raiva	82	4,15
Surpresa	252	12,60
Tristeza	456	22,80
Neutro	541	27,05

A transformação do *dataset* para uma Matriz Documento e redução de dimensionalidades será descrita na Seção 3.3.

3.3 Preparação dos dados

A etapa de preparação de dados é responsável pela transformação do *dataset* na Matriz Documento. Para tal lançou-se mão de tarefas típicas e bem conhecidas de pré-processamento tais como limpeza, tokenização, remoção de *stopwords*, *stemming* e por fim a extração de TF-IDF. A sequência lógica da execução das tarefas de pré-processamento pode ser visualizada na Figura 15.

Figura 15 – Estratégia de preparação dos dados - *dataset* para Matriz Documento

A tarefa de limpeza compreende a remoção de pontuação, símbolos e todos os tipos de caracteres de espaços brancos (incluindo quebras de linhas). O objetivo dessa tarefa é a redução de dimensionalidade e minimização de ruídos, uma vez que estes caracteres são irrelevantes no contexto. A Tabela 6 exhibe os caracteres envolvidos na limpeza e seus respectivos valores decimais na tabela ASCII.

A tarefa consecutiva à limpeza envolve a eliminação de palavras que não devem ser consideradas no documento, conhecidas como *stopwords*. Tratam-se palavras irrelevantes na análise de textos, justamente por não traduzirem sua essência. Normalmente fazem parte desta lista as preposições, pronomes, artigos, advérbios, e outras classes de palavras auxiliares. Além

Tabela 6 – Lista de caracteres removidos na operação de limpeza

ASCII	DEC	ASCII	DEC	ASCII	DEC	ASCII	DEC
!	33	+	43	?	63	}	125
"	34	,	44	@	64	~	126
#	35	-	45	[91	\t	9
\$	36	.	46	\	92	\n	10
%	37	/	47]	93	\v	11
&	38	:	58	^	94	\f	12
'	39	;	59	_	95	\r	13
(40	<	60	`	96		20
)	41	=	61	{	123		
*	42	>	62		124		

dessas, existem também palavras cuja frequência na coleção de documentos é muito alta. A Tabela 7 relaciona a lista de palavras removidas do documento.

Tabela 7 – Lista de *stopwords* idioma Português

de	ele	aos	tenho	delas	estivesse	houveremos	seremos	tivermos
a	das	seus	lhe	esta	estivéssemos	houveria	serão	tiverem
o	seu	quem	deles	estes	estivessem	houveríamos	seria	terei
que	sua	nas	essas	estas	estiver	houveriam	seríamos	terá
e	ou	me	esses	aquele	estivermos	sou	seriam	teremos
do	é	esse	pelas	aquela	estiverem	somos	tenho	terão
da	quando	eles	este	aqueles	hei	são	tem	teria
em	muito	estão	fosse	aquelas	há	era	temos	teríamos
um	há	você	dele	isto	havemos	éramos	têm	teriam
para	nos	tinha	tu	aquilo	hão	eram	tinha	
com	já	foram	te	estou	houve	fui	tínhamos	
não	é	essa	vocês	está	houvemos	foi	tinham	
uma	eu	num	vos	estamos	houveram	fomos	tive	
os	também	nem	lhes	estão	houvera	foram	teve	
no	só	suas	meus	estive	houvéramos	fora	tivemos	
se	pelo	meu	minhas	estive	haja	seja	tiveram	
na	pela	ás	teu	estivemos	hajamos	sejamos	tivera	
por	até	minha	tua	estiveram	hajam	sejam	tivéramos	
mais	isso	têm	teus	estava	houvesse	fosse	tenha	
as	ela	numa	tuas	estávamos	houvéssemos	fossem	tenhamos	
dos	entre	pelos	nosso	estavam	houvessem	for	tenham	
como	era	ela	nossa	estivera	houver	formos	tivesse	
mas	depois	qual	nossos	esteja	houvermos	forem	tivéssemos	
foi	sem	será	nossas	estejamos	houverem	serei	tivessem	
ao	mesmo	nós	dela	estejam	houverei	será	tiver	

De modo a reduzir palavras ou transformar palavras em um único termo no processo de indexação, é interessante eliminar as variações morfológicas das palavras do documento. Elas são eliminadas através da identificação do radical da palavra. Os prefixos e os sufixos são retirados e os radicais resultantes são adicionados ao índice. Essa técnica de identificação de radicais é denominada lematização ou *stemming*, que em inglês significa reduzir uma palavra ao seu radical (ou raiz). Além da eliminação dos prefixos e sufixos, características de gênero, número e grau das palavras são eliminados.

Neste estágio, a redução da dimensionalidade foi feita com uso de *threshold* e Ganho de Informação selecionando os termos mais representativos. Cada sentença é representada como

um vetor de termos com o TF-IDF calculado conforme as Equações 3.1 e 3.2. Ferreira et al. [90] descrevem esta estratégia como uma eficiente representação textual na forma de vetor. Ao término desta etapa é gerada uma Matriz Documento $D = \{d_1, d_2, d_3, \dots, d_n\}$, onde n representa o número de documentos do *dataset* e $d_n = \{w_1, w_2, w_3, \dots, w_m, class\}$. Em d_n , w é o TF-IDF para o termo m , e m é tamanho do *corpora*, e classe é a emoção anotada para a sentença.

$$w_{m,n} = tf_{m,n} * idf_m \quad (3.1)$$

$$idf_m = LOG \frac{N}{df_m} \quad (3.2)$$

Com a Matriz Documento devidamente formada é possível realizar a extração de características com base no HOE e SL. Estes últimos, assim como as estratégias de decomposição e agregação, são os elementos centrais deste estudo.

3.4 Extração de Características

A extração de características foi realizada por meio da conversão da Matriz Documento em dois grupos conforme ilustrado na Figura 16. O primeiro grupo é formado pelas vinte e quatro percepções de emoções descritas na HOE nos seis diferentes níveis de ativação dentro das quatro dimensões afetivas, também conhecidas como SL1; o segundo grupo consiste das 16 combinações possíveis entre as quatro dimensões, conhecida como SL2. Para cada grupo foram geradas 3 formas derivadas relacionadas aos pesos TF-IDF, *Strength* e número de palavras ou *Words* para cada documento.

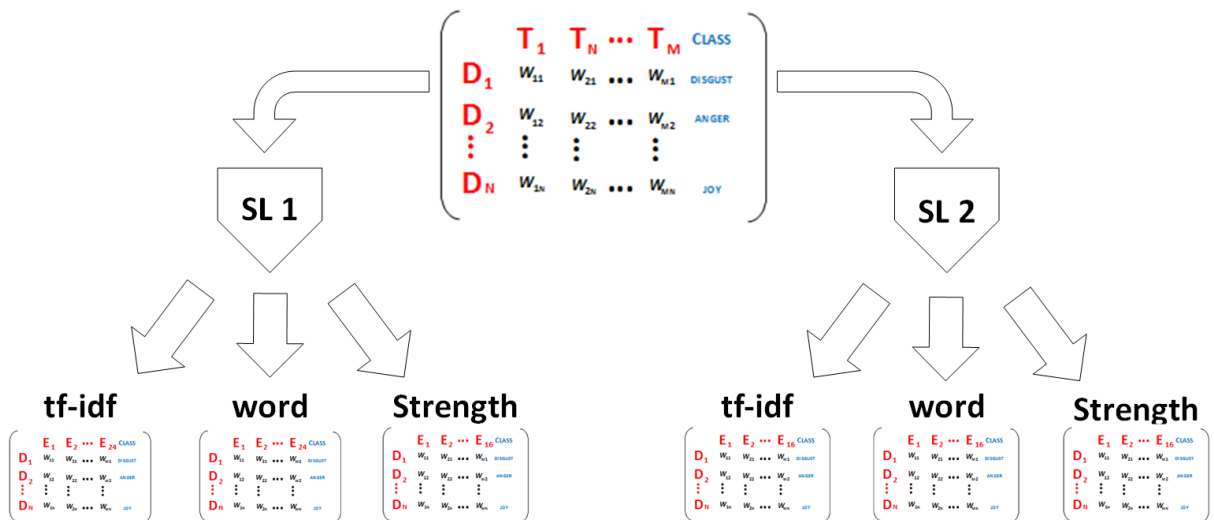


Figura 16 – Esquema de extração das características derivadas relativas ao TF-IDF, *Word* e *Strength* em função dos SL1 e SL2.

Tabela 8 – Sentic Level 1 - *strength* extraídos de 5 sentenças

Emoção	Sentença				
	1	2	3	4	5
excitação	0,00	0,00	0,71	0,85	0,00
vigilância	0,00	0,38	0,00	0,00	1,15
fúria	0,10	0,30	0,05	0,06	0,00
admiração	-0,11	-0,06	0,00	-0,14	-0,20
alegria	-0,55	0,00	0,00	0,00	0,00
antecipação	-0,82	-2,85	0,00	0,00	0,00
raiva	0,00	0,00	0,00	0,00	0,67
confiança	0,00	0,39	0,00	0,00	0,00
serenidade	0,68	0,46	0,35	0,27	0,34
interesse	0,00	0,00	0,00	0,00	0,00
aborrecimento	0,00	0,00	0,00	0,00	0,00
reconhecimento	0,00	0,00	0,00	-0,70	-0,82
pensativo	1,85	0,80	0,00	0,00	0,00
distração	0,52	1,20	0,00	0,00	0,00
apreensão	0,00	0,04	0,01	0,00	0,00
tédio	-0,63	-0,42	0,00	-0,25	-0,49
tristeza	0,00	0,00	0,00	0,00	0,00
surpresa	0,00	0,00	0,00	0,00	0,00
medo	0,00	0,00	0,00	1,83	0,95
desgosto	0,38	0,00	0,00	0,00	0,00
angústia	0,32	0,72	0,27	0,31	0,84
perplexidade	-0,13	0,00	0,00	0,00	0,00
terror	0,00	-0,57	0,00	0,00	0,00
repugnância	-2,31	-0,98	0,00	0,00	0,00
Classe	Desgosto	Tristeza	Tristeza	Tristeza	Surpresa

Para extrair as características desejadas (SL1 e SL2), utilizou-se a API SenticNet 3 [91]. Esta API devolve, para cada termo, quatro valores $dim_{t_n} = (ple, att, sen, apt)$ correspondentes a cada dimensão, onde $\{ple, att, sen, apt\} \in [-1, 1]$. Estas dimensões afetivas são identificadas por Amabilidade, Atenção, Sensibilidade e Atitude e associadas a cada termo. Estes valores são extraídos para cada emoção contidas nos SL1 e SL2 em função das Tabelas 2 e 3. Usando como exemplo a emoção “excitação”, o cálculo de cada grupo (Strength, TF-IDF e Word) é realizado conforme as Equações 3.3, 3.4, 3.5.

$$excitação_{Strength} = \sum_{t=1}^n \begin{cases} dim_{t_n}(ple) & 1 > dim_{t_n}(ple) > \frac{2}{3}; \\ 0 & \text{caso contrário.} \end{cases} \quad (3.3)$$

$$excitação_{TF-IDF} = \sum_{t=1}^n \begin{cases} w_{mn} \frac{dim_{t_n}(ple)}{|dim_{t_n}(ple)|} & 1 > dim_{t_n}(ple) > \frac{2}{3}; \\ 0 & \text{caso contrário.} \end{cases} \quad (3.4)$$

$$excitação_{word} = \sum_{t=1}^n \begin{cases} 1 \frac{dim_{t_n}(ple)}{|dim_{t_n}(ple)|} & 1 > dim_{t_n}(ple) > \frac{2}{3}; \\ 0 & \text{caso contrário.} \end{cases} \quad (3.5)$$

É importante notar que as Equações 3.3, 3.4, 3.5 referem-se somente à extração da emoção “excitação” que é identificada pelos termos contemplados pela dimensão *Pleasantness* com nível de ativação entre 1 e 2/3 conforme informado na Tabela 2 caracterizando uma emoção de SL1. Para extração das demais características é mandatório repetir este processo para todas emoções identificadas nas Tabelas 2 e 3. Ao término deste processo temos como produto os dois grupos de características compostos de *Strength*, TF-IDF e *Word*. Para o grupo SL1 tem-se $d(sl1)_n = \{e_1, e_2, e_3, \dots, e_{24}, class\}$ e para SL2 tem-se $d(sl2)_n = \{e_1, e_2, e_3, \dots, e_{16}, class\}$. Na Tabela 8 é possível visualizar uma parte da matriz derivada *Strength* para o grupo de 24 emoções do SL1. Findada esta etapa é necessário preparar os conjuntos de dados de treinamento e teste que serão elucidados na Seção 3.5.

3.5 Preparação OVA e OVO

O último estágio de preparação antes do treinamento é a adequação das bases de treinamento para os métodos de decomposição OVA e OVO. Para a preparação OVA, primeiramente, o *dataset* deve ser segmentado em seis ou sete partes (sete no caso da inclusão das sentenças neutras) com o objetivo de agregar as sentenças rotuladas com a mesma emoção, conforme demonstrado na Figura 5.

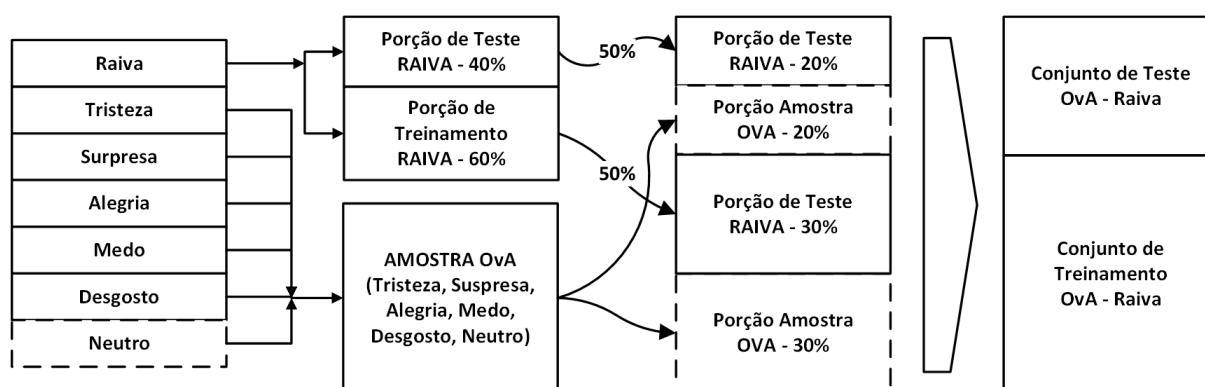


Figura 17 – Preparação OVA para emoção Raiva

Uma vez segmentado o *dataset*, em função dos rótulos, para formar o conjunto de treinamento e teste, cada um destes subconjuntos é montado de forma a atender o método *holdout*, estratificados nas proporções 60% para conjunto de treinamento e 40% para o conjunto de teste em função da classe alvo. Em outras palavras, os conjuntos de treinamento e teste são construídos de forma que se mantenham balanceados, conforme sugere o exemplo da preparação do conjunto de treinamento para a emoção Raiva demonstrada na Figura 17. A preparação OVO, o processo é similar ao OVA, exceto pela re-rotulação das classes não alvo que ocorrem no OVA.

No método de decomposição OVO cada conjunto de treinamento é formado pela junção da Classe alvo com a demais classes de emoções. Ao término deste processo obtém-se para OVA um $K = 7$ conjuntos de treinamentos (K é o número de classes/emoções), enquanto para OVO obtém-se $k(k - 1)/2$ conjuntos de dados de treinamentos. Neste estágio tem-se um conjunto de dados de treinamento a submeter aos algoritmos de AM de modo a treinar os classificadores que por sua vez receberão os dados de teste.

3.6 Decomposição e Agregação

O processo de decomposição, baseado na estratégia dividir para conquistar, consiste no desmembramento de um problema complexo em diversos problemas de menor complexidade, uma vez que ao seu término tem-se múltiplos classificadores binários devidamente treinados e especializados na segregação de duas classes (Emoções).

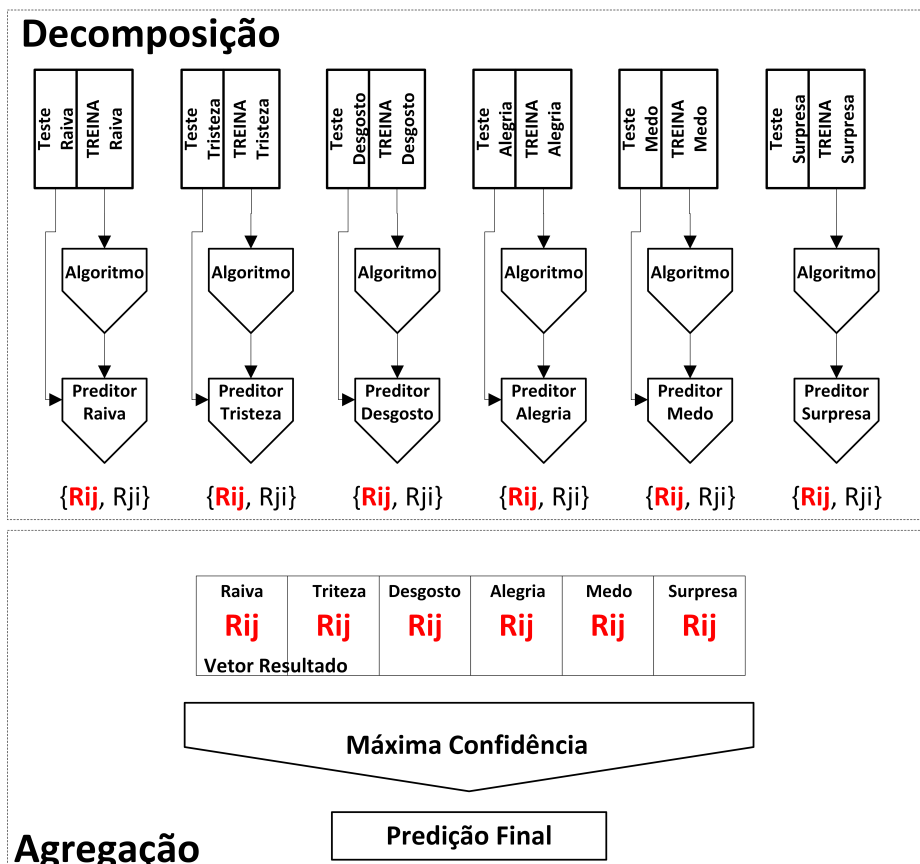


Figura 18 – Esquema geral das tarefas de decomposição e Agregação.

Neste contexto verifica-se a importância da preparação dos dados porque é na preparação que são selecionadas as sentenças rotuladas com as classes envolvidas no problema. Deste modo é possível submeter as porções de treinamento aos algoritmos básicos e produzir os classificadores binários especializados em segregar as classes alvos. Quanto aos métodos de Agregação, estes serão responsáveis pela predição final das classes, uma vez que os mesmos fazem

uso dos graus de confiança gerados por cada classificador. A classificação final é produzida pela aplicação dos métodos de agregação descritos na Seção 2.3.2. Na Figura 18 é possível visualizar as tarefas contidas nos métodos de Decomposição e Agregação sob uma abordagem OVA.

O principal objetivo das tarefas de decomposição é treinar classificadores binários especialistas em discriminar as classes envolvidas no problema binário. Os algoritmos (SVM, AD, RF e NB) serão submetidos a um processo de treinamento, tendo como dados de entrada a porção de dados de treinamento (60%) contendo sentenças rotuladas com as classes envolvidas no problema, por exemplo Raiva contra as demais classes (para uma abordagem OVA). Uma vez com os classificadores devidamente treinados, a porção de dados de teste (40%) é submetida a todos classificadores treinados que por sua vez produzirão os graus de confiança para cada uma das classes envolvidas nos problemas binários e possibilitando a formação da Matriz Resultado (OVO) ou Vetor Resultado (OVA).

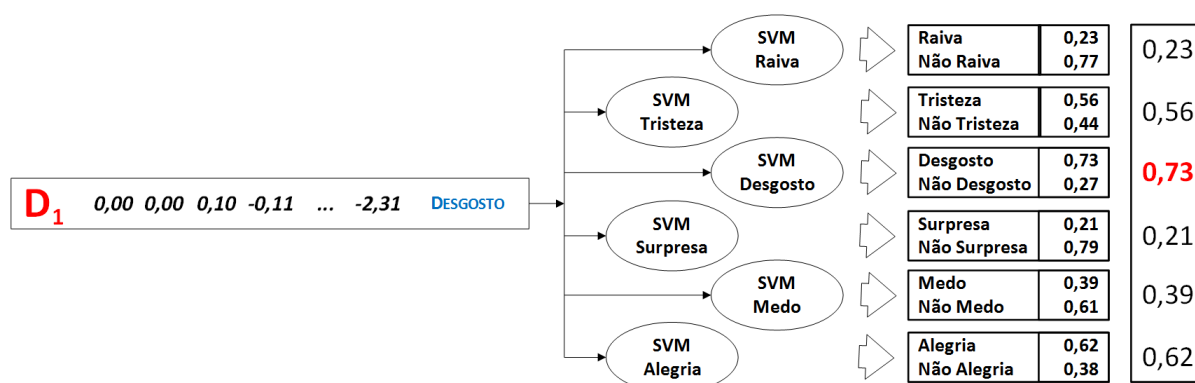


Figura 19 – Aplicação do método de agregação Máxima Confidência (OVA) para sentença preparada com Características SL1/Strength.

Tomando como exemplo a primeira sentença exibida na Tabela 4 e o conjunto de dados de testes preparados e baseados no SL1, e derivado do *Strength* conforme exibido na Tabela 8. Admitindo que submetida cada amostra aos k ou $k(k-1)/2$ (OVA e OVO respectivamente) classificadores, previamente treinados pelo algoritmo SVM, produz-se o grau de confiança para cada uma das classes envolvidas no problema.

A Figura 19 mostra a aplicação do método de agregação Máxima Confidência (OVA) no qual pode-se notar que cada classificador produz os graus de confiança para classes envolvidas. Neste caso a maior confiança é 0,79, implicando que a sentença não pertence a classe “Surpresa”, seguido por 0,77 que não se trata de “Raiva” e por fim 0,73 de confiança que pertence a classes Desgosto. O método de Máxima Confiança atribui que a classes predita deve ser aquela cuja classe contida no conjunto de classes K possui maior grau de confiança, logo

Tabela 9 – Classificadores SVM treinados com dados de treinamento com características *SL1/strength* e respectivos graus de confidências

Classificador SVM		Confidência	
		R_{ij}	R_{ji}
Alegria	Desgosto	0,13	0,87
Alegria	Tristeza	0,75	0,25
Alegria	Medo	0,12	0,88
Alegria	Raiva	0,09	0,91
Alegria	Surpresa	1,00	0,00
Desgosto	Tristeza	0,69	0,31
Desgosto	Medo	0,03	0,97
Desgosto	Raiva	0,65	0,35
Desgosto	Surpresa	0,91	0,09
Tristeza	Medo	0,75	0,25
Tristeza	Raiva	0,45	0,55
Tristeza	Surpresa	0,82	0,18
Medo	Raiva	0,34	0,66
Medo	Surpresa	0,10	0,90
Raiva	Surpresa	0,76	0,24

a classe predita é “Desgosto” e neste caso em particular houve acerto na predição uma vez que o rótulo original é “Desgosto”.

Tabela 10 – Matriz Resultado e cômputo para predição final pelo método de agregação de Votação

	Alegria	Desgosto	Tristeza	Medo	Raiva	Surpresa	Votos
Alegria	—	0,13	0,75	0,12	0,09	1,00	2
Desgosto	0,87	—	0,69	0,03	0,65	0,91	4
Tristeza	0,25	0,31	—	0,75	0,45	0,82	2
Medo	0,88	0,97	0,25	—	0,34	0,1	2
Raiva	0,91	0,35	0,55	0,66	—	0,76	4
Surpresa	0,00	0,09	0,18	0,90	0,24	—	1

Considerando ainda o mesmo cenário, quanto a sentença e características, demonstrando a aplicação dos métodos de agregação de Votação (OVO), são apresentados na Tabela 9 os graus de confiança produzidos pelos classificadores treinado pelo algoritmo SVM. Com as predições produzidas pelos classificadores binários é possível formar a Matriz Resultado conforme demonstrado na Tabela 10.

No caso do método de agregação por Votação assume-se para o resultado de cada Classificador, a classe vencedora é aquela com maior confidência. Ao observar a Tabela 10, o cômputo dos votos se dá ao considerar que um classificador binário, por exemplo ao tomar o classificador Alegria X Tristeza, o mesmo resulta num grau de confidência 0,75 para “Alegria” contra 0,25 para “Tristeza”, assim este classificador produz um voto para a emoção “Alegria”. A coluna

Votos lista os votos para cada um dos sentimentos envolvidos no problema multi-classe. Neste caso tem-se um empate entre as emoções “Desgosto” e “Raiva”, cada qual com 4 votos. Como critério de desempate é usado o classificador com as duas classes empatadas, e será a classe predita aquela com maior confiança, nesta situação temos “Desgosto” com 0,65 contra 0,35 para “Raiva” e assim como no método de Confidência Máxima OVA o método de agregação OVO por votação predisse a classe correta.

Ao término das etapas de Decomposição e Agregação torna-se possível avaliar os resultados e aplicar os testes estatísticos.

4 RESULTADOS

Este capítulo apresenta os experimentos realizados e a discussão dos resultados obtidos dos experimentos. A discussão dos resultados tem como ponto de partida a questão: “As sentenças neutras podem influenciar na classificação?”. A Tabela 11 mostra os 10 melhores resultados em termos de acurácia média e nota-se a inexistência de configurações contendo sentenças neutras. Observando-se ainda o *boxplot* na Figura 20 podemos notar uma diferença de 6,08 pontos percentuais, em acurácia média, na distribuição normal dos resultados. A mediana dos resultados para os experimentos sem a presença das sentenças neutras é de 37,55 % em acurácia enquanto para os experimentos com sentenças neutras é expressivamente menor com 31,47 %. É também relevante destacar que mesmo os experimentos contendo neutros tenham produzido melhores resultados, não conferem maior estabilidade. Uma vez que a distância entre o primeiro quartil e terceiro quartil para os experimentos com neutro dista em 4,48 pontos percentuais em acurácia contra 6,03 dos experimentos sem sentenças neutras.

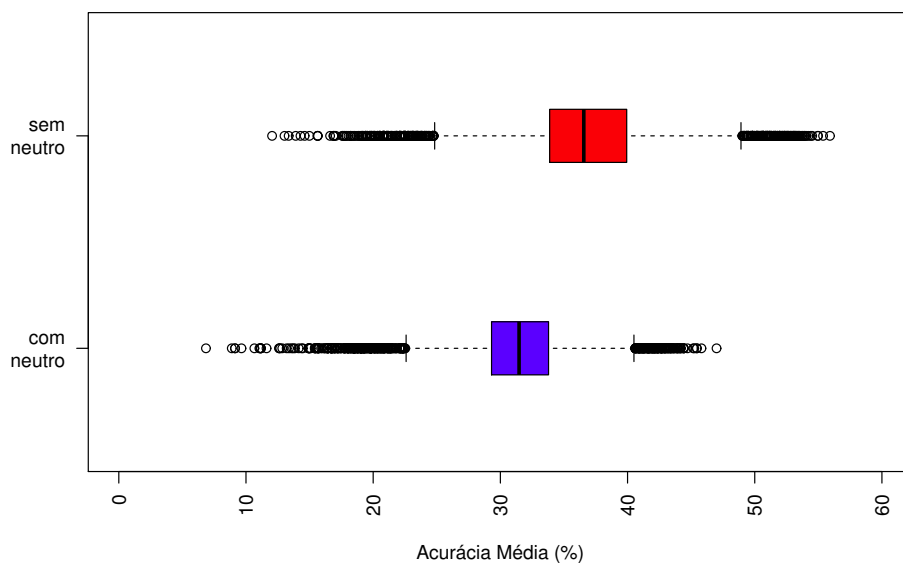


Figura 20 – Comparando influência das sentenças com presença do neutro.

Assim, é possível afirmar que as configurações com presença das sentenças neutras (uma sentença neutra expressa uma informação fatural, enquanto uma sentença subjetiva pode expressar uma opinião, emoção, crença, alegação, etc), interferem de forma negativa na classificação, principalmente devido ao acréscimo de uma classe na solução do problema. Assim é possível justificar a existência da Análise de Subjetividade que tem por meta a separação de sentenças subjetivas e objetivas, conforme destaca Bing Liu em [1].

Tabela 11 – Resultado das 10 melhores configurações segundo acurácia média em 30 repetições

Neutro	Característica	SL	A.M.	Dec	Agr	Acurácia(%)
não	Word	1	SVM	OVO	WENG	55,91
não	TF-IDF	1	SVM	OVO	WENG	55,37
não	Strength	2	SVM	OVO	WENG	54,96
não	Strength	2	SVM	OVO	WENG	54,96
não	TF-IDF	2	NB	OVO	KULE	54,51
não	TF-IDF	1	SVM	OVO	WENG	54,39
não	Strength	1	SVM	OVO	WENG	54,22
não	TF-IDF	1	SVM	OVO	WENG	54,07
não	Strength	2	SVM	OVO	WENG	53,97
não	Strength	1	SVM	OVO	WENG	53,89

Para responder à segunda questão, “É possível descrever as sentenças textuais por meio dos *Sentic Levels* e obter resultados aceitáveis?”, é necessário visualizar os melhores resultados sob a ótica da acurácia média, conforme a Tabela 11. Os melhores resultados estão distribuídos entre os SL1 e SL2 na razão de 60/40 sugerindo uma leve vantagem para o SL1 ainda que branda. Reforçando estes resultados observa-se que a acurácia média para cada SL e os três diferentes descritores visualizados na Tabela 13 demonstram variações tênues entre os SLs e os descritores.

Tabela 12 – Complexidade computacional dos Algoritmos, onde n é número de amostras e m número de características.

Algoritmo	Complexidade
RF	$O(mn \log n)$ [92]
SVM	$O(mn^2)$ [93]
DT	$O(mn)$ [94]
NB	$O(mn)$ [95]

Estes resultados corroboram na afirmação de que tanto SL como os descritores (*Strength*, *TF-IDF* e *Word*) não influenciam na classificação. Contudo, mesmo admitindo o fato da inexistência de prevalência de uma combinação específica de características sobre outra qualquer, a melhor escolha seria o uso do SL2. A escolha do SL2 deve-se ao fato de conter 16 atributos (emoções de ordem secundária) contra os 24 do SL1, considerando ainda a complexidade dos algoritmos, visualizados na Tabela 12, onde uma menor quantidade de atributos proporciona menor esforço computacional independentemente do algoritmo.

Tabela 13 – Resultado SL x Descritores

	Strength	TF-IDF	Words
SL1	33,54%	33,97%	34,42%
SL2	34,05%	33,66%	34,05%

Quanto à questão: “Existe um algoritmo mais apropriado para identificação de emo-

ções?”, a Figura 22(a) e a Tabela 15 ilustram os resultados referentes a esta questão. O algoritmo SVM produz os melhores resultados quanto à classificação de emoções, considerando que o *ranking* médio de 9,44 para SVM (oriundo do teste Friedman) é melhor que o 11,04 para o RF, 14,47 para AD e 15,05 para NB. Referente a acurácia, conforme exibido na Tabela 15, para o algoritmo SVM a menor acurácia é superior em relação aos demais algoritmos, fato que se repete para as acurácias maiores e médias. Quanto à estabilidade, a Figura 21(a) mostra que tanto SVM quanto RF são os algoritmos mais estáveis, ou seja, a distância entre os primeiros e terceiros quartis dos *boxplots* são os mais estreitos, com vantagem para o algoritmo RF. Por outro lado, os quartis do algoritmo SVM posicionam-se acima dos quartis do RF. Pode-se ainda arguir, conforme a Tabela 11, que 90% dos melhores resultados, quanto à acurácia são obtidos com o algoritmo SVM como classificador básico. Assim é possível afirmar que o algoritmo SVM é a melhor escolha de algoritmo básico para classificação de emoções.

Tabela 14 – Relação da menor, média e maior acurácia por Método de Agregação.

	VOTE	WENG	KULE	WVOT	LVPC	MCON
Menor	14,91%	15,42%	6,85%	16,92%	20,71%	14,91%
Média	36,29%	37,47%	31,97%	32,48%	34,87%	36,29%
Maior	52,84%	55,91%	53,51%	46,36%	54,52%	52,84%

Tabela 15 – Relação da menor, média e maior acurácia por Algoritmo.

	SVM	RF	DT	NB
Menor	25,38%	22,63%	6,85%	14,91%
Média	37,00%	34,73%	31,91%	31,59%
Maior	55,91%	52,27%	48,68%	54,52%

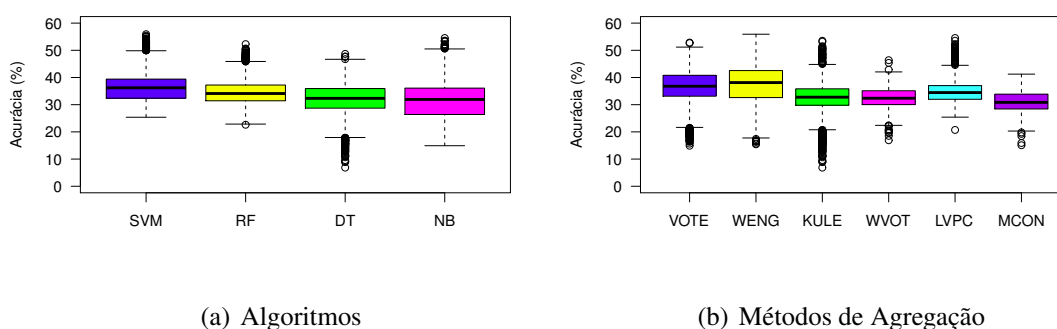


Figura 21 – Acurácia para algoritmos e métodos de agregação para classificar o sentimento da sentença

Por fim, para responder à questão “Existe método de agregação mais adequado?”, é possível observar na Figura 22(b), apresentando o *ranking* médio que os métodos WENG com 7,45 e VOTE com 8,01 são muito similares. Relativo a acurácia, na Tabela 14 é possível observar

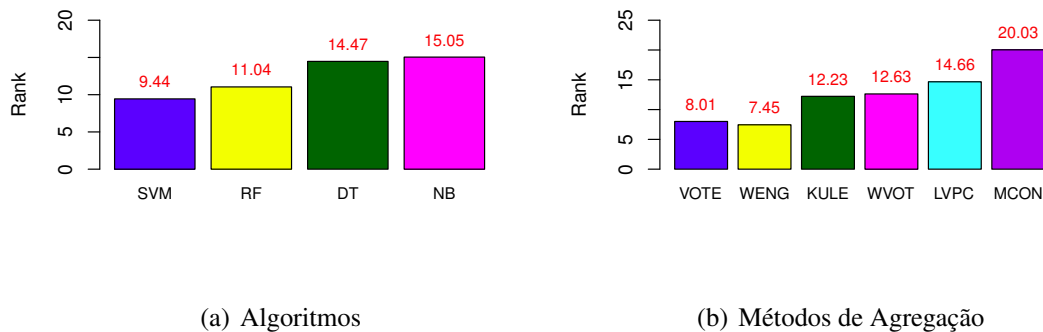


Figura 22 – *Ranking* médio para algoritmos e métodos de agregação para classificar o sentimento da sentença

que o método de agregação WENG obteve melhores resultados em acurácia média e maior em relação aos demais métodos. Tais resultados sugerem que tanto os métodos de agregação WENG como VOTE produzem os melhores resultados. Pela Figura 21(b) pode-se observar que apesar de não apresentar o melhor resultado quanto a estabilidade (distância entre o primeiro e terceiro quartil), o melhor resultado obtido pelo método WENG não se trata de um *outlier* implicando num resultado dentro da distribuição normal dos resultados. Observando ainda a Tabela 11 verifica-se que 90% dos melhores resultados são obtidos com o método de agregação WENG, somando-se essa constatação às demais torna possível afirmar que o método WENG proporciona os melhores resultados para classificação de emoções.

Baseado no teste Friedman com valor- $p < 2.2^{-16}$, verifica-se uma diferença significativa entre os classificadores multi-classes avaliados nos experimentos. Desta forma, descarta-se a hipótese de diferença insignificante para todas as configurações, assim como pode ser vista na Figura 23. Pode-se notar que a configuração classificador base SVM em conjunto com o método de agregação WENG obtém o melhor resultado situando-se no topo do *ranking*. Consequentemente podemos afirmar que conjunto SVM-WENG é a melhor configuração. Especificamente, esta configuração produziu o resultado de 55,91% quanto à acurácia, conforme também visualizado na Tabela 11.

A acurácia de 55,91% obtida neste estudo é consideravelmente inferior às acurácias de 81,45% obtidas no trabalho, referente a AS, de [10] e 90% no trabalho de [96] referente a PCMC. Entretanto ao analisar este trabalho sob o foco PCMC a diferença quanto à acurácia pode ser atribuída às significativas diferenças entre os *datasets*, uma vez que este trabalho adotou um *dataset* de notícias reais rotuladas com emoções enquanto trabalhos no estado da arte fizeram uso de *datasets* consagrados na academia. Quando tratamos do foco AS, trabalhos recentes do estado da arte têm trabalhado eminentemente com problemas binários, ou seja, tipicamente polaridade enquanto este trabalho abordou a problemática multi-classes. Considerando estes fatores a acurácia de 55,91% é um resultado significativo.

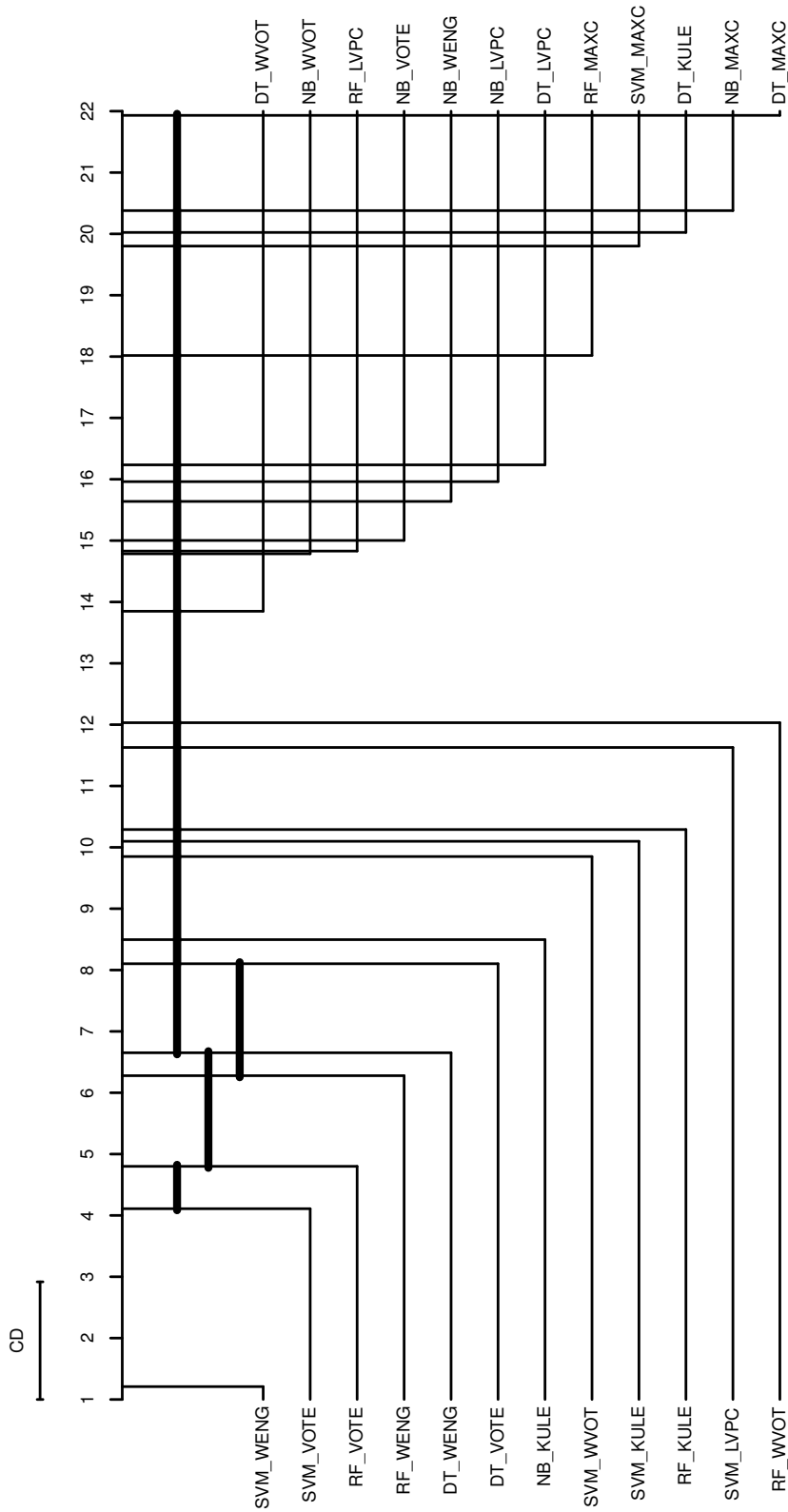


Figura 23 – Comparação entre as acurácias médias por configuração (algoritmo e agregação) de acordo o teste Nemenyi. Grupos de classificadores que não apresentam diferença significativa ($\alpha = 0,05$) estão conectados pelas linhas horizontais

5 CONCLUSÃO

Este trabalho expôs um estudo utilizando técnicas de AM para classificação multi-classes. Foram utilizados como características os SL1 e SL2 oriundos do HOE, bem como as derivações de descritores de TF-IDF, *Strength*, *Word* de uma base de dados de notícias do mundo real.

Como primeiro pilar dos experimentos temos a verificação da influência das sentenças neutras na classificação. Conforme visualizado na Tabela 11, que apresenta os dez melhores experimentos em termos de acurácia, nenhum dos resultados consta da presença de sentenças neutras. Somando-se a essa constatação, na Figura 20 é possível verificar uma diferença de 6 pontos percentuais em acurácia média de vantagem para os experimentos realizados sem a presença de sentenças neutras ante os experimentos com a presença de sentenças neutras. Deste modo podemos afirmar que a ausência das sentenças neutras influencia positivamente a classificação em termos de acurácia, podendo-se ainda atribuir a diferença negativa de performance ao acréscimo da classe neutra no problema.

Referente ao desempenho quanto ao uso das características extraídas com base nos *Sentic Levels* e derivações em função do *Strength*, TF-IDF e *Word*, ao observar a Tabela 11 nota-se, quanto ao uso dos *Sentic levels*, que o *Sentic Level* está presente em 60% dos melhores resultados ante o Sentic Level 2 e na Tabela 13. No tocante às demais características elas dividem-se em proporções muito próximas umas das outras. Assim não é possível afirmar que exista prevalência na escolha da característica, ou seja, a escolha da combinação de características não é determinante na obtenção de melhores resultados, entretanto os experimentos com SL2 são mais apropriados porque o menor número de características (16 contra 24 do SL1) é fator relevante quanto à complexidade computacional.

Dos algoritmos utilizados nos classificadores básicos (SVM, AD, NB e RF) pode-se notar que o SVM produziu os melhores resultados. Do ranqueamento produzido pelo Teste Friedman e *post hoc* Nemenyi, o algoritmo SVM obteve 9,44 ante 11,04 para RF, 14,47 para AD e 15,05 para NB, somado ao melhor ranqueamento o fato de 90% dos melhores resultados apresentados na Tabela 11 corroboram em afirmar que o algoritmo SVM influencia positivamente o resultado final em acurácia.

Da adequação do método de agregação empregado na classificação segundo a Figura 22(b), constata-se o ranqueamento 7,45 para WENG contra 8,01 para VOTE que a princípio impossibilita determinar qualquer prevalência, mas ao se observar na Tabela 11 que 90% dos melhores resultados são obtidos com o método de agregação WENG é possível aferir que o mesmo produz contribui positivamente em maior intensidade em relação aos demais.

Diante das constatações da contribuição para a classificação quanto aos métodos de

agregação evidenciou-se que os métodos WENG e VOTE se destacaram diante dos demais, podendo afirmar que são os melhores métodos para este experimento. Os experimentos revelaram ainda que a melhor combinação de configuração é o conjunto algoritmo SVM e método de agregação WENG com uma acurácia de 55,91%.

A principal contribuição deste trabalho é a proposição de um novo conjunto de descritores baseados em emoções adequado para resolução de problemas multi-classes quando o objetivo é classificação de emoções. Adicionalmente, as estratégias de binarização se mostraram eficientes na solução de problemas multi-classes. Quanto a trabalhos futuros, este estudo abre um leque de perspectivas que consistem em explorar várias direções. Podemos citar dentre as possíveis uma abordagem Multi-rótulos e modelos de pre-agregação.

REFERÊNCIAS

- [1] LIU, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.
- [2] DARWIN, C. The expression of the emotions in man and animals. *The American Journal of the Medical Sciences*, LWW, v. 232, n. 4, p. 477, 1956.
- [3] FAYAZI, A. et al. Uncovering crowdsourced manipulation of online reviews. In: ACM. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.], 2015. p. 233–242.
- [4] LI, X. et al. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, Elsevier, v. 69, p. 14–23, 2014.
- [5] RILL, S. et al. Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. *Knowledge-Based Systems*, Elsevier, v. 69, p. 24–33, 2014.
- [6] CAMBRIA, E. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, IEEE, v. 31, n. 2, p. 102–107, 2016.
- [7] RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, Elsevier, v. 89, p. 14–46, 2015.
- [8] WEICHSELBRAUN, A.; GINDL, S.; SCHARL, A. Enriching semantic knowledge bases for opinion mining in big data applications. *Knowledge-based systems*, Elsevier, v. 69, p. 78–85, 2014.
- [9] EKMAN, P. An argument for basic emotions. *Cognition & emotion*, Taylor & Francis, v. 6, n. 3-4, p. 169–200, 1992.
- [10] GALAR, M. et al. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, Elsevier, v. 44, n. 8, p. 1761–1776, 2011.
- [11] FERNÁNDEZ-DELGADO, M. et al. Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, JMLR. org, v. 15, n. 1, p. 3133–3181, 2014.
- [12] SESMERO, M. P. et al. An ensemble approach of dual base learners for multi-class classification problems. *Information Fusion*, Elsevier, v. 24, p. 122–136, 2015.
- [13] CAMBRIA, E.; HUSSAIN, A. *Sentic computing: Techniques, tools, and applications*. [S.l.]: Springer Science & Business Media, 2012. v. 2.
- [14] PLUTCHIK, R. *Emotion: A psychoevolutionary synthesis*. [S.l.]: Harpercollins College Division, 1980.
- [15] CAMBRIA, E.; LIVINGSTONE, A.; HUSSAIN, A. The hourglass of emotions. In: *Cognitive behavioural systems*. [S.l.]: Springer, 2012. p. 144–157.

- [16] JOACHIMS, T. Text categorization with support vector machines: Learning with many relevant features. In: SPRINGER. *European conference on machine learning*. [S.l.], 1998. p. 137–142.
- [17] YE, Q.; ZHANG, Z.; LAW, R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, Elsevier, v. 36, n. 3, p. 6527–6535, 2009.
- [18] CHEN, C. et al. Visual analysis of conflicting opinions. In: IEEE. *2006 IEEE Symposium On Visual Analytics Science And Technology*. [S.l.], 2006. p. 59–66.
- [19] SHIMADA, K.; ENDO, T. Seeing several stars: A rating inference task for a document containing several evaluation criteria. In: SPRINGER. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. [S.l.], 2008. p. 1006–1014.
- [20] TAN, S.; ZHANG, J. An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*, Elsevier, v. 34, n. 4, p. 2622–2629, 2008.
- [21] WILSON, T.; WIEBE, J.; HOFFMANN, P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, MIT Press, v. 35, n. 3, p. 399–433, 2009.
- [22] HASTIE, T.; TIBSHIRANI, R. et al. Classification by pairwise coupling. *The annals of statistics*, Institute of Mathematical Statistics, v. 26, n. 2, p. 451–471, 1998.
- [23] SEJNOWSKI, T. J.; ROSENBERG, C. R. Parallel networks that learn to pronounce english text. *Complex systems*, v. 1, n. 1, p. 145–168, 1987.
- [24] CERRI, R. *Técnicas de classificação hierárquica multirrótulo*. Tese (Doutorado) — Universidade de São Paulo, 2010.
- [25] JAMES, W. Ii.—what is an emotion? *Mind*, Mind Assoc, n. 34, p. 188–205, 1884.
- [26] DARWIN, C.; EKMAN, P.; PRODGER, P. *The expression of the emotions in man and animals*. [S.l.]: Oxford University Press, USA, 1998.
- [27] TURKLE, S. *The second self: Computers and the human spirit*. [S.l.]: Mit Press, 2005.
- [28] SCHERER, K. R. Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion*, Taylor & Francis, v. 7, n. 3-4, p. 325–355, 1993.
- [29] PICARD, R. W.; PICARD, R. *Affective computing*. [S.l.]: MIT press Cambridge, 1997. v. 252.
- [30] MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.
- [31] MINSKY, M. *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. [S.l.]: Simon and Schuster, 2007.
- [32] DALGLEISH, T. The emotional brain. *Nature Reviews Neuroscience*, Nature Publishing Group, v. 5, n. 7, p. 583–589, 2004.

- [33] BROCA, P. *Anatomie comparée des circonvolutions cérébrales: le grand lobe limbique et la scissure limbique dans la série des mammifères*. [S.l.: s.n.], 1878.
- [34] PAPEZ, J. W. A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, American Medical Association, v. 38, n. 4, p. 725–743, 1937.
- [35] MACLEAN, P. D. Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroencephalography and clinical neurophysiology*, Elsevier, v. 4, n. 4, p. 407–418, 1952.
- [36] LEDOUX, J. E. *Synaptic self: How our brains become who we are*. [S.l.]: Penguin, 2003.
- [37] AVERILL, J. R. A constructivist view of emotion. *Emotion: Theory, research, and experience*, v. 1, p. 305–339, 1980.
- [38] OSGOOD, C. E. The nature and measurement of meaning. *Psychological bulletin*, American Psychological Association, v. 49, n. 3, p. 197, 1952.
- [39] RUSSELL, J. A. Affective space is bipolar. *Journal of personality and social psychology*, American Psychological Association, v. 37, n. 3, p. 345, 1979.
- [40] WHISSELL, C. The dictionary of affect in language. *Emotion: Theory, research, and experience*, New York, NY: Academic Press, v. 4, n. 113-131, p. 94, 1989.
- [41] FRIJDA, N. H. The laws of emotion. *American psychologist*, American Psychological Association, v. 43, n. 5, p. 349, 1988.
- [42] SMITH, C. A.; LAZARUS, R. S. *Emotion and adaptation*. Guilford Press, 1990.
- [43] CANNON, W. B. *Bodily changes in pain, hunger, fear, and rage: An account of recent researches into the function of emotional excitement*. [S.l.]: D. Appleton and company, 1915.
- [44] BARRETT, L. F. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, Sage Publications, v. 10, n. 1, p. 20–46, 2006.
- [45] DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.
- [46] FÜRNKRANZ, J. Round robin classification. *Journal of Machine Learning Research*, v. 2, n. Mar, p. 721–747, 2002.
- [47] CLARK, P.; BOSWELL, R. Rule induction with cn2: Some recent improvements. In: SPRINGER. *European Working Session on Learning*. [S.l.], 1991. p. 151–163.
- [48] KNERR, S.; PERSONNAZ, L.; DREYFUS, G. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: *Neurocomputing*. [S.l.]: Springer, 1990. p. 41–50.
- [49] HÜLLERMEIER, E.; BRINKER, K. Learning valued preference structures for solving classification problems. *Fuzzy Sets and Systems*, Elsevier, v. 159, n. 18, p. 2337–2352, 2008.

- [50] WU, T.-F.; LIN, C.-J.; WENG, R. C. Probability estimates for multi-class classification by pairwise coupling. *The Journal of Machine Learning Research*, JMLR. org, v. 5, p. 975–1005, 2004.
- [51] VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer Science & Business Media, 2013.
- [52] SCHÖLKOPF, B. et al. Support vector method for novelty detection. In: CITeseer. *NIPS*. [S.l.], 1999. v. 12, p. 582–588.
- [53] MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of machine learning*. [S.l.]: MIT press, 2012.
- [54] SAFAVIAN, S. R.; LANDGREBE, D. A survey of decision tree classifier methodology. 1990.
- [55] QUINLAN, J. R. *C4. 5: programs for machine learning*. [S.l.]: Elsevier, 2014.
- [56] TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introdução ao datamining: mineração de dados*. [S.l.]: Ciencia Moderna, 2009.
- [57] JIANFANG LIU ZONGTIAN (AUTH.), B. L. M. M. J. C. e. S. *Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings*. 1. ed. [S.l.]: Springer-Verlag Berlin Heidelberg, 2012. (Lecture Notes in Computer Science 7473). ISBN 9783642340611,364234061X.
- [58] LOPES, K. M. de O. et al. Modelos baseados em data mining para classificação multitemporal de culturas no mato grosso utilizando dados de ndvi/modis. Campinas, SP, 2013.
- [59] HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.
- [60] JIANG, L. et al. Survey of improving naive bayes for classification. In: SPRINGER. *International Conference on Advanced Data Mining and Applications*. [S.l.], 2007. p. 134–145.
- [61] THEODORIDIS, S. et al. *Introduction to pattern recognition: a matlab approach*. [S.l.]: Academic Press, 2010.
- [62] SOKOLOVA, M.; JAPKOWICZ, N.; SZPAKOWICZ, S. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In: *AI 2006: Advances in Artificial Intelligence*. [S.l.]: Springer, 2006. p. 1015–1021.
- [63] ALPAYDIN, E. *Introduction to machine learning*. [S.l.]: MIT press, 2014.
- [64] RABELO, J. C.; PRUDÊNCIO, R. B.; BARROS, F. A. Using link structure to infer opinions in social networks. In: IEEE. *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2012. p. 681–685.
- [65] CHEN, H.; ZIMBRA, D. Ai and opinion mining. *IEEE Intelligent Systems*, IEEE, v. 25, n. 3, p. 74–80, 2010.
- [66] LI, Y.-M.; SHIU, Y.-L. A diffusion mechanism for social advertising over microblogs. *Decision Support Systems*, Elsevier, v. 54, n. 1, p. 9–22, 2012.

- [67] RUI, H.; LIU, Y.; WHINSTON, A. Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, Elsevier, v. 55, n. 4, p. 863–870, 2013.
- [68] WAWRE, S. V.; DESHMUKH, S. N. Sentiment classification using machine learning techniques.
- [69] MCDONALD, R. et al. Structured models for fine-to-coarse sentiment analysis. In: CITESEER. *Annual Meeting-Association For Computational Linguistics*. [S.l.], 2007. v. 45, n. 1, p. 432.
- [70] RABINER, L. R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, IEEE, v. 77, n. 2, p. 257–286, 1989.
- [71] CRAMMER, K.; SINGER, Y. Ultraconservative online algorithms for multiclass problems. *Journal of Machine Learning Research*, v. 3, n. Jan, p. 951–991, 2003.
- [72] MCDONALD, R.; CRAMMER, K.; PEREIRA, F. Online large-margin training of dependency parsers. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 43rd annual meeting on association for computational linguistics*. [S.l.], 2005. p. 91–98.
- [73] DANG, Y.; ZHANG, Y.; CHEN, H. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, IEEE, v. 25, n. 4, p. 46–53, 2010.
- [74] BLITZER, J. et al. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: *ACL*. [S.l.: s.n.], 2007. v. 7, p. 440–447.
- [75] SALEH, M. R. et al. Experiments with svm to classify opinions in different domains. *Expert Systems with Applications*, Elsevier, v. 38, n. 12, p. 14799–14804, 2011.
- [76] PANG, B.; LEE, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. [S.l.], 2004. p. 271.
- [77] TABOADA, M.; GRIEVE, J. Analyzing appraisal automatically. In: *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS# 04# 07)*, Stanford University, CA, pp. 158q161. AAAI Press. [S.l.: s.n.], 2004.
- [78] ZHANG, Z. et al. Sentiment classification of internet restaurant reviews written in cantonese. *Expert Systems with Applications*, Elsevier, v. 38, n. 6, p. 7674–7682, 2011.
- [79] WANG, G. et al. Sentiment classification: The contribution of ensemble learning. *Decision support systems*, Elsevier, v. 57, p. 77–93, 2014.
- [80] MORAES, R.; VALIATI, J. F.; NETO, W. P. G. Document-level sentiment classification: An empirical comparison between svm and ann. *Expert Systems with Applications*, Elsevier, v. 40, n. 2, p. 621–633, 2013.
- [81] FEI, B.; LIU, J. Binary tree of svm: a new fast multiclass training and classification algorithm. *IEEE transactions on neural networks*, IEEE, v. 17, n. 3, p. 696–704, 2006.

- [82] HONG, J.-H. et al. Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive bayes classifiers. *Pattern Recognition*, Elsevier, v. 41, n. 2, p. 662–671, 2008.
- [83] HÜHN, J. C.; HÜLLERMEIER, E. Fr3: a fuzzy rule learner for inducing reliable classifiers. *IEEE Transactions on Fuzzy Systems*, IEEE, v. 17, n. 1, p. 138–149, 2009.
- [84] FRIEDMAN, J. *Another approach to polychotomous classification*. [S.l.], 1996.
- [85] HSU, C.-W.; LIN, C.-J. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, IEEE, v. 13, n. 2, p. 415–425, 2002.
- [86] LORENA, A. C.; CARVALHO, A. C. D.; GAMA, J. M. A review on the combination of binary classifiers in multiclass problems. *Artificial Intelligence Review*, Springer, v. 30, n. 1-4, p. 19–37, 2008.
- [87] FÜRNKRANZ, J. Round robin ensembles. *Intelligent Data Analysis*, IOS Press, v. 7, n. 5, p. 385–403, 2003.
- [88] JAPKOWICZ, N.; SHAH, M. *Evaluating learning algorithms: a classification perspective*. [S.l.]: Cambridge University Press, 2011.
- [89] DOSCIATTI, M.; FERREIRA, L.; PARAISO, E. C. Anotando um corpus de notícias para a análise de sentimentos: um relato de experiência. In: *STIL*. [S.l.: s.n.], 2015.
- [90] FERREIRA, L. et al. Using a genetic algorithm approach to study the impact of imbalanced corpora in sentiment analysis. In: *The Twenty-Eighth International Flairs Conference*. [S.l.: s.n.], 2015.
- [91] CAMBRIA, E.; OLSHER, D.; RAJAGOPAL, D. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: *Twenty-eighth AAAI conference on artificial intelligence*. [S.l.: s.n.], 2014.
- [92] BIAU, G. Analysis of a random forests model. *Journal of Machine Learning Research*, v. 13, n. Apr, p. 1063–1095, 2012.
- [93] YANG, J. et al. Linear spatial pyramid matching using sparse coding for image classification. In: IEEE. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. [S.l.], 2009. p. 1794–1801.
- [94] SU, J.; ZHANG, H. A fast decision tree learning algorithm. In: *AAAI*. [S.l.: s.n.], 2006. v. 6, p. 500–505.
- [95] MOORE, A. W.; ZUEV, D. Internet traffic classification using bayesian analysis techniques. In: ACM. *ACM SIGMETRICS Performance Evaluation Review*. [S.l.], 2005. v. 33, n. 1, p. 50–60.
- [96] PEÑALVER-MARTINEZA FRANCISCO GARCIA-SANCHEZA, R. V.-G. M. R.-G. V. M. A. F. J. L. S.-C. I. Feature-based opinion mining through ontologies. *Expert Systems with Applications*, Elsevier, v. 41, p. 5995—6008, 2014.

TRABALHOS PUBLICADOS PELO AUTOR

1. Almeida, A.M.G.; Paraiso, E. C.; Barbon, S., **Multi-class Emotions classification by Sentic Levels as features in Sentiment Analysis**, Intelligent Systems (BRACIS), 2016 Brazilian Conference on, 2016, IEEE
2. Almeida, A.M.G.; Barbon, S.; Igawa, R.A.; Paraiso, E.C.; Moriguchi, S.N., **Mineração de opiniões em microblogs com abordagem CESA**, XII Escola Regional de Informática de Banco de Dados - ISSN 2177-4226, 2016
3. Almeida, A.M.G.; Barbon, S.; Igawa, R.A.; Moriguchi, S.N., **Opinion Mining: A Comparison of Hybrid Approaches**, International Conferences on Advances in Multimedia (MMedia), 2016, IEEE (Qualis CC 2012, B4)
4. Campos, G.F.C.; Igawa, R.A.; Seixas, J.L.; Almeida, A.M.G.; Guido, R.C.; Barbon, S., **Supervised Approach for Indication of Contrast Enhancement in Application of Image Segmentation**, International Conferences on Advances in Multimedia (MMedia), 2016, IEEE (Qualis CC 2012, B4)
5. Igawa, R.A.; Almeida, A.M.G.; Zarpelão B.B.; Barbon, S., **Recognition of Compromised Accounts on Twitter**, Brazilian Symposium on Information Systems (SBSI), 2015, ACM, (Best Paper Candidate - Qualis CC 2012, B4)
6. Igawa, R.A.; Almeida, A.M.G.; Zarpelão B.B.; Barbon, S., **Recognition on Online Social Network by user's writing style**, Revista Brasileira de Sistemas de Informação (iSys), 2015, SBC, (Qualis CC 2012, B5)