



Universidade Federal da Bahia
Escola Politécnica
Departamento de Engenharia Elétrica
Programa de Pós-Graduação em Engenharia Elétrica



MODELO COMPUTACIONAL PARA IDENTIFICAÇÃO DE
COMPORTAMENTOS ESTEREOTIPADOS E DETERMINAÇÃO
DO NÍVEL DE ATIVAÇÃO DE PSEUDO AUTISTAS

Autor: Marcos Yuzuru de Oliveira Camada
Orientadores: Prof. Dr. Jês de Jesus Fiais Cerqueira - UFBA
Prof. Dr. Antonio Marcus Nogueira Lima - UFCG

*Tese submetida ao Programa de Pós-Graduação em
Engenharia Elétrica, para preencher os requisitos parciais à
obtenção do Título de*
DOUTOR EM ENGENHARIA ELÉTRICA

Banca Examinadora

Dr. Jês de Jesus Fiais Cerqueira – UFBA
Dr. Antonio Marcus Nogueira Lima – UFCG
Dr. Eduardo Furtado de Simas Filho – UFBA
Dr. Augusto Cesar Pinto Loureiro da Costa – UFBA
Dr. Pablo Javier Alsina – UFRN
Dr. Ivan Nunes da Silva – USP - São Carlos

Salvador – BA

2020

MARCOS YUZURU DE OLIVEIRA CAMADA

**MODELO COMPUTACIONAL PARA
IDENTIFICAÇÃO DE COMPORTAMENTOS
ESTEREOTIPADOS E DETERMINAÇÃO DO
NÍVEL DE ATIVAÇÃO DE PSEUDO
AUTISTAS**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal da Bahia, como parte dos requisitos para obtenção do grau de Doutor em Engenharia Elétrica.

Orientadores:

Prof. Dr. Jés de Jesus Fiais Cerqueira

Prof. Dr. Antonio Marcus Nogueira Lima

Savaldor – BA

2020

C17
2

Camada, Marcos Yuzuru de Oliveira.

Modelo computacional para identificação de comportamentos estereotipados e determinação do nível de ativação de pseudo autistas/ Marcos Yuzuru de Oliveira Camada. - Salvador, 2020.

145 f. : il. color.

Orientador: Prof. Dr. Jês de Jesus Fiais Cerqueira.

Coorientador: Prof. Dr. Antonio Marcus Nogueira Lima.

Tese (doutorado) - Universidade Federal da Bahia. Escola Politécnica, 2020.

1. Inteligência computacional. 2. Comportamento estereotipado. 3. Autismo. I. Cerqueira, Jês de Jesus Fiais. II. Lima, Antonio Marcus Nogueira. III. Universidade Federal da Bahia. IV. Título.

CDD: 006.3

MARCOS YUZURU DE OLIVEIRA CAMADA

**COMPUTATIONAL MODEL FOR
IDENTIFYING STEREOTYPED BEHAVIORS
AND DETERMINING THE ACTIVATION
LEVEL OF PSEUDO AUTISTIC**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal da Bahia, como parte dos requisitos para obtenção do grau de Doutor em Engenharia Elétrica.

Orientadores:

Prof. Dr. Jés de Jesus Fiais Cerqueira

Prof. Dr. Antonio Marcus Nogueira Lima

Savaldor – BA

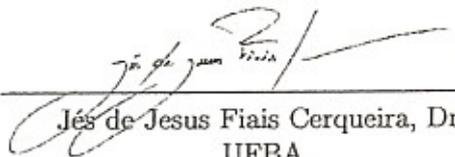
2020

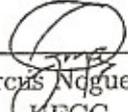
MARCOS YUZURU DE OLIVEIRA CAMADA
MODELO COMPUTACIONAL PARA IDENTIFICAÇÃO
DE COMPORTAMENTOS ESTEREOTIPADOS E
DETERMINAÇÃO DO NÍVEL DE ATIVAÇÃO DE
PSEUDO AUTISTAS

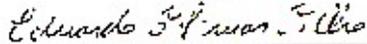
Tese apresentada à Universidade Federal da Bahia, como parte das exigências do Programa de Pós-graduação em Engenharia Elétrica, para a obtenção do título Doutor em Engenharia Elétrica.

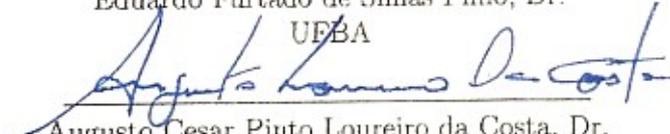
APROVADA em 05 de Junho de 2020.

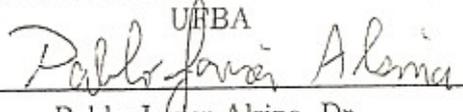
BANCA EXAMINADORA

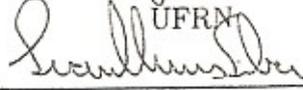

Jês de Jesus Fiais Cerqueira, Dr.
UFBA


Antonio Marcus Nogueira Lima, Dr.
UFMG


Eduardo Furtado de Simas Filho, Dr.
UFBA


Augusto Cesar Pinto Loureiro da Costa, Dr.
UFBA


Pablo Jayler Alsina, Dr.
UFRN


Ivan Nunes da Silva, Dr.
USP - São Carlos

Às minhas flores: Talita e Nina Sayuri...

AGRADECIMENTOS

Ao meu orientador professor Dr. Jés de Jesus Fiais Cerqueira pela orientação, pela paciência, pelo apoio, pela confiança, pelas contribuições e críticas durante todo o processo de desenvolvimento deste trabalho.

Ao meu coorientador professor Dr. Antonio Marcus Nogueira Lima, pela paciência, pelas contribuições e críticas fundamentais.

À minha esposa Talita Leitzke pela compreensão, apoio e pelo suporte emocional durante todos estes anos.

À minha filha Nina Sayuri por apresentar seu encanto motivador nos momentos mais difíceis da reta final deste trabalho.

Aos meus pais Ernesto Hitochi e Solange Camada pelo apoio.

À minha irmã Ilza Mitsuko Camada pelo apoio.

Ao meu grande amigo Diego pelas contribuições e sugestões para o desenvolvimento deste trabalho.

Ao IFBaiano por ter permitido a minha licença para me dedicar a este trabalho por um determinado período.

Aos professores do Programa de Pós-Graduação em Engenharia Elétrica da UFBA (PPGEE – UFBA) pelos ensinamentos.

Aos Técnicos Administrativos da secretaria do PPGEE – UFBA pela prestação no atendimento.

Resumo

Os comportamentos estereotipados são movimentos atípicos e repetitivos do corpo, os quais podem estar relacionados a transtornos mentais. Em pacientes com Transtorno do Espectro Autista (TEA), estes comportamentos podem ocasionar o aumento dos níveis de ativação. Deste modo, o desenvolvimento de sistemas capazes de reconhecer automaticamente tanto comportamentos estereotipados como inferir seu nível de ativação pode ajudar abordagens terapêuticas. Este trabalho propõe um sistema para inferir os níveis de ativação de comportamentos estereotipados típicos do autismo, em que diferentes Algoritmos de Aprendizado de Máquina (AAMs) são avaliados para o processo de classificação dos comportamentos estereotipados e para determinar os níveis de ativação relacionados. Uma métrica de desempenho denominado de Índice de Desempenho Temporal (IDT) é também proposta para avaliar o desempenho dos AAMs, considerando o tempo de classificação dos comportamentos estereotipados relacionando-o com os níveis de acurácias e precisão. Para classificação dos comportamentos estereotipados, o *Hidden Markov Model* (HMM) e o *Multilayer Percetron* (MLP) apresentaram melhores desempenho que a Support Vector Machine (SVM) e *Convolutional Neural Network* (CNN). A aplicação de técnicas de Adaptive Neuro-Fuzzy baseados no algoritmo de agrupamento *Fuzzy C-Means* permitiram determinar e diferenciar os níveis de ativação dos comportamentos estereotipados considerados neste trabalho.

Palavras Chave

algoritmos de aprendizado de máquina, análise de desempenho, estado afetivo, comportamento estereotipado, autismo, TEA

Abstract

Stereotyped Behaviors (SBs) are atypical and repetitive movements of the body, which can be related to low mental health condition. These behaviors can lead the patients with Autism Spectrum Disorder (ASD) to increase their activation levels. The development of systems able to both recognize SBs and inferring activation level automatically can aid some therapeutic approaches. In this paper, a system is proposed to infer activation levels from recognized SBs, where different Machine Learning Algorithms (MLAs) are evaluated for classifying the SBs and determining the related activation levels. A performance metric, called Temporal Performance Index (TPI), is also proposed to evaluate the performance of MLAs that consider the time for classification of SBs by relating it to accuracy and precision. For classifying the SB, the Hidden Markov Models and Multilayer Perceptron presented the best performance than Support Vector Machine and Convolutional Neural Network. The Adaptive Neuro-Fuzzy technique based on the Fuzzy C-Means Clustering algorithm allowed one to determine and differentiate the activation levels of the three stereotyped behaviors considered in the present study.

Key Words

machine learning algorithm, performance analysis, affective state, stereotyped behavior, autism, ASD

Sumário

Resumo	vii
Abstract	ix
Lista de Abreviaturas	xxi
Lista de Símbolos	xxv
1 Introdução	1
1.1 Contexto da Tese	4
1.2 Proposta da Tese	7
1.3 Objetivos	8
1.3.1 Objetivo geral	8
1.3.2 Objetivos específicos	8
1.4 Contribuições da Tese	9
1.5 Publicações e Participação em Eventos	9
1.6 Organização do Texto	10
2 Abordagens de Reconhecimento de Comportamentos Estereotipa-	
dos e Níveis de Ativação	13
2.1 Reconhecimento Automático de Comportamentos Estereotipados	13

2.2	Reconhecimento Automático de Níveis de Ativação	18
2.3	Considerações Finais	21
3	Comportamentos Estereotipados e Autismo	23
3.1	Transtorno do Espectro Autista	24
3.2	TEA e Comportamentos Estereotipados	27
3.3	Considerações Finais	29
4	Estado Afetivo e Expressões Corporais	31
4.1	Fenômenos Afetivos	32
4.2	Modelos Emocionais	32
4.3	Reconhecimento Afetivo e Expressão Corporal	35
4.3.1	Elementos afetivos na expressão corporal	37
4.3.2	Caracterização emocional na expressão corporal	39
4.4	Aspectos Emotivos nos Comportamentos Estereotipados Autistas . .	43
4.5	Considerações Finais	43
5	Arquitetura do Sistema e Experimentos	45
5.1	Visão Geral	46
5.2	Metodologia	51
5.2.1	Bibliotecas de software auxiliares para extração de caracterís- ticas	54
5.2.2	Configuração dos experimentos	56
5.3	Experimentos para Classificação dos Comportamentos Estereotipados	58
5.3.1	Experimentos de classificação com HMM	58
5.3.2	Experimentos de classificação com SVM	61
5.3.3	Experimentos de classificação com MLP	63

5.3.4	Experimentos de classificação com CNN	65
5.3.5	Discussão sobre os resultados experimentais dos classificadores	68
5.4	Experimentos de Inferência do Nível de Ativação	70
5.4.1	Desempenho de treino do ANFIS	70
5.4.2	Desempenho de teste do ANFIS	72
5.4.3	Discussão sobre os resultados experimentais do ANFIS	73
5.5	Considerações Finais	74
6	Conclusão	77
6.1	Trabalhos Futuros	79
A	Grupos Associados TEA de acordo com CID	95
A.1	Grupos Associados TEA de acordo com CID	95
A.2	Áreas do CIF Associados ao TEA	96
A.3	Áreas do CARS2 Associados ao TEA	97
B	Ferramentas de Inteligência Computacional	99
B.1	Modelos Ocultos de Markov	99
B.2	Máquina de Vetores de Suporte	103
B.3	Perceptron Multicamadas	106
B.4	Redes Neurais Convolutionais	107
B.5	Sistemas de Inferência Neuro-Fuzzy Adaptativo	110
B.5.1	Sistemas de Inferências Nebulosos	110
B.5.2	Estratégia de Treinamento do ANFIS	116

Lista de Figuras

1.1	Arquitetura geral do robô socialmente assistivo HiBot constituído por um conjunto de sensores e atuadores afetivos.	5
1.2	Dispositivo ARM modelo BeagleBoard-xM.	7
3.1	Comportamentos estereotipados típicos do autismo.	28
4.1	Modelos dimensionais.	33
4.2	Exemplo de discordância entre as expressões da face e do corpo para emoções de “medo” e “raiva” (Meeren <i>et al.</i> , 2005).	36
4.3	Determinação da extensão espacial formado pela área do triângulo das (Meijer, 1989).	41
4.4	Eixo referencial da cabeça.	43
5.1	Arquitetura geral do Módulo de Reconhecimento de Expressões Corporais (MREC).	46
5.2	Recursos presentes no Kinect [®] : emissor de infravermelho, receptor de profundidade infravermelho, sensor RGB, conjunto de microfones e um motor de inclinação vertical.	47
5.3	Articulações consideradas pelas bibliotecas OpenNI/NiTE.	48
5.4	Operações para simplificar a entrada do classificador. (a) sistema de eixos referencial; (b) deslocamento da origem dos eixos para a articulação do tronco; (c) normalização das articulações.	50

5.5	Definição dos grupos QoMs. QoM Topo [Azul] : cabeça (J_1), pescoço (J_2), tronco (J_9); QoM Membros Superiores [Verde]: ombros (J_3, J_4), cotovelos (J_5, J_6); QoM Membros Inferiores [Amarelo]: quadril (J_{10}, J_{11}), e joelhos (J_{12}, J_{13}).	51
5.6	Comportamentos estereotipados típicos obtidos pela câmera Kinect [®] nos seus respectivos quadros de imagem e profundidade: BR (a) e (b); HF (c) e (d) e TS (e) e (f).	52
5.7	Arquitetura das bibliotecas de software OpenNI (PrimeSense, 2011).	55
5.8	Definição das regiões das articulações dos segmentos dos membros do corpo (Shotton <i>et al.</i> , 2011).	56
5.9	Erro quadrático médio do algoritmo de treino do ANFIS em relação ao número de épocas.	71
5.10	Teste do modelo FCM ANFIS para os Participantes 3 e 4.	75
B.1	Caracterização da arquitetura geral de um Modelo Oculto de Markov.	100
B.2	Uma separação do hiperplano para dados de duas dimensões, onde os quadrados e circunferências representam pontos de dados nas classes -1 e +1, respectivamente. Os elementos nas linhas tracejadas pertencem ao vetor de suporte.	104
B.3	Modelo da arquitetura do MLP constituído por neurônios organizados em camadas. Há três tipos de camadas: entrada, ocultas e de saída.	106
B.4	Modelo geral da arquitetura da CNN.	109
B.5	Tipos de Sistemas Nebulosos de acordo com Wang (1997): (A) Puro, (B) TSK e (C) Mamdani.	112

Lista de Tabelas

1.1	Comparativo das principais características dos dispositivo de implantação e de experimentos.	6
2.1	Abordagens de reconhecimento automático de comportamentos estereotipados autistas.	15
2.2	Sistemas de reconhecimento de estado afetivo automático para autista.	18
5.1	Valores estatísticos das coordenadas das articulações no eixo referencial $x - y - z$ do Kinect [®]	49
5.2	Resultados experimentais do HMM com matriz de covariância diagonal variando o número de estados (n_s), baseando nas amostras de treino e teste do mesmo participante (Part).	59
5.3	Resultados experimentais do HMM com matriz de covariância esférica variando o número de estados (n_s), baseando nas amostras de treino e teste do mesmo participante (Part).	59
5.4	Resultados experimentais do HMM com matriz de covariância diagonal variando o número de estados (n_s), baseando nas amostras de treino e teste de diferentes participantes (Part).	60
5.5	Resultados experimentais do HMM com matriz de covariância esférica variando o número de estados (n_s), baseando nas amostras de treino e teste de diferentes participantes (Part).	60

5.6	Resultados experimentais da SVM com <i>kernel</i> radial variando o valor de γ , baseando nas amostras de treino e teste do mesmo participante (Part).	62
5.7	Resultados experimentais da SVM com <i>kernel</i> polinomial variando o valor de γ , baseando nas amostras de treino e teste do mesmo participante (Part).	62
5.8	Resultados experimentais da SVM com <i>kernel</i> radial variando o valor de γ , baseando nas amostras de treino e teste de diferentes participantes (Part).	63
5.9	Resultados experimentais da SVM com <i>kernel</i> polinomial variando o valor de γ , baseando nas amostras de treino e teste de diferentes participantes (Part).	63
5.10	Resultados experimentais do MLP variando a quantidade de neurônios ($n_h^{[1]}$) baseando nas amostras de treino e teste do mesmo participante (Part).	64
5.11	Resultados experimentais do MLP a quantidade de neurônios ($n_h^{[1]}$), baseando nas amostras de treino e teste de diferentes participantes (Part).	64
5.12	Resultados experimentais da CNN variando o número de filtros (n_k) baseando nas amostras de treino e teste do mesmo participante (Part).	66
5.13	Resultados experimentais da CNN variando o número de camadas (n_L) baseando nas amostras de treino e teste do mesmo participante (Part).	66
5.14	Resultados experimentais da CNN variando o número de filtros (n_k), baseando nas amostras de treino e teste de diferentes participantes (Part).	67
5.15	Resultados experimentais da CNN variando o número de camadas (n_L), baseando nas amostras de treino e teste de diferentes participantes (Part).	67

5.16	Sumário dos melhores resultados experimentais dos AAM baseados nas amostras dos mesmos participantes para treino e teste.	68
5.17	Sumário dos melhores resultados experimentais dos AAM baseados nas amostras dos participantes de treino diferente de teste.	69

Lista de Abreviaturas

AAM	Algoritmos de Aprendizado de Máquina.
ANN	<i>Artificial Neural Network.</i>
ARM	<i>Advanced RISC Machine.</i>
ANFIS	<i>Adaptive-Neural-Network-Based Fuzzy Inference System.</i>
APA	<i>American Psychiatric Association.</i>
API	<i>Application Programming Interface.</i>
AVM	Aprendizado por Votação Majoritária.
BR	<i>Body Rocking.</i>
CARS2	<i>Childhood Autism Rating Scale - 2 edition.</i>
CEP	Comitê de Ética em Pesquisa.
CID-10	Classificação Internacional de Doenças - 10ª revisão.
CIF	Classificação Internacional de Funcionalidades.
CNA	Componente de Nível de Ativação.
CNN	<i>Convolutional Neural Network.</i>
CRCE	Componente de Reconhecimento de Comportamento Estereotipado.
DSM-5	<i>Diagnostic and Statistical Manual of Mental Disorders - 5 edition.</i>
DT	<i>Decision Tree.</i>
DTW	<i>Dynamic Time Warping.</i>
ECG	Eletrocardiograma.
EEG	Eletroencefalograma.
EMG	Eletromiograma.
FCM	<i>Fuzzy C-Means Clustering Method.</i>

FDP	Função de Densidade de Probabilidade.
FK	Filtro de Kalman.
GP	<i>Grid Partition.</i>
HB	<i>Head Banging.</i>
HF	<i>Hand Flapping.</i>
HMM	<i>Hidden Markov Models.</i>
HR	<i>Hand Rotation.</i>
IDT	Índice de Desempenho Temporal.
IHR	Interação Humano-Robô.
KNN	<i>K-nearest Neighbours.</i>
LDA	<i>Linear Discriminant Analysis.</i>
MEF	Módulo de Expressão Facial.
MLP	<i>Multilayer Perceptron.</i>
MREC	Módulo de Reconhecimento de Expressões Corporais.
MREEG	Módulo de Reconhecimento de EEG.
MREF	Módulo de Reconhecimento de Expressões Faciais.
MRF	Módulo de Reconhecimento da Fala.
MSV	Módulo de Síntese de Voz.
OCC	<i>Ortony, Clore e Collins.</i>
QoM	<i>Quantify of Motion.</i>
RBF	<i>Radial Basis Function.</i>
RBS-R	<i>Repetitive Behavior Scale-Revised.</i>
RGB	<i>Red Green Blue.</i>
RGB-D	<i>Red Green Blue - Depth.</i>
SC	<i>Subtractive Clustering.</i>
SNC	Sistema Nervoso Central.
SUS	Sistema Único de Saúde.
SVM	<i>Support Vector Machine.</i>
TEA	Transtorno do Espectro Autista.
TS	<i>Top Spinning.</i>
TSK	Takagi-Sugeno-Kang.
WEKA	<i>Waikato Environment for Knowledge Analysis.</i>

WHO *World Health Organization.*

Lista de Símbolos

J_j	Articulação de índice j .
f	Índice sequencial de um determinado quadro de imagem.
χ_j^f	Módulo da coordenada da posição da articulação j no quadro f .
ψ_j^f	Módulo da orientação dada em quatérnio da articulação j no quadro f .
$v_l(f)$	Módulo da velocidade linear de um determinado membro do corpo l em um instante f .
x	Posição na coordenada do eixo da abcissa.
y	Posição na coordenada do eixo da ordenada.
z	Posição na coordenada do eixo da profundidade.
m_l	Massa aproximada de um membro l do corpo.
$E_{tot}(f)$	Energia cinética de um membro do corpo em um instante f .
k	Índice de suavidade do corpo.
SI_h	Índice de simetria horizontal das mãos.
SI_v	Índice de simetria vertical das mãos.
$SI_{espacial}$	Índice de simetria espacial das mãos.
S_Ω	Soma dos QoMs de um grupo articular Ω .
w_j	Velocidade angular de uma articulação j .
IC	Índice de inclinação da cabeça.
\mathcal{A}	Acurácia.
\mathcal{P}	Precisão.
\mathcal{T}	Tempo de reconhecimento do comportamento estereotipado.
IDT	Índice de desempenho temporal.

TP	Verdadeiro-Positivo.
FP	Falso-Positivo.
TN	Verdadeiro-Negativo.
FN	Falso-Negativo.
$f(a)$	Função de distribuição normal.
σ	Desvio padrão.
\varkappa	Média.
μ_i	Função pertinência i .
A	Matriz de probabilidade de transição de estado.
a_{ij}	Transição de um estado i para o estado j .
B	Matriz de probabilidade de observações.
b_j	Distribuição de probabilidade de símbolos observáveis j .
c	Conjunto de coeficientes.
π	Distribuição de estado inicial.
U	Matriz de covariância.
V	Conjunto discreto de símbolos observáveis;
\mathbf{v}_j	Símbolo observável j ;
n_S	Número de estados do HMM.
γ	Parâmetro do <i>kernel</i> do SVM.
$\mathfrak{K}(x, y)$	Função <i>kernel</i> .
$\mathcal{S}(x)$	Função sigmóide.
$n_h^{[l]}$	Quantidade de neurônios de uma camada l .
n_k	Número de filtros.
n_L	Número de camadas convolucionais.
S	Conjunto de estados de um HMM.
$S^{[i]}$	Um determinado estado i de um HMM.
\mathfrak{R}	Conjunto dos números reais.
\mathbb{U}	Universo de discurso.
c_{jm}	Coefficiente de mistura para m -ésima mistura no estado j .
Λ	HMM contínuo.
$k(x, z)$	Função <i>kernel</i> .
O_n^l	Saída do n -ésimo nodo da l -ésima camada de um ANFIS.

O	Tamanho da saída após a aplicação do filtro.
\mathcal{K}	Tamanho do filtro.
\mathcal{N}	Tamanho da entrada do filtro.
ς	Tamanho do passo (<i>stride</i>).
ρ	Espaçamento das margens internas da matriz (<i>padding</i>).
F	Conjunto nebuloso.
N	Nome da variável linguística.
\mathcal{V}	Valores possíveis para a variável linguística.
R	Regra semântica.

Capítulo 1

Introdução

Os comportamentos estereotipados podem ser observados nos seres humanos e outros em animais tais como, roedores (Hudson *et al.*, 2020), equinos (Kirsty *et al.*, 2015), primatas não humanos (Péter *et al.*, 2017; Moore *et al.*, 2015), entre outros (Polanco *et al.*, 2018). Este comportamentos são caracterizados por obedecer um padrão de movimentos repetitivos e invariantes (Mason, 1991).

De um modo geral, os comportamentos estereotipados podem estar associados às experiências de estresse, medo ou frustração em decorrência de ambientes pobres ou nascidos em cativeiros. Nos seres humanos, os comportamentos estereotipados podem também estar relacionados ao uso de drogas psicomotores estimulantes, danos cerebrais ou transtorno mental (Imeh-Nathaniel *et al.*, 2017; Mason, 1991).

O Transtorno do Espectro Autista (TEA) é um caso particular de transtorno mental que tem os comportamentos estereotipados como um dos principais sintomas. O TEA é um distúrbio neurológico pertencente a categoria dos transtornos do desenvolvimento pervasivo. Os seus principais sinais clínicos são percebidos ainda durante a infância a partir dos 3 anos de idade. Todavia, deficiências no desenvolvimento típico da linguagem podem dificultar a identificação dos principais sintomas (Levy *et al.*, 2009; WHO, 2020b).

O transtorno mental supracitado é considerado como um distúrbio de desenvolvimento complexo, pois apresenta diversos graus de severidade. Ele é diagnosticado através da observação comportamental e habilidades cognitivas do indivíduo

em ambiente clínico multidisciplinar. De um modo geral, o TEA pode apresentar deficiência qualitativa da interação social, comunicação e comportamentos estereotipados (ou incomuns), um repertório de atividades de interesse restrito e deficiência no contato visual (Levy *et al.*, 2009; Kyriakopoulos, 2020).

Pesquisas recentes têm investigado as motivações e mecanismos que disparam os comportamentos estereotipados nos autistas. Kuhn (1999) considera que tais comportamentos podem estar correlacionados ao mecanismo de defesa dos autistas ou de contrabalanço do nível de ativação. Este mecanismo visa reduzir o nível de ansiedade ou estresse motivado por situações que ocasionam o aumento do nível de ativação. O nível de ativação de um indivíduo está associado ao nível de alerta mental ou atividade física que influencia seu estado afetivo (Russell, 2003). Mudanças no nível de ativação pode ser percebidas nos comportamentos estereotipados (Troisi, 2002; Gaffary *et al.*, 2018). Cunningham e Schreibman (2008) revelam também que estes comportamentos podem ser um mecanismo de auto-estimulação. Em ambas considerações, os comportamentos estereotipados são disparados devido a condição de hipersensibilidade ambiental do autista. Alguns dos comportamentos estereotipados clássicos do autismo são: (i) balançar do corpo (*body rocking* – *BR*), (ii) balançar das mãos (*hand flapping* – *HF*), (iii) rodopiar (*top spinning* – *TS*) e (iv) bater a cabeça (*head banging* – *HB*).

Os comportamentos estereotipados podem causar problemas ao autista e aos seus cuidadores. Albinali *et al.* (2009) afirma que os comportamentos estereotipados podem levar o indivíduos com TEA a diminuição de repertório de atividades (Lovaas *et al.*, 1971), como também problemas de relacionamento social (Tse *et al.*, 2018) e, em alguns casos, podem ser autodestrutivos (Flowers *et al.*, 2020).

O desenvolvimento de ferramentas capazes de reconhecer automaticamente os comportamentos estereotipados e determinar seu nível de ativação pode contribuir com suporte no processo de tratamento destes indivíduos. Assim, tais ferramentas podem ser capazes de monitorar um autista durante o seu processo terapêutico ou na sua vida cotidiana visando a detecção de situações que causem estresse para este indivíduo. Pesquisas recentes têm aplicado ferramentas de inteligência computacional com o objetivo de reconhecer de maneira automática estes comportamentos. As

abordagens encontradas na literatura fazem uso de *Decision Tree* (DT) (Goodwin *et al.*, 2011), *Dynamic Time Warping* (DTW) (Gonçalves *et al.*, 2012), *Artificial Neural Network* (ANN) (Coronato *et al.*, 2014) e *Convolutional Neural Network* (CNN) (Rad *et al.*, 2015) ou modelos estatísticos (Rodrigues *et al.*, 2013). Estes trabalhos são discutidos na Seção 2.1 do Capítulo 2.

Aspectos afetivos podem ser percebidos a partir expressões corporais. Laban e Lawrence (1974) foram pioneiros em qualificar as expressões corporais a partir de coreografia de dança. Contudo, devido a interpretação artística tais expressões podem ser exageradas ou depender da qualidade interpretativa do artista. Deste modo, Savva *et al.* (2012) e Atkinson *et al.* (2007) propuseram abordagens capazes de qualificar as expressões corporais a partir de movimentos naturais do corpo.

A observação do nível de ativação de indivíduos autistas é possível através dos seus comportamentos estereotipados (Kuhn, 1999; Cunningham e Schreibman, 2008). Comumente, pesquisas não utilizam os comportamentos estereotipados como fonte de determinação de nível de ativação de autistas. A principal abordagem consiste na utilização de sinais biológicos tais como, Eletrocardiograma (ECG), Eletromiograma (EMG) e Eletroencefalograma (EEG). Estas abordagens fazem uso de sensores invasivos que podem ser inadequado em autistas hipersensíveis a estes dispositivos. Estas abordagens utilizam ferramentas de inteligência computacional para classificar o nível de ativação a partir desses sinais. As principais ferramentas presentes na literatura relacionada são *Support Vector Machine* (SVM) (Liu *et al.*, 2007; Sarabadani *et al.*, 2018), *Multilayer Perceptron* (MLP) (Othman e Wahab, 2010; Razali e Wahab, 2011), filtro de Kalman (Kushki *et al.*, 2014), *K-nearest Neighbours* (KNN) (Sarabadani *et al.*, 2018) e *Linear Discriminant Analysis* (LDA) (Sarabadani *et al.*, 2018). Estes trabalhos são discutidos na Seção 2.2 do Capítulo 2.

O processo de reconhecimento automático do estado afetivo através de comportamentos típicos autistas apresenta alguns desafios, tais como:

- Uso de sensores de baixa intrusividade. Indivíduos autistas podem apresentar uma hipersensibilidade sensorial. Deste modo, a utilização de sensores em contato ao corpo dos autistas pode causar rejeição a estes dispositivos, irrita-

ção, ocasionando interferência nos experimentos ou nos tratamentos (Black *et al.*, 2017). Portanto, o uso de sensores tipo câmera pode ser mais adequado, pois não há contato com o indivíduo;

- Efetividade em ambiente clínico real. As abordagens presentes nos trabalhos realizam a análise de desempenho em relação a acurácia e a precisão. Embora estas duas variáveis estatísticas sejam importante para avaliação de desempenho, elas podem ser insuficientes quando considerado o processo de interação entre o dispositivo e o paciente. Um tempo longo para classificação de comportamentos estereotipados pode prejudicar este processo de interação.

O problema de reconhecimento automático dos comportamentos estereotipados autistas apresenta similaridade com as abordagens relacionadas ao reconhecimento de padrões de movimentos corporais, como classificação de gestos, linguagem de sinais (Cheok *et al.*, 2019) e, especificamente para comportamento estereotipado de rodopiar, do andar (Singh *et al.*, 2018). Nestas classes de problemas como também no processo de reconhecimento dos comportamentos estereotipados, são utilizados dados cinemáticos (posição, velocidade e aceleração).

1.1 Contexto da Tese

A área de Interação Humano-Robô (IHR) tem contribuído com o auxílio de tratamentos de indivíduos autistas através de robôs socialmente assistivos. Um dispositivo robótico denominado de HiBot têm sido desenvolvido no Laboratório de Robótica do Departamento de Engenharia Elétrica da Universidade Federal da Bahia. O HiBot tem como finalidade ser uma plataforma de experimentos de IHR. A Figura 1.1 ilustra a arquitetura geral deste dispositivo com seus sensores e atuadores afetivos.

No contexto de auxílio no tratamento de indivíduos autistas, o HiBot deve se comportar como um mediador entre o terapeuta, o cuidador e o indivíduo autista. O HiBot é um dispositivo semi-autônomo, pois os terapeutas ou responsáveis podem determinar as atividades e o comportamento a depender do perfil do autista.

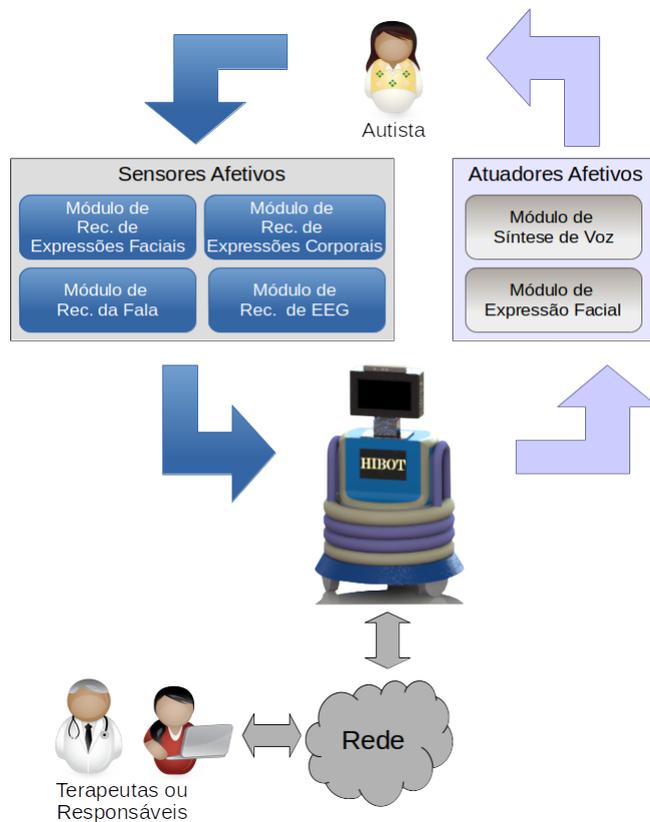


Figura 1.1: Arquitetura geral do robô socialmente assistivo HiBot constituído por um conjunto de sensores e atuadores afetivos.

Para tornar o processo de interação social efetivo com o autista, este dispositivo robótico deve interagir usando protocolos emocionais. Para isto, ele apresenta um conjunto de sensores e atuadores afetivos organizados em módulos. Embora não seja o foco deste trabalho, a arquitetura destes módulos do HiBot é brevemente descrita a seguir.

Os módulos são organizados em uma arquitetura distribuída. A conexão entre estes módulos é baseada em uma infraestrutura de rede *Ethernet* através de um *switch*. Cada módulo pode ser constituído por um ou mais dispositivos computacionais embarcados integrados com seu respectivo sensor ou atuador. Assim, cada interface de sensoriamento ou de atuação do HiBot é autônoma.

Os atuadores afetivos agem nos modais de voz e face. O Módulo de Síntese de Voz (MSV) e o Módulo de Expressão Facial (MEF) atuam nestes modais respecti-

Tabela 1.1: Comparativo das principais características dos dispositivo de implantação e de experimentos.

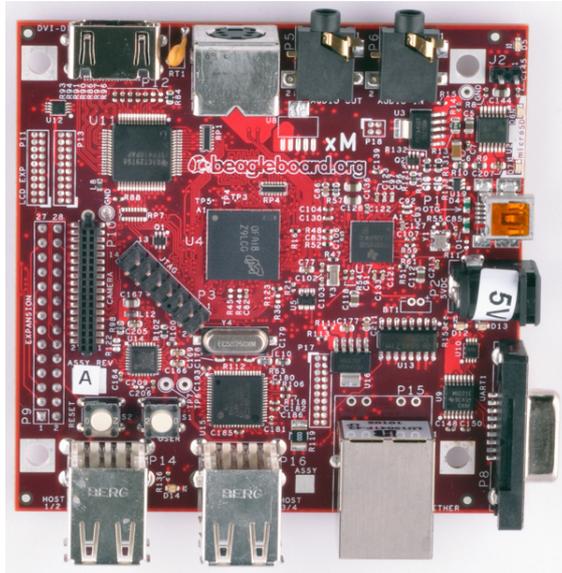
Características	Dispositivo de Implantação	Dispositivo de Experimentos
Arquitetura	ARM	CISC
Processador	Cortex-A8 1GHz	Intel i5-8400 2,8GHz – 4GHz
Memória <i>Cache</i>	32 KiB (nível 1) / 512 KiB (nível 2)*	9 MB
Número de núcleos	1	6
Suporte a ponto-flutuante	Dupla precisão	Dupla precisão
Memória principal	SDRAM 512 MB 200MHz	8GB RAM DDR4 2666Mhz
Memória secundária	MicroSD	1TB 7200rpm
Sistema Operacional	Linux/Android	Linux

* $40,96 \times 10^{-4}$ MB (nível 1) / $65,54 \times 10^{-3}$ MB (nível 2).

vamente. O MSV deve emitir sons e síntese de voz através de um conjunto de saída de áudio. O MEF deve projetar faces humanoides através de um monitor de vídeo.

Os sensores afetivos devem captar pistas afetivas de um determinado indivíduo por meio dos modais afetivos face (Módulo de Reconhecimento de Expressões Faciais – MREF), fala (Módulo de Reconhecimento da Fala – MRF), bio-sinais de EEG (Módulo de Reconhecimento de EEG – MREEG) e expressões corporais (Módulo de Reconhecimento de Expressões Corporais – MREC). Este trabalho aborda este último módulo.

O MREC é planejado a ser implementado em um agrupamento (*cluster*) de dispositivos *Advanced RISC Machine* (ARM) modelo BeagleBoard-xM[®] (ver a Figura 1.2). A arquitetura baseada em *cluster* é uma abordagem distribuída, que permite obter níveis computacionais adequados, através da subdivisão de processamento complexo em dispositivos computacionais menores com recursos computacionais mais limitados. A escolha deste dispositivo se deve aos seus recursos computacionais adequados mesmo com sua pequena dimensão (8,26 cm x 8,26 cm), o que facilita embarcamento no HiBot. A Tabela 1.1 apresenta as principais especificações deste dispositivo. Este modelo é constituído por um processador Cortex-A8 1 GHz ARM com *cache* de 32 KiB (nível 1) e 512 KiB (nível 2), e uma memória RAM de 512 MB e 200 MHz. O processamento gráfico é realizado por uma aceleradora gráfica 2D/3D integrada. A memória secundária é representada por um cartão MMC/SD. Este dispositivo é compatível com Sistemas Operacionais de alto nível como arquitetura Linux e Android[®] (Foundation, 2020).



Fonte: <http://beagleboard.org/beagleboard-xm>

Figura 1.2: Dispositivo ARM modelo BeagleBoard-xM.

Os experimentos baseados na arquitetura MREC proposta neste trabalho foram realizados em um computador pessoal com processador i5-8400 com frequência mínima de 2,8 GHz e máxima de 4 GHz e *cache* de 9 MB, com 6 núcleos e memória RAM de 8 GB e 2666 MHz. Além disto, tais experimentos foram executados sobre sistema operacional Linux (Mint 19).

1.2 Proposta da Tese

Este trabalho realiza um estudo sobre desempenho de algoritmos de inteligência computacional dirigidos à arquitetura de um componente de sistema proposto neste trabalho denominado Módulo de Reconhecimento de Expressões Corporais (MREC). Este módulo deve inferir o nível de estado afetivo de um indivíduo a partir da expressão corporal. O contexto da aplicação deste módulo está relacionado à tecnologia assistiva, visando determinar o nível de ativação de autistas através de determinados comportamentos estereotipados típicos do autismo. Este componente e os demais, que fazem parte da arquitetura robótica, são descritos na Seção 1.1.

O MREC apresenta um sensor tipo câmera para capturar os comportamentos estereotipados. Este dispositivo obtém imagens com informações de profundidade (*RGB-D – Red Green Blue – Depth*) e apresenta baixo nível de intrusividade com indivíduo observado. As informações sobre articulações de um indivíduo (orientação e posição de um determinado quadro) são extraídas através das bibliotecas de *software* OpenNI/NiTE[®].

Este estudo visa determinar o algoritmo de inteligência computacional e sua configuração mais adequada para este componente. Para este propósito, é também proposto uma variável de desempenho denominada de Índice de Desempenho Temporal (IDT) que relaciona a acurácia, precisão e tempo para o reconhecimento dos comportamentos estereotipados autistas no mundo real.

1.3 Objetivos

1.3.1 Objetivo geral

Realizar um estudo de desempenho sobre algoritmos de inteligência computacional para inferência do nível de ativação a partir de comportamentos estereotipados típicos do autismo baseados numa arquitetura de uma interface de visão para o dispositivo robótico socialmente assistivo denominado HiBot.

1.3.2 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Classificar comportamentos estereotipados autistas a partir de articulações-chaves;
- Inferir o nível de ativação de um determinado comportamento estereotipado identificado;
- Realizar um estudo de desempenho dos Algoritmos de Aprendizado de Máquina (AAM) aplicados ao reconhecimento de comportamentos estereotipados;

- Realizar um estudo sobre a capacidade de diferenciação do nível de ativação do *Adaptive-Network-Based Fuzzy Inference System* (ANFIS) para inferir o nível de ativação de um indivíduo.

1.4 Contribuições da Tese

As principais contribuições desta tese são:

1. Proporcionar uma ferramenta capaz de determinar níveis de ativação a partir de comportamentos estereotipados típicos do autismo com baixa intrusividade;
2. Determinação dos Algoritmos de Aprendizado de Máquina e sua configuração que apresentam melhor desempenho para classificação de comportamentos estereotipados do autismo para aplicação em plataformas computacionais de baixo custo;
3. Determinação de uma métrica de desempenho que considera o tempo de classificação dos comportamentos estereotipados;
4. Determinação de um ANFIS capaz de distinguir níveis de ativação.

1.5 Publicações e Participação em Eventos

A seguir são listados as participações em eventos e as publicações de artigos relacionadas a este trabalho até este momento:

- Apresentação da arquitetura geral em poster no Workshop Internacional de Engenharia Biomédica no período do dia 12 a 14 de Novembro de 2014, Vitória - Espírito Santo;
- Camada, M. Y. O., Ferreira, D. S. F., Cerqueira, J. J. F., Costa, A. L., Conceição, A. G. Body Language Recognition System using HMM and Fuzzy Inference System, International Workshop on Assistive Technology (IWAT 2015), 2015, Vitória – Espírito Santo;

- Camada, M. Y. O., Stéfano, D., Cerqueira, J. J. F., Lima, A. M. N., Conceição, A. G. S., Costa, A. C. P. L. (2016). Recognition of Affective State for Austist from Stereotyped Gestures. In Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (Vol. 1, pp. 197?204). SCITEPRESS - Science and and Technology Publications. <https://doi.org/10.5220/0005983201970204>;
- Camada, M. Y. O., Cerqueira, J. J. F., Lima, A. M. N. (2017). Stereotyped gesture recognition: An analysis between HMM and SVM. In 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA) (pp. 328 – 333). IEEE. <https://doi.org/10.1109/INISTA.2017.8001180>.
- Camada, M. Y. O., Cerqueira, J. J. F., Lima, A. M. N. (2020). A Performance Analysis of Machine Learning Algorithms for Affective State Recognition: Application for Autistic Behaviors. In Special Issue on Soft Computing for Sentiment Analysis. Applied Soft Computing – Elsevier. **(Em análise)**

1.6 Organização do Texto

Este trabalho está organizado da seguinte forma:

Capítulo 2. Apresentação dos trabalhos recentes presentes na literatura relacionada sobre reconhecimento automático de comportamentos estereotipados e estado afetivo.

Capítulo 3. Contextualização do autismo e seu relacionamento com comportamentos estereotipados típicos.

Capítulo 4. Breve discussão sobre os fenômenos afetivos e como tais fenômenos estão presentes nos comportamentos estereotipados dos autistas.

Capítulo 5. Descrição dos experimentos realizados com as ferramentas de inteligência computacional para classificar os comportamentos estereotipados típicos do autismo e a inferência do nível de ativação e discussão sobre seus resultados.

Capítulo 6. Discussão sobre as contribuições deste trabalho, como também a proposta dos trabalhos futuros.

Apêndice A. Descrição das principais classificações clínicas do TEA.

Apêndice B. Breve descrição das ferramentas de inteligência computacional aplicadas neste trabalho.

Capítulo 2

Abordagens de Reconhecimento de Comportamentos Estereotipados e Níveis de Ativação

Pesquisas recentes têm investigado o processo de reconhecimento automático de comportamentos estereotipados e do nível de ativação em indivíduos com autismo. A utilização de Algoritmos de Aprendizado de Máquina (AAM) para estas finalidades tem apresentado resultados importantes.

Neste capítulo, os trabalhos relacionados ao reconhecimento automático de comportamentos estereotipados são tratados na Seção 2.1. As propostas de ferramentas para reconhecimento do nível de ativação são apresentadas na Seção 2.2. Uma discussão sobre as abordagens descritas neste capítulo é feita na Seção 2.3.

2.1 Reconhecimento Automático de Comportamentos Estereotipados

Os comportamentos estereotipados são caracterizados por serem movimentos corporais anormais, repetitivos e sem finalidade óbvia (Mason, 1991). Estes comportamentos e sua relação com indivíduos com TEA são abordados na Seção 3.2 do

Capítulo 3. O uso de ferramentas computacionais para o reconhecimento automático deste tipo de comportamento pode contribuir para tratamentos mais eficazes destes indivíduos.

A Tabela 2.1 sumariza os principais trabalhos presentes na literatura sobre classificação automática dos comportamentos estereotipados autistas. Nesta tabela, as abordagens são descritas em relação classes de comportamentos estereotipados, tipo de sensores, algoritmo de aprendizado de máquina, características corporais consideradas, quantidade de participantes e acurácia.

As análises de desempenho comparativo entre AAM nos trabalhos relacionados são baseados principalmente na acurácia (\mathcal{A}) e na precisão (\mathcal{P}). Estas duas variáveis são definidas, respectivamente, como,

$$\mathcal{A} = \frac{TP + TN}{TP + FN + TN + FP}, \quad (2.1)$$

$$\mathcal{P} = \frac{TP}{TP + FP} \quad (2.2)$$

onde os verdadeiros-positivos (TP) e verdadeiros-negativos (TN) são resultados em que o modelo prediz corretamente a classe positiva e negativa, respectivamente. Os falsos-positivos (FP) e falsos-negativos (FN) são resultado em que o modelo prediz incorretamente a classe positiva e a classe negativa, respectivamente. Portanto, acurácia (\mathcal{A}) determina o quanto próximo um resultado de medição está do valor verdadeiro. A precisão (\mathcal{P}) refere-se o quanto próximo um resultado de medição está do outro.

Uma análise sobre qualidade dos comportamentos estereotipados *Body Rocking* (BR) e *Hand Flapping* (HF) no ambiente de sala de aula e laboratório foi realizada por Goodwin *et al.* (2011). Neste trabalho, sensores do tipo acelerômetro foram colocados em determinados pontos dos punhos e do tronco (região do tórax) de 6 crianças autistas. Para realizar a classificação dos dados destes sensores foi aplicado o algoritmo de *Decision Tree* (DT). Os resultados dos experimentos foram organizados em relação aos ambientes e não aos tipos de comportamentos. Deste modo, o desempenho médio no ambiente de laboratório e sala de aula foram 91,1% e 90,4%, respectivamente.

Tabela 2.1: Abordagens de reconhecimento automático de comportamentos estereotipados autistas.

Referência	Comportamento Estereotipado	Sensor	AAM	Características Corporais	Participante	Acurácia (%)
Goodwin <i>et al.</i> (2011)	BR, HF	acelerômetro	DT	aceleração do punho e tronco.	6	91,1 ^L , 90,4 ^C
Gonçalves <i>et al.</i> (2012)	HF	RGB-D, acelerômetro	DTW, estatístico	aceleração e posição angular do punho e tronco	4	51 76
Rodrigues <i>et al.</i> (2013)	HF	acelerômetro, batimentos cardíacos	estatístico	aceleração articular. frequência cardíaca	7	82
Rad <i>et al.</i> (2018)	BR, HF	acelerômetro	CNN	aceleração, dos punhos e tronco.	5 ^S 5 ^T 6 ^R	95 ^S 93 ^T 75 ^R
Coronato <i>et al.</i> (2014)	HF,HR	acelerômetro	ANN	posição dos punhos	2	99
Sadouk <i>et al.</i> (2018)	BR, HF	acelerômetro	CNN	posição dos membros	6	84,59 ^E , 93,45 ^F
Nossa abordagem	BR, [*] HF, ^{**} TS ^{**}	RGB-D	HMM MLP	posição angular das articulações.	6	98,3, [*] 99,3, ^{**} 98,3 ^{**}

^L Ambiente de laboratório.

^C Ambiente escolar.

^S Simulado.

^T Tempo simulado.

^R Dados reais.

^E Domínio do tempo.

^F Domínio da frequência.

* Melhores resultados com HMM.

** Melhores resultados com MLP.

Um estudo comparativo sobre o processo de reconhecimento de comportamento estereotipado (HF) usando duas abordagens diferentes para captura de dados corporais, câmera RGB-D e sensor acelerômetro, foi proposto por Gonçalves *et al.* (2012). Para a primeira abordagem, os autores aplicaram *Dynamic Time Warping* (DTW) para reconhecer o comportamento estereotipado através de dados de coordenadas das articulações do corpo. A segunda abordagem foi baseada em métodos estatísticos (desvio padrão, raiz quadra média, picos e vales) em relação à aceleração do punho e tronco. Em ambas abordagens foram utilizadas 4 crianças autistas. A acurácia da primeira e da segunda abordagem foram de 51% e 76%, respectivamente. Uma extensão deste trabalho foi realizada por Rodrigues *et al.* (2013) em que os estudos foram realizados apenas com um único sensor (acelerômetro) e o aperfeiçoamento dos parâmetros dos métodos estatísticos para classificação, alcançando 82% de acurácia com 7 autistas em um ambiente de laboratório.

Uma abordagem usando *Convolutional Neural Network* (CNN) foi proposta por (Rad *et al.*, 2018). Neste trabalho, dados de acelerômetros fixados no punho de autistas foram usados para classificação de comportamentos estereotipados (BR e HF). Três abordagens foram usadas: (i) dados simulados; (ii) dados com tempo simulado, e (iii) dados reais. Esta abordagem alcançou o melhor desempenho com 95% (dados simulados). Os desempenhos para demais abordagens foram: 93% (dados com tempo simulado) e 75% (dados reais).

Um protótipo de sistema para detecção de comportamentos estereotipados também a partir de acelerômetros no punho foi proposto por (Coronato *et al.*, 2014). Os comportamentos estereotipados considerados foram HF e o Rotação das Mãos (*Hand Rotation – HR*). A arquitetura desta proposta é formada por três componentes principais de classificação. O primeiro componente obtém o sinal dos sensores presos no punho direito do indivíduo, realiza a extração de características estatísticas. O classificador ANN gera eventos temporais relacionados ao status anormal do comportamento (comportamentos estereotipados) do indivíduo observador. Para reduzir a quantidade de erros do primeiro componente, é definido dois componentes que são baseados no paradigma consciência substitucional. O segundo componente classifica em eventos normais e anormais. O terceiro componente disponibiliza informações clínicas mais detalhadas, como duração e frequência dos movimentos dos comportamentos estereotipados. Os experimentos foram realizados com 2 pacientes autistas. Os processos de classificação *online* e *offline* alcançaram um nível de acurácia superior a 92% e 99%, respectivamente.

Um estudo sobre transferência de aprendizado em CNNs para o processo de reconhecimento de comportamentos estereotipado da abordagem no domínio da frequência para o domínio do tempo é realizado por Sadouk *et al.* (2018). Este trabalho utiliza dados de 6 participantes autistas em relação aos comportamentos estereotipados BR e HF, usando para isto acelerômetros embutidos nos braços, nas pernas e cintura. Os resultados dos desempenhos não são discriminados em relação a cada um dos comportamentos estereotipados, determinando assim, o desempenho geral para ambos os comportamentos. Deste modo, a acurácia no domínio do tempo e da frequência foram de 84,59% e 93,45% respectivamente.

Os trabalhos recentes relacionados ao processo de reconhecimento automático dos comportamentos estereotipados descritos na Tabela 2.1 apresentam limitações. Dentre elas, pode-se afirmar que as análises dos resultados de desempenho nesses estudos são restritas a precisão e acurácia. Embora estas duas variáveis estatísticas sejam importantes para o estudo sobre desempenho de algoritmos de aprendizado de máquina, elas não são suficientes para testes práticos de interação com indivíduos em ambientes clínicos reais. Neste contexto, abordagens com tempo longo para classificação pode prejudicar o processo interativo entre os participantes. Portanto, o tempo para reconhecimento de comportamentos estereotipados é um importante requisito para aplicação prática em um ambiente clínico real. As análises de tempo de reconhecimento não foram feitas em Goodwin *et al.* (2011), Gonçalves *et al.* (2012), and Rodrigues *et al.* (2013). Rad *et al.* (2018) definiu 2.5 s para os dados dos experimentos de tempo simulado. Uma comparação de desempenho entre estas abordagens é imprecisa devido a variação dos tipos de comportamento estereotipados considerados.

Ademais, esses trabalhos apresentam um repertório de classes de comportamentos estereotipados restritos principalmente ao BR e HF. Contudo, o comportamento estereotipado *Top Spinning* (TS) também é considerado como comportamento estereotipado típico e comum do autismo. Além disto, a utilização de sensores fixados no corpo do autista pode ser inapropriada para alguns destes indivíduos, devido a possibilidade de apresentarem hipersensibilidade ambiental.

Os experimentos que envolvem a utilização de participantes autistas nos trabalhos relacionados devem considerar condições ambientais. Devido a hipersensibilidade dos autistas, estas condições podem influenciar nos resultados. Goodwin *et al.* (2011) realizaram um estudo destas condições considerando ambiente de laboratório e sala de aula.

Em vista disto, este trabalho realiza um estudo de desempenho do processo de classificação de comportamentos estereotipados típicos do autismo (BR, HF e TS) usando uma câmera RGB-D (Kinect[®]). Para este propósito, foram utilizados dados de posições articulares extraídas de 6 participantes. Os melhores resultados foram atingidos com HMM para reconhecimento do BR (98,3%) e MLP para reconheci-

Tabela 2.2: Sistemas de reconhecimento de estado afetivo automático para autista.

Referência	Abordagem	Sensor	Rótulo Afetivo	Características Corporais	Participante
Liu <i>et al.</i> (2007)	Robô, RBF SMV	ECG, EMG	afeição, ansiedade, e engajamento	sinal biológico	3 ^{AU}
Othman e Wahab (2010)	MLP	EEG	valência, excitação	sinal biológico	6 ^{AU}
Kushki <i>et al.</i> (2014)	Filtro de Kalman	ECG	nível de ativação	sinal biológico	24 ^{AU}
Sarabadani <i>et al.</i> (2018)	KNN, LDA, SVM Linear, SVM Polinomial, RBF SVM	ECG	valência, excitação	sinal biológico	15 ^{AU}
Razali e Wahab (2011)	MLP	EEG	alegria, tristeza, calma	sinal biológico	6 ^{NA} 6 ^{AU}
Nossa abordagem	ANFIS	RGB-D câmera	nível de ativação	comportamento estereotipado	6 ^{NA}

^{NA} Não-autista.

^{AU} Autista.

mento do HF (99,3%) e TS (98,2%).

2.2 Reconhecimento Automático de Níveis de Ativação

O nível de ativação acompanhado da valência formam o plano cartesiano denominado de dimensão afetiva definida por Russell (2003) (ver Seção 4.1 do Capítulo 4). O nível de ativação determina o nível de excitação do estado mental de um indivíduo.

Ferramentas capazes de reconhecer automaticamente o nível de ativação de seres humanos têm sido desenvolvidas para diferentes propósitos. No cenário de tratamento de pessoas com TEA, a percepção do nível de ativação pode ser útil para verificar a efetividade ou aceitabilidade do autista perante uma abordagem terapêutica. As propostas de ferramentas para este propósito podem variar em relação ao tipo de sensores empregados, aos rótulos afetivos considerados, às características corporais utilizadas, à quantidade e tipo de participantes nos experimentos. A Tabela 2.2 sumariza a principais abordagens encontradas na literatura relacionada.

Um dispositivo robótico assistivo com objetivo auxiliar no processo de reabili-

tação de crianças autistas foi proposto por (Liu *et al.*, 2007). Esta abordagem utilizou 15 participantes autistas e um conjunto de sensores (Eletromiograma [EMG] e Eletrocardiograma [ECG]) para detectar pistas de expressões emocionais. Um componente é responsável por extração de características de sinais fisiológicos. O componente de reconhecimento de afeto utiliza tais características para classificar rótulos afetivos tais como, simpatia, ansiedade, engajamento com acurácia de 85,23%, 80,16%, e 83,63%, respectivamente. Este componente implementa um *Support Vector Machine* (SVM) com *kernel* tipo *Radial Basis Function* (RBF). Estes rótulos alimentam um componente deste dispositivo que é responsável pela interação do robô com o indivíduo autista. Um banco de dados armazena o rótulo emocional inferido e o comportamento do dispositivo robótico com um indivíduo autista.

Uma investigação sobre o comportamento do cérebro de um autista no processo de reconhecimento afetivo facial é realizado por Othman e Wahab (2010). A análise deste trabalho é baseado no modelo bidimensional (valência e excitação). Para isto, é utilizado um sensor EEG para verificar estímulos cerebrais no processo de reconhecimento afetivo da face em crianças autistas. O processo de classificação do estado emocional a partir dos sinais de EEG é realizado por um modelo de *Multilayer Perceptron* (MLP), alcançando uma acurácia de aproximadamente 90%.

Um proposta baseada em Filtro de Kalman (FK) é proposto por Kushki *et al.* (2014). Neste trabalho, um componente extrai um sinal em um intervalo de batimentos cardíacos obtidos a partir de um sensor ECG. Esta característica é utilizada por um componente que implementa uma FK modificado realimentado. Este componente deve detectar níveis de ativação (ansiedade) em tempo real. Os experimentos deste trabalho são baseado em 24 autistas e alcançou um acurácia de 95%.

Uma abordagem multimodal (ECG, respiração e condutância da pele) foi proposta por Sarabadani *et al.* (2018) para determinar o nível de excitação e valência em autistas. Para isto, utilizou-se uma combinação de classificadores integrados através de um sistema de voto majoritário. Os classificadores combinados foram *K-nearest Neighbours* (KNN), *Linear Discriminant Analysis* (LDA), e SVM de *kernel* Linear, Polinomial e RBF. Esta abordagem alcançou uma acurácia média de 80%.

Em Razali e Wahab (2011), é proposto um sistema para classificar os estados

afetivos alegria, tristeza, calma através de sinais de EEG. Os experimentos conduzidos neste trabalho foram feitos com 6 crianças autistas e outras 6 não-autistas. A arquitetura geral desta proposta é constituída de dois grande grupos de componentes. O primeiro grupo visa realizar o pré-processamento dos dados coletados pelo dispositivo EEG. Deste modo, este grupo de componente diminui o tamanho da amostragem e remove possíveis ruídos através de um filtro passa-baixa. Além disto, o sinal foi transformado do domínio do tempo para o domínio da frequência para se observar a distribuição do espectro. O segundo grupo de componentes é responsável pelo processo de classificação basendo-se no modelo do MLP. Assim, neste componente realiza a classificação dos dados em termos do plano de valência e excitação, alcançando uma acurácia média de 85%.

Igualmente como ocorre nas abordagens de classificação de comportamentos estereotipados visto na seção anterior, os trabalhos relacionados ao reconhecimento de níveis de ativação são baseados em sensores de sinais biológicos. Este tipo de sensores são capazes de extrair melhor informações afetivas de um determinado indivíduo. Contudo, estes sensores podem ser inapropriados em indivíduos autistas hipersensíveis, dificultando a incorporação destes dispositivos no corpo destas pessoas.

As abordagens sumarizadas na Tabela 2.2 são baseadas na utilização de participantes autistas. A Subseção 4.3.1 do Capítulo 4 descreve a utilização de estímulo corporal espontâneo obtêm-se movimentos corporais mais realísticos. Contudo, há necessidade de um especialista da área comportamental para definir as expressões corporais em termos de estado emocional. No contexto dos autistas, também há dificuldade de se obter um banco de amostras de comportamentos estereotipados amplo e bem caracterizado.

A abordagem proposta neste trabalho tem como objetivo inferir o nível de ativação a partir de comportamentos estereotipados. Este trabalho utiliza estímulos corporais não espontâneos através de participantes voluntários não autistas.

Para inferir o nível de ativação a partir de comportamentos estereotipados é utilizado um modelo de ANFIS para inferir este nível de ativação a partir de Quantificadores de Movimento (ver Subseção 4.3.2 do Capítulo 4) de grupos articulares. Os

resultados dos experimentos realizados permitem observar que este modelo é capaz de diferenciar níveis de ativação seguindo a notação de regras nebulosas.

2.3 Considerações Finais

O desenvolvimento de ferramentas capazes de reconhecer os comportamentos estereotipados e inferir seu nível de ativação pode ser útil para o processo terapêutico de indivíduos com TEA. Estudos clínicos têm indicado que os comportamentos estereotipados são capazes de transmitir sugestões afetivas de autistas (Mason, 1991; Kuhn, 1999; Cunningham e Schreibman, 2008).

A Seção 4.4 do Capítulo 4 aborda que tais comportamentos podem estar relacionados ao mecanismo de defesa ou de auto-estimulação devido a hipersensibilidade sensorial em relação ao ambiente do autista. Neste estado, ocorre um aumento do nível de ativação do autista. Contudo, os trabalhos na literatura relacionada tratam os dois fenômenos separadamente.

Além disto, nota-se nas Tabelas 2.1 e 2.2 que as principais abordagens fazem uso de sensores embutidos em indivíduos autistas. Devido a possibilidade destes pacientes apresentarem hipersensibilidade ao toque ou hostilidade em usar objetos estranhos, estes sensores podem ser invasivos e podem impossibilitar a sua aplicação mais ampla ou em um ambiente clínico real. A utilização de sensores menos intrusivos é um importante requisito para aplicações envolvendo indivíduos com TEA. Deste modo, um sensor câmera pode ser mais adequado para este tipo de aplicação como foi proposto por Gonçalves *et al.* (2012).

Uma outra característica presente nos trabalhos de reconhecimento de comportamentos estereotipados e determinação do estado afetivo de indivíduos autistas refere-se a utilização de autistas nos experimentos. A Subseção 4.3.1 do Capítulo 4 descreve que o uso de estímulos corporais espontâneos geram fenômenos afetivos reais. Contudo, esta abordagem necessita do acompanhamento de um especialista da área comportamental. No contexto dos autistas, existe também a dificuldade de se criar um banco de amostras e com características bem-definidas dos comportamentos estereotipados rotulados amplo que é mais prático de ser obtido em abordagens de

estímulos corporais não espontâneos através de uso de atores. Contudo, esta última abordagem apresenta a dependência da qualidade interpretativa dos participantes em executar os movimentos semelhantes aos gestos reais.

Capítulo 3

Comportamentos Estereotipados e Autismo

Os Comportamentos Estereotipados ou Comportamentos Motores Estereotípicos são padrões de movimentos corporais anormais, repetitivos e sem finalidade óbvia (Mason, 1991). Este tipo de comportamento pode ser observado em diferentes espécies criadas em cativeiros ou confinados tais como, roedores (Hudson *et al.*, 2020), equinos (Kirsty *et al.*, 2015), primatas não humanos (Péter *et al.*, 2017), humanos (Mason, 1991; Cunningham e Schreibman, 2008), dentre outros (Polanco *et al.*, 2018; Moore *et al.*, 2015).

Condições diversas podem desencadear este tipo de comportamentos tanto nos humanos como em outros animais. Nos seres humanos, os comportamentos estereotipados podem estar associados ao mal-estar de um indivíduo decorrente de condições de estresse, frustração ou medo (Mason, 1991), influência de drogas estimulantes psicomotoras (Imeh-Nathaniel *et al.*, 2017), ou transtornos mentais. Um caso de transtorno mental que pode apresentar os comportamentos estereotipados como uma das características sintomáticas é o Transtorno do Espectro Autistas (TEA).

Em vista disto, este capítulo visa contextualizar e caracterizar o autismo e sua manifestação através dos comportamentos estereotipados. Assim, a Seção 3.1 descreve o TEA e os seus principais critérios de diagnósticos. O relacionamento

do TEA e comportamentos estereotipados é abordado na Seção 3.2. Uma breve discussão sobre este capítulo está presente na Seção 3.3.

3.1 Transtorno do Espectro Autista

O autismo e seus sintomas foram primeiramente descritos pelos médicos austríacos Leo Kanner e Hans Asperger entre os anos de 1943 e 1944 (Kyriakopoulos, 2020). Atualmente, o autismo ou Transtorno do Espectro do Autismo (TEA) é definido como um distúrbio neuropsiquiátrico que pode estar associado à diversos fatores tais como, influência genética, problemas neurológicos (anomalias anatômicas ou fisiológicas do Sistema Nervoso Central – SNC) e de riscos psicossociais (Levy *et al.*, 2009; Lima *et al.*, 2017). Os fatores de riscos psicossociais podem ser disfunções na vida familiar ou conjugal, eventos de vida estressantes, exposição aos maus-tratos, dentre outros (Paula *et al.*, 2007). A natureza espectral do autismo esta relacionada a diferentes níveis de severidade que este tipo de transtorno mental pode se manifestar.

Em 2017, a *World Health Organization* (WHO) estimou que 1 a cada 160 crianças no mundo apresentava TEA (WHO, 2020a). No Brasil, pesquisas censitárias relacionadas ao TEA são escassas, o que dificulta a adoção de políticas públicas para este segmento da sociedade. A mais recente pesquisa censitária sobre autismo foi realizada pelo Instituto de Psiquiatria do Hospital das Clínicas da Universidade de São Paulo em 2007 revelando a existência de 1 milhão de casos de autismo na população de 190 milhões (Paiva Junior e Ribeiro, 2010). Além disto, segundo dados epidemiológicos internacionais, a incidência do TEA na população de crianças do sexo masculino é maior que do sexo feminino (Loomes *et al.*, 2017).

O autismo está presente desde o nascimento do indivíduo, mas os seus principais sinais e sintomas se manifestam por volta dos 3 anos de idade (Moss *et al.*, 2016). Assim, um diagnóstico precoce e um tratamento¹ singular são os principais fatores

¹De um modo geral, não há um consenso nos trabalhos científicos, de saúde e governamentais sobre a utilização das palavras “*tratamento*”, “*terapia*” e “*intervenção*” em relação às abordagens médicas no autismo. Com finalidade de facilitar a compreensão, tais palavras entendam-se como equivalentes neste trabalho.

para alcançar resultados promissores.

Atualmente, o diagnóstico do TEA é realizado unicamente por análise clínica através da observação do indivíduo e entrevistas com os pais e/ou responsáveis. Esta ação é acompanhada por uma equipe de especialistas multidisciplinares.

O diagnóstico do autismo e a determinação de sua severidade podem ser baseados em diferentes critérios. Os principais critérios associados a literatura médica são descritas na 10ª edição da Classificação Internacional das Doenças (CID-10) (WHO, 2020b), Classificação Internacional de Funcionalidades (CIF) (WHO, 2020c), a 5ª edição do Manual Estatístico e Diagnóstico de Transtorno Mentais (*Diagnostic and Statistical Manual of Mental Disorders – DSM-5*) (APA, 2013) e a 2ª edição da Escala de Avaliação do Autismo na Infância (*Childhood Autism Rating Scale – CARS2*) (Schopler *et al.*, 2010).

A CID-10 foi elaborada pela Organização Mundial da Saúde (OMS) e é atualmente adotada no Brasil pelo Sistema Único de Saúde (SUS). Este documento relaciona doenças e outros problemas de saúde (WHO, 2020b). Ela tem como objetivo classificar estudos epidemiológicos gerais e avaliar a assistência à saúde. Neste documento, o autismo é abordado no *Capítulo V – Transtornos Mentais e Comportamentais (F00–F99)* – nas seções *Transtornos do Desenvolvimento Psicológico (F80–F89)* e *Transtornos Invasivos do Desenvolvimento (F84)*. Neste grupo, os indivíduos são caracterizados por alterações qualitativas das interações sociais recíprocas e modalidades de comunicação, comportamento estereotipado, restrito e obsessivo repertório de interesse de atividades, e isolamento. A classificação sumariada do TEA de acordo com o CID-10 está presente na Seção A.1 do Apêndice A.

A avaliação com CID-10 pode ser realizada acompanhada da CIF. A CIF é um sistema de classificação também mantida pela OMS como objetivo geral dispor de uma linguagem unificada e padronizada para descrição da saúde e de seus estados relacionados. Além disto, ela descreve a funcionalidade e as incapacidades identificadas relacionadas às condições de saúde em uma pessoa (WHO, 2020c). Através da CIF é possível identificar os facilitadores e obstáculos, dentre os fatores ambientais (físico, social e de atitude) peculiares a cada indivíduo em diferentes momentos da vida. O TEA é classificado considerando a atividade e participação

individual, qualificadores de desempenho e de capacidade em 9 áreas (ver Seção A.2 do Apêndice A).

Os critérios com maior ênfase para o diagnóstico de transtornos mentais estão presentes no DSM-5 propostos pela Associação Americana de Psiquiatria (*American Psychiatric Association – APA*). De acordo com o DSM-5, os critérios para diagnóstico do TEA são (APA, 2013):

- Deficiência persistente na comunicação e interesse social. Estes indivíduos podem apresentar limitações na reciprocidade social, emocional, nos comportamentos de comunicação verbal e não verbal úteis para interação social;
- Padrões restritivos e repetitivos de comportamento, interesses ou atividades. Seus comportamentos respeitam padrões restritivos, repetidos e estereotipados. Além disso, pode-se também apresentar insistência ou inflexibilidade em coisas ou rotinas, como também apresentar comportamentos verbais e não verbais ritualísticos;
- Hiper ou hipo-reativos a estímulos sensoriais ambientais.

A determinação da severidade do autismo é comumente baseada na Escala de Avaliação do Autismo Infantil (*Childhood Autism Rating Scale – CARS2*) (Schopler *et al.*, 2010). Esta escala é representada por um formulário que categoriza 15 grupos de questões (ver Seção A.3 do Apêndice A). Cada questão pode ser atribuída uma resposta de valor numérico associados aos rótulos “dentro do intervalo normal para a idade da criança”, “ligeiramente anormal”, “moderadamente anormal”, “severamente anormal”. As pontuações das categorias são somadas e associadas a uma escala de três níveis dependendo da faixa de idade (2 a 12 anos e maiores que 13): “sintomas mínimo para ausência de TEA”, “sintomas suave para moderado de TEA” e “sintomas severos de TEA”.

A avaliação comportamental do indivíduo com TEA faz parte do processo clínico de diagnóstico e prognóstico, como também no processo terapêutico. Uma das ferramentas utilizadas para avaliação destes comportamentos é a Escala de Comportamento Repetitivo Revisada (*Repetitive Behavior Scale-Revised – RBS-R*) (Martínez-González e Piqueras, 2018).

A RBS-R é uma extensão da RBS (Hooker *et al.*, 2019) a qual categoriza o autismo em 6 sub-escalas comportamentais baseadas em experiências clínicas: (i) comportamento estereotipado (movimentos sem propósito óbvio que são repetidos de forma semelhante), (ii) comportamento auto-injuriioso (ações que causam ou tem potencial de causar vermelhidão, contusões ou outras lesões no corpo), (iii) comportamento compulsivo (comportamento que é repetido e executado de acordo com uma regra ou envolvem ações sendo feitas com gratificação emocional), (iv) comportamento ritualístico (realiza atividades cotidianas de maneira similar), (v) comportamento uniforme (resistência a mudança, insistindo que coisas permaneçam iguais), e (vi) comportamento restrito (limitada variedade de foco, interesse ou atividade).

A RBS-R é utilizada para determinar o nível de severidade do autismo a partir da análise comportamental do indivíduo. É atribuída a cada item da escala valores numéricos de intensidade entre 0 e 3 referentes aos rótulos “*comportamento não ocorre*” e “*comportamento ocorre e é um problema severo*”, respectivamente.

3.2 TEA e Comportamentos Estereotipados

No que concerne aos indivíduos com TEA, os comportamentos estereotipados tornam-se mais comuns a partir dos 2 anos de idade podendo ser considerados como um sintomas do domínio social e comunicação (Watt *et al.*, 2008). Estudos indicam diferentes fatores para o desencadeamento dos comportamentos estereotipados nos autistas.

Os comportamentos estereotipados podem estar relacionados a um mecanismo de defesa (Kuhn, 1999) ou de auto-estimulação (Cunningham e Schreibman, 2008) ocasionados pela hipersensibilidade [ambiental] destes indivíduos. Neste sentido, a determinação da abordagem terapêutica efetiva necessita ser personalizada para cada paciente e acompanhada por uma equipe multidisciplinar.

Os comportamentos estereotipados podem prejudicar a qualidade de vida e o

¹<https://www.youtube.com/watch?v=GY-Di41j5Wk>

²<https://www.youtube.com/watch?v=iDuAugAykcc>

³<https://www.youtube.com/watch?v=Gt-WMpTP7IE>

⁴<https://www.youtube.com/watch?v=Pa-pdBF4FFA>



(a) Balançar do corpo¹



(b) Balançar das mãos²



(c) Rodopiar³



(d) Bater da cabeça⁴

Figura 3.1: Comportamentos estereotipados típicos do autismo.

processo terapêutico dos indivíduos autistas. Alguns motivos para isto são:

- Os autistas podem se engajar nos comportamentos estereotipados por longo tempo, tornando-os dominantes, diminuindo o repertório de atividades de interesse (Lovaas *et al.*, 1971);
- Os comportamentos estereotipados podem ser considerados como inapropria-

dos por outras pessoas, assim os indivíduos autistas podem ser estigmatizados, prejudicando o processo de interação e integração social (Tse *et al.*, 2018);

- Determinados comportamentos estereotipados podem evoluir para comportamentos que causam lesões (Flowers *et al.*, 2020).

Indivíduos com TEA podem apresentar comportamentos estereotipados típicos. Alguns destes comportamentos estereotipados podem ser caracterizados da seguinte forma (Kuhn, 1999):

1. Balançar do corpo. Movimento do torso-superior para cima e para baixo (ver Figura 3.1(a));
2. Balançar das mãos. Movimento das mãos para cima e para baixo que pode ser acompanhado de movimentos dos dedos (ver Figura 3.1(b));
3. Rodopiar. Girar sobre o mesmo eixo andando ou correndo (ver Figura 3.1(c));
4. Bater da cabeça. Movimento semelhante ao balanço do corpo mas na presença de obstáculos como paredes ou chão (ver Figura 3.1(d)).

3.3 Considerações Finais

Estudos relacionados ao Transtorno do Espectro Autista (TEA) visam melhorar o entendimento deste distúrbio que, hodiernamente, não é ainda plenamente bem definido. As pesquisas multidisciplinares, integrando as áreas tradicionalmente clínicas com a engenharia e a computação, têm buscado aprimorar os tratamentos de indivíduos com TEA, integrar socialmente estas pessoas visando torná-los independentes e produtivas, como também melhorar o seu bem-estar.

Uma intervenção precoce e personalizada é importante para se obter resultados promissores. Deste modo, o diagnóstico deve ser realizado por uma equipe multidisciplinar. Este diagnóstico é baseado em diferentes critérios, tais como CID-10 (WHO, 2020*b*), CIF (WHO, 2020*c*) e mais especificamente o DSM-5 (APA, 2013),

este último concerne os transtornos mentais. A determinação da severidade do autismo pode ser apoiada pelo CARS2 (Schopler *et al.*, 2010) e pelo RBS-R (Hooker *et al.*, 2019). Este último é direcionado em características comportamentais para determinar a severidade do autismo.

Os comportamentos estereotipados podem estar relacionados ao mecanismo de defesa ou de auto-estimulação devido a hipersensibilidade de pessoas com TEA (Kuhn, 1999; Cunningham e Schreibman, 2008). Assim, a sua detecção e atribuição de severidade de maneira automática sob apoio de ferramentas computacionais pode contribuir no processo terapêutico dos autistas.

Capítulo 4

Estado Afetivo e Expressões Corporais

O trabalho de Darwin intitulado “*A Expressão da Emoção nos Homens e Animais*” foi pioneiro no estudo sobre as formas de expressões afetivas do ser humano, principalmente, através dos modais facial e corporal. Este trabalho tem influenciado pesquisas na área da neurociência relacionadas aos estudos sobre percepção e representação afetiva no ser humano.

As pesquisas sobre os fenômenos afetivos no ser humano avançaram desde os primeiros estudos realizados por Darwin. Este capítulo aborda os principais fenômenos afetivos e os relacionam com a expressividade corporal humano. Como prelúdio, a Seção 4.1 esclarece a definição de fenômenos afetivos com o intuito de evitar a ambiguidade dos termos comuns na literatura. Os modelos emocionais presentes na literatura relacionada são descritos na Seção 4.2. O processo de reconhecimento afetivo através da expressão corporal é abordado na Seção 4.3. A relação de elementos emocionais e os comportamentos estereotipados são esclarecidos na Seção 4.4. Por último, as considerações finais deste capítulo são discutidas na Seção 4.5.

4.1 Fenômenos Afetivos

Estudos recentes relacionados ao reconhecimento automático e geração de expressões afetivas têm utilizado de maneira inconsistente ou imprecisa termos como “afeto”, “humor” e “emoções” (Karg *et al.*, 2013). De acordo com Ortony (2009) e Karg *et al.* (2013), “afeto” refere-se a qualquer aspecto mental de um indivíduo em relação a uma resposta ou avaliação do ambiente, seja ele espontâneo e automático ou deliberado e consciente, podendo ser benéfico ou nocivo, simpático ou hostil. De um modo geral, o “afeto” apresenta dois principais fenômenos: “emoção” e “humor”. A “emoção” é um fenômeno intenso e de curta duração. Enquanto o “humor” é de longa duração e que pode variar a intensidade ao longo do tempo.

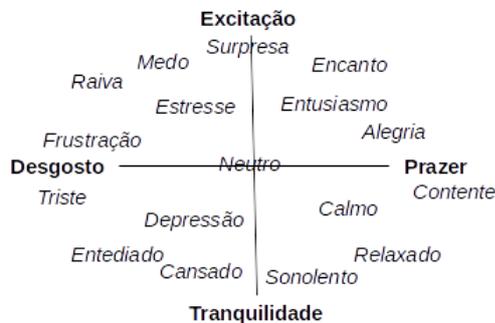
Com o intuito de evitar o uso dos termos de maneira ambígua em seus trabalhos, Karg *et al.* (2013) denominou “estado afetivo” como qualquer situação em que um fenômeno pode ser interpretado como de longa ou de curta duração. Em casos particulares nos quais estejam claro ser de curta ou de longa duração, foram utilizadas, explicitamente, as denominações “emoção” e “humor” respectivamente. Da mesma forma, neste trabalho adotar-se-ão as mesmas regras de denominação dos termos propostas por Karg *et al.* (2013).

4.2 Modelos Emocionais

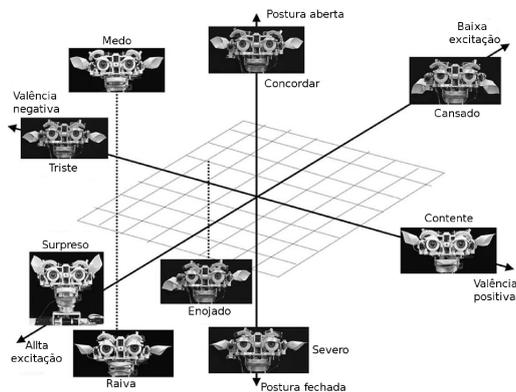
O ser humano pode reconhecer e diferenciar o estado afetivo de um outro indivíduo através da observação de diferentes canais verbais e não verbais. A partir desta diferenciação, o observador, inconscientemente ou conscientemente, é capaz de estabelecer rótulos emocionais sobre o indivíduo, como “alegria”, “tristeza”, “aborrecimento”, dentre outros a partir de pistas afetivas.

Deste modo, pesquisas relacionadas ao estudo de reconhecimento ou representação de estados afetivos devem ser capazes de determinar tais rótulos emocionais. Os principais modelos emocionais são:

- Discreto;



(a) Modelo de bi-dimensional baseado no Núcleo Afetivo proposto por Russell (2003) com eixos: nível de satisfação (prazer/desgosto) e ativação (excitação/tranquilidade).



(b) Modelo multidimensional proposto por Breazeal (2004) estende o modelo bi-dimensional adicionando a dimensão da expressão facial.

Figura 4.1: Modelos dimensionais.

- Dimensão afetiva;
- Orientado ao significado;
- Orientado aos componentes.

O modelo discreto é influenciado pelo trabalho de Darwin, que considerava que as emoções foram diferenciadas durante o curso evolutivo das espécies juntamente com o desenvolvimento de estratégias emotivas adaptadas ao ambiente. Neste modelo, cada emoção básica apresenta uma condição própria e obedece um padrão de reação comportamental, expressividade e fisiológica (Scott, 1988). Os rótulos das emoções variam nos trabalhos presentes na literatura. Por exemplo, Ekman (1992) determinou 15 (quinze) tipos de emoções que podem ser expressas pela face e corpo: “prazer”, “raiva”, “desprezo”, “contentamento”, “nojo”, “vergonha”, “excitação”, “medo”, “culpa”, “orgulho”, “alívio”, “tristeza/angústia”, “satisfação”, “prazer sensorial” e “compaixão”. Contudo, Izard (1994) observou desvantagens no modelo discreto, tais como, dificuldade ou falta de exatidão em determinar tais rótulos emocionais a partir da face, diferença na interpretação de uma expressão em diferentes

culturas ou línguas, e a dificuldade em capturar uma grande variedade de experiências emocionais distintas que possam ser intuitivamente postuladas.

No modelo dimensão afetiva, o estado afetivo de um indivíduo é determinado a partir de um espaço contínuo de emoções. As emoções podem ser definidas sobre um espaço bi-dimensional ou multidimensional. O modelo bi-dimensional foi proposto por Russell (2003) denominado por ele como “*núcleo afetivo*”. Nesta representação, os eixos da abscissa e da ordenada representam o nível de satisfação do indivíduo (prazer/desgosto) e o nível de ativação (excitação/tranquilidade) (ver Figura 4.1(a)), respectivamente. No modelo multidimensional, é acrescido um ou mais eixos relacionando-os com outros aspectos emotivos. O terceiro eixo proposto por Breazeal (2004) é relacionado com a expressão facial de um robô denominado de *Kismet* (ver Figura 4.1(b)).

O nível de ativação, eixo vertical do Núcleo Afetivo, é definido por Russell (2003) como sendo uma mobilização geral do estado mental de um indivíduo para uma ação vigorosa. Esta ação irá depender de uma avaliação das circunstâncias e recursos atuais, da criação de um objetivo e da formação de um plano para alcançá-lo. Da mesma forma, o Núcleo Afetivo influencia a qualidade e o tipo de processamento cognitivo do indivíduo. Evidências têm mostrado que o nível de ativação pode influenciar o desempenho cognitivo. Dessa forma, o desempenho ótimo cognitivo ocorre em níveis intermediários do nível de ativação, com níveis de ativação altos para simples tarefas, baixos para as mais complexas. Os níveis de ativação podem ser refletidas através das expressões corporais como por exemplo, movimentos mais energéticos ou explosivos (Wallbott, 1998; Shu *et al.*, 2018).

O modelo orientado ao significado organiza a semântica de maneira hierárquica em relação aos termos emocionais. Devido a esta característica, este modelo permite interpretar padrões emocionais comuns em diferentes culturas. Contudo, a estruturação destes rótulos emocionais pode ser inconsistente a depender de uma língua em particular (Scott, 1988). Colby *et al.* (1989) propôs um modelo orientado ao significado denominado de *OCC* (*Ortony, Clore e Collins*). Este trabalho propõe uma estrutura semântica para diferentes conceitos de emoção com base nas condições cognitivas obtidas envolvendo eventos, objetos e outros agentes.

O modelo orientado aos componentes visa esclarecer a relação entre a elicitación e caracterização das emoções e a padronização da resposta de uma maneira mais explícita. Neste modelo, uma resposta emocional começa com uma avaliação do significado pessoal de um evento ou situações antecedentes, que por sua vez dão origem a uma resposta emocional que envolve experiência subjetiva, fisiologia e comportamental (Scott, 1988).

4.3 Reconhecimento Afetivo e Expressão Corporal

O ser humano é capaz de reconhecer o estado afetivo de outros indivíduos através de informações sugestivas de modais visuais e auditivos, tais como face (Zhao *et al.*, 2003), movimentos dos olhos (O'Dwyer *et al.*, 2017), membros do corpo (Kleinsmith e Bianchi-Berthouze, 2013), variação na intonação da voz (Lassalle *et al.*, 2019). Além destes, o estado afetivo de uma pessoa também pode ser percebido através de sinais biológicos, tais como Eletroencefalograma (EEG), Electrocardiograma (ECG), Eletromiograma (EMG), respiração, batimentos cardíacos, volume sanguíneo, variação da temperatura ou condutividade da pele (Shu *et al.*, 2018).

As abordagens mais comuns envolvem as modalidades de expressão emocional da face (Zhao *et al.*, 2003) e da voz (Lassalle *et al.*, 2019). Recentemente, o interesse da comunidade científica tem aumentado em relação ao uso de linguagem corporal para percepção emocional. Conforme Kleinsmith e Bianchi-Berthouze (2013), e Ekman e Friesen (1974), existem algumas vantagens da aplicação da expressão corporal para o reconhecimento de emoções. Algumas delas são (Kleinsmith e Bianchi-Berthouze, 2013):

- As expressões corporais são capazes de transmitir emoções tão bem quanto as expressões faciais;
- Necessidade das tecnologias atuais melhorarem o entendimento de diferentes modalidades com o intuito de promover uma interação mais ubíqua;

- Aumento da relevância das expressões corporais em diversas áreas da sociedade como entretenimento, educação, saúde e segurança;
- As expressões corporais permitem uma melhor diferenciação entre os estados emotivos de “medo” e “raiva” ou “medo” e “alegria”, como também a possibilidade do reconhecimento dos estados afetivos a uma distância maior do observador.

Além dessas considerações, Meeren *et al.* (2005) e den Stock *et al.* (2007) também observaram situações em que as expressões corporais são conflitantes com expressões faciais, o observador tende, inconscientemente, a priorizar as expressões corporais no julgamento do estado afetivo da pessoa observada. A Figura 4.2 ilustra esta situação.

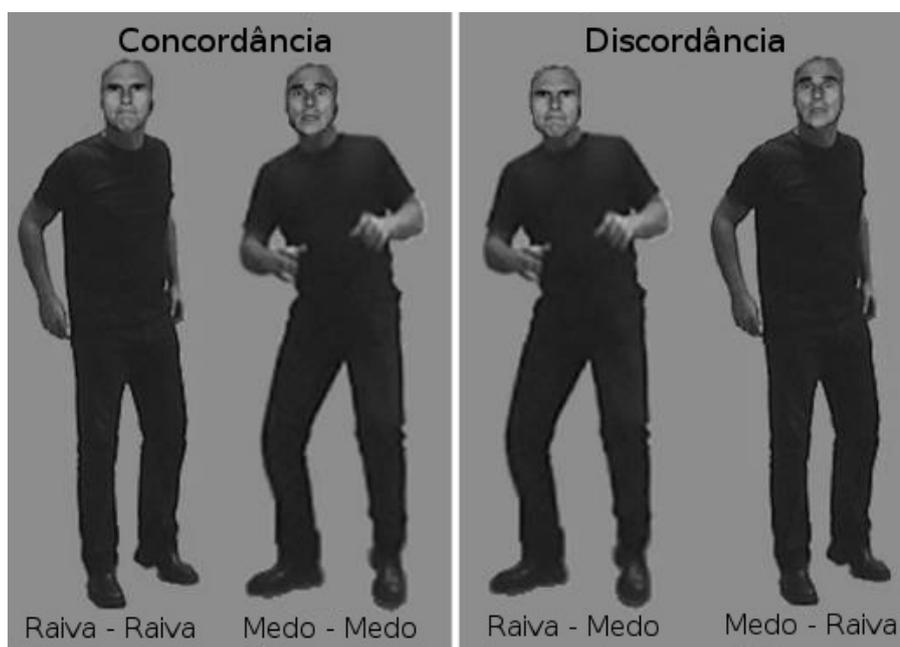


Figura 4.2: Exemplo de discordância entre as expressões da face e do corpo para emoções de “medo” e “raiva” (Meeren *et al.*, 2005).

4.3.1 Elementos afetivos na expressão corporal

O processo de reconhecimento o estado afetivo de um indivíduo a partir das expressões corporais é complexo em termos da neurociência. Diferentes elementos afetivos devem ser considerados neste processo, tais como: (i) tipo de configuração do corpo; (ii) propósito do movimento corporal e (iii) tipo de estímulo emocional.

O corpo é capaz de transmitir pistas afetivas a partir diferentes formas de configurações corporais. Assim, duas categorias amplas destas configurações são:

- Configuração estática;
- Configuração dinâmica.

A configuração estática ou postura é capaz de transmitir informações através de partes ou de todo corpo. Determinadas posturas são capazes de transmitir informações emocionais com precisão (Thakur e Han, 2018). Assim, informações anatômicas e geométricas do corpo, tais como posicionamento do tronco (Meijer, 1989), cabeça (Schouwstra e Hoogstraten, 1995) e braços (Wallbott, 1998), podem ser utilizadas para realizar associação com as classes de emoções.

A configuração dinâmica consiste em mudanças da postura corporal ao longo do tempo. De um modo geral, a configuração dinâmica transmite melhor informações relacionadas a intensidade (excitação/relaxamento) e não da categorização da emoção (Ekman e Friesen, 1974), embora exista certos movimentos corporais capazes de transmitir determinadas emoções (Poyo Solanas *et al.*, 2020). As informações cinemáticas do corpo, ou parte dele, são a base para determinação do estado afetivo através da configuração dinâmica do corpo (Hartmann *et al.*, 2019; Glowinski *et al.*, 2011).

A análise das configurações estática e dinâmica do corpo pode resultar em informações redundantes (Roether *et al.*, 2009). Assim, estas redundâncias devem ser tratadas de maneira adequada, podendo ser atribuído um nível de certeza maior nas informações coincidentes de um indivíduo observado.

As configurações estática e dinâmica do corpo nem sempre estão relacionadas às ações sociais e expressões emocionais. O desenvolvimento de abordagens para

o reconhecimento automático do estado afetivo de um indivíduo a partir de sua expressão corporal deve ser também capaz de diferenciar tais ações. Para isto, Karg *et al.* (2013) categorizou os movimentos corporais em:

Movimentos comunicativos. Relacionados a uma ampla variedade de movimentos que são realizados no cotidiano e podem transmitir emoção;

Movimentos funcionais. Objetivam realizar alguma tarefa, sendo que a expressividade emocional não é o objetivo principal;

Movimentos artísticos. Podem ser movimentos coreografados ou não coreografados e são caracterizados por expressar a emoção de maneira exagerada;

Movimentos abstratos. Movimentos que não são relacionados explicitamente para tarefas do dia-a-dia e nem para comunicar algo. Estes movimentos podem ser analisados independentemente de uma eventual significado simbólico, por exemplo um levantar de braço.

O processo de extração de características emocionais a partir da observação da expressão corporal necessita de fontes de modelos realísticos e precisos. Duas fontes de estímulos corporais possíveis são:

- (i) Não espontâneo;
- (ii) Espontâneo.

Nos estímulos corporais emocionais não espontâneos, a fonte de estímulos consiste na observação de pessoas agindo de maneira artificial, fazendo-se uso de atores. Estes são instruídos a expressar especificamente um estado afetivo ou emoção. A principal crítica nesta abordagem é que tais estímulos podem parecer não naturais ou exagerados, dependendo assim, da qualidade interpretativa do profissional (Yacoub *et al.*, 2003). Contudo, a principal vantagem está relacionada a possibilidade da definição de maneira mais específica dos estímulos corporais desejados (Castellano *et al.*, 2010). Esta abordagem é também recomendada em situações em que há uma barreira ética na aquisição de emoções intensas e genuínas no ambiente de laboratório (Kleinsmith e Bianchi-Berthouze, 2013).

Devido às desvantagens apresentadas no modelo de representação de expressões emocionais não espontâneas, pesquisadores veem desenvolvendo trabalhos sobre reconhecimento emocional a partir de expressões corporais espontâneas. A principal vantagem obtida nesta abordagem é que os gestos e movimentos ocorrem de forma natural, logo, mais realísticos. Entretanto, há a necessidade da presença de um especialista da área comportamental para definir as expressões corporais em termos de estado emocional (Kleinsmith e Bianchi-Berthouze, 2013).

Tanto nos estímulos corporais espontâneos como não espontâneos podem ser influenciados pela idiosincrasia. A idiosincrasia consiste na maneira específica de agir de cada indivíduo. O trabalho de Wallbott (1998) observou que a idiosincrasia está presente no momento em que os atores interpretavam seus movimentos. Assim, o estudo da expressividade emocional através do corpo deve prever esta particularidade dos movimentos em diferentes atores (Ahmed *et al.*, 2020). Além da idiosincrasia, tais estímulos corporais também podem ser influenciados por aspectos culturais (Kounine, 2017). Contudo, há determinadas expressões corporais que podem ser consideradas independentes da cultura (Kounine, 2017).

4.3.2 Caracterização emocional na expressão corporal

As pesquisas sobre o uso das expressões corporais como modal de expressividade afetiva apresentam como desafios a definição e a extração de características emocionais a partir das configurações estáticas e dinâmicas do corpo do indivíduo. Para caracterizar determinadas expressões corporais, pode-se utilizar descritores emocionais de baixo-nível ou alto-nível (Kleinsmith e Bianchi-Berthouze, 2013).

Os descritores baixo-nível relacionam os estados afetivos com informações detalhadas sobre a configuração do corpo e não depende de um contexto. Deste modo, são utilizadas informações descritivas mais detalhadas e mensuráveis, como por exemplo, dados da orientação, posicionamento, cinemático (velocidade, aceleração e energia cinética), força, fluidez e regularidade (ou periodicidade) das articulações dos membros do corpo. (Kleinsmith e Bianchi-Berthouze, 2013).

Por outro lado, os descritores alto-nível abordam características gerais sobre

a configuração estática (forma) e dinâmica (movimento) do corpo para a discriminação dos estados afetivos através de um observador humano, baseando-se em um contexto. Os descritores mais comuns são abertura e posicionamento do corpo, forma do movimento dos membros e retidão dos movimentos (Meijer, 1989; Gross *et al.*, 2010; Kleinsmith e Bianchi-Berthouze, 2013).

Esses descritores relacionam informações corporais com a intenção emocional. Normalmente, estas informações são analisadas em conjunto para obter maior precisão. De acordo com Schouwstra e Hoogstraten (1995) e Wallbott (1998), algumas destas informações podem estar associadas a mais de um estado emocional. Na literatura atual, não há um consenso dos nomes ou quais informações mais úteis para caracterização emocional a partir da expressão corporal.

Glowinski *et al.* (2011) propõem algumas equações que definem as características de expressividade que podem ser usadas em sistemas de reconhecimento automático de estado afetivo por meio da expressão corporal. Estas equações estão associadas aos elementos cinemáticos do corpo humano, tais como posição (no espaço 3D), velocidade e aceleração. Assim, tais categorias são:

Energia. Caracteriza a expressividade emocional relacionando a energia gasta pelo indivíduo para realizar um movimento tais com, extensão espacial, suavidade, simetria e inclinação da cabeça. Este descritor foi denominado de Quantificador de Movimento (QoM) por Camurri *et al.* (2003). Altos valores estão associados ao estado de alta ativação como estado de terror, raiva e entusiasmo. Valores baixos estão relacionados ao estado de baixa ativação, como tédio, tristeza. Seja $v_l(f)$ o módulo da velocidade de cada segmento do membro do corpo l e um instante f :

$$v_l(f) = \sqrt{\dot{x}_l(f)^2 + \dot{y}_l(f)^2 + \dot{z}_l(f)^2}, \quad (4.1)$$

onde \dot{x}_l , \dot{y}_l e \dot{z}_l são as variações de posições das coordenadas x , y e z entre os instantes $f - 1$ e f .

Deste modo, a energia cinética aproximada, E_{tot} , é determinada pela soma

ponderada de cada energia cinética dos membros do corpo:

$$E_{tot}(f) = \frac{1}{2} \sum_{l=1}^n m_l v_l(f)^2, \quad (4.2)$$

onde m_l é a massa aproximada da articulação de segmentos de membros do corpo baseada na tabela antropofométrica (Liang *et al.*, 2016).

Extensão espacial (limite triangular). Relaciona o nível abertura do corpo com a área do triângulo formada entre as mãos e a cabeça (ver Figura 4.3). A abertura do corpo é associada ao nível de acessibilidade e ligação com o comunicador, estabelecendo assim empatia (Meijer, 1989).

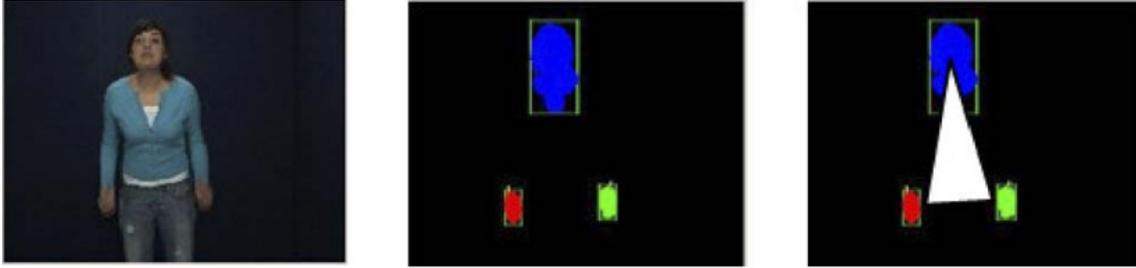


Figura 4.3: Determinação da extensão espacial formado pela área do triângulo das (Meijer, 1989).

Suavidade. Descreve os movimentos ondulares e circulares com as mãos com velocidade média ou alta, mas sem mudanças abruptas na velocidade e aceleração (Wallbott, 1998). A suavidade dos movimentos está associada ao nível de ativação dos indivíduo. Deste modo, a curvatura (k) mede a taxa de alteração da direção de um vetor tangente em uma trajetória curva. Então, k é calculada, considerando uma observação 2D, como,

$$k = \frac{\dot{x} \cdot \ddot{y} - \dot{y} \cdot \ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}, \quad (4.3)$$

onde \dot{x} , \ddot{x} e \dot{y} , \ddot{y} são as derivadas de primeira, segunda ordem da trajetória das mãos em relação aos componentes horizontais e verticais, respectivamente.

Simetria. A assimetria lateral relacionada a expressão emocional tem sido estudada através das expressões faciais (Wang *et al.*, 2017). O trabalho de Giese *et al.* (2008) demonstra que as assimetrias laterais de todo o corpo humano também são capazes de transmitir sugestões emotivas. Neste trabalho, observou-se que para expressão de “raiva”, “alegria” e “tristeza” do lado esquerdo do corpo apresentou significantes valores de maior amplitude e energia. Camurri *et al.* (2003) consideram características emocionais a partir da simetria dos gestos das duas mãos. Esta simetria é calculada através do eixo horizontal (SI_h) e do eixo vertical (SI_v). O SI_h é calculada a partir da posição baricentral das mãos,

$$SI_h = \frac{||x_B - x_L| - |x_B - x_R||}{|x_R - x_L|} \quad (4.4)$$

onde x_B é coordenada da abscissa do baricentro, x_L é a coordenada da borda esquerda do limite triangular, e x_R é a coordenada da borda direita da abscissa do limite triangular.

A assimetria vertical (SI_v) é calculada pela diferença entre as coordenadas do eixo da ordenadas das mãos. Assim, a simetria espacial ($SI_{espacial}$) é calculada como,

$$SI_{espacial} = \frac{SI_h}{SI_v} \quad (4.5)$$

Inclinação para cima e para baixo da cabeça. O movimento da cabeça é estudado no campo da música. A importância deste movimento demonstra intenções emocionais (Hartmann *et al.*, 2019). Deste modo, Glowinski *et al.* (2011) quantificam a velocidade da Inclinação da Cabeça (IC) em relação ao componente de profundidade (z) considerando o eixo referencial ilustrado pela Figura 4.4,

$$IC(f) = \dot{z}. \quad (4.6)$$

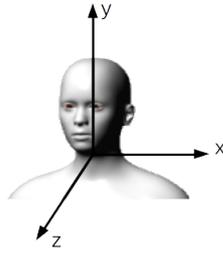


Figura 4.4: Eixo referencial da cabeça.

4.4 Aspectos Emotivos nos Comportamentos Estereotipados Autistas

Os comportamentos estereotipados em indivíduos com TEA estão relacionados com o aumento dos níveis do estado mental (nível de ativação) (Ohnishi *et al.*, 2000). Indivíduos autistas afetados por tais comportamentos podem se tornar ansiosos, agitados ou agressivos (Troisi, 2002; Rodrigues *et al.*, 2013). Pacientes sob estas condições mentais podem sofrer prejuízos no processo terapêutico, na qualidade de vida e no convívio social. Assim, retrocessos terapêuticos e formação de estigma ou isolamento social são algumas das consequências deste estado mental em indivíduos autistas ao manifestarem comportamento estereotipados (Mason, 1991). O TEA é descrito no Capítulo 3.

4.5 Considerações Finais

As pesquisas relacionadas aos fenômenos afetivos apresentam uma abordagem multidisciplinar, envolvendo as áreas da neurociência, engenharia e computação. Embora muitos avanços tenham ocorrido a fim de elucidar o comportamento afetivo humano, nota-se ainda que o domínio da compreensão dos mecanismos emocionais ainda não está bem esclarecido.

A maioria das pesquisas tem focado na expressão facial e análise da voz, muitas vezes influenciadas pelo trabalho de Darwin (Spunt *et al.*, 2017). Pesquisas também têm estudado os modais de expressividade emocional através dos sinais biológicos

(EEG, ECG, EMG, respiração e condutividade da pele). Além destes modais, a obtenção de pistas afetivas a partir das expressões corporais é uma abordagem que tem sido tratado em trabalhos mais recentes.

O processo de compreensão de fenômenos afetivos através da expressão corporal se baseia na extração de características de duas variáveis: a configuração estática (pose) e dinâmica (movimento) do corpo. Contudo, é importante considerar que nem sempre tais variáveis estão relacionadas à expressão emocional. A descrição do estado afetivo através dessas variáveis podem fazer uso de descritores de baixo-nível (ângulo e posicionamento das articulações ou membros), como também o uso de descritores de alto-nível (cinemática do movimento do corpo ou nível de abertura do corpo).

A análise das poses e movimentos do corpo pode ser realizada a partir de estímulos espontâneos ou não-espontâneos. Tais abordagens apresentam vantagens e desvantagens. Nos estímulos não espontâneos, por fazer uso de atores, a análise emocional pode depender da idiosincrasia ou interpretação artística do ator, que pode ser não tão condizente com a realidade. Contudo, esta abordagem é adequada em situações em que questões éticas podem ser impeditivas, como por exemplo, simulação de dor extrema. Já na espontânea, é possível obter informações corporais realísticas, mas há a dificuldade de obter, de forma isolada, a expressividade emocional, necessitando muitas vezes de um especialista em teoria das emoções.

Capítulo 5

Arquitetura do Sistema e Experimentos

Neste capítulo, é apresentado o sistema proposto para reconhecimento de estado afetivo a partir das expressões corporais denominado Módulo de Reconhecimento de Expressões Corporais (MREC). No contexto da robótica assistiva aplicada ao Hi-bot, esta ferramenta visa inferir o nível de ativação a partir de comportamentos estereotipados típicos do autismo. A arquitetura geral do MREC é detalhada na Seção 5.1.

Uma análise de desempenho baseada na arquitetura do MREC tem como objetivo identificar os AAMs e determinar a sua configuração mais adequada, possibilitando uma escolha mais sistemática destes algoritmos. A definição metodológica da execução deste experimento é abordada na Seção 5.2. Os experimentos são divididos em duas etapas: (i) classificação dos comportamentos estereotipados (Seção 5.3); e (ii) inferência do nível de ativação (Seção 5.4). Esta divisão dos experimentos permite averiguar os resultados individualizados de tais ferramentas. Uma breve discussão sobre este capítulo é realizada na Seção 5.5.

5.1 Visão Geral

O MREC é um módulo que faz parte de um conjunto de interfaces de sensoriamento afetivo previstas para plataforma robótica de experimentos de Interação Humano-Robô (IHR) denominada HiBot (ver Seção 1.1 do Capítulo 1). O MREC deve ser capaz de inferir o estado afetivo de um indivíduo a partir de suas expressões corporais. No contexto de indivíduos com TEA, o MREC deve ser capaz de reconhecer o nível de ativação de indivíduos autistas a partir dos seus comportamentos estereotipados típicos. A arquitetura deste módulo é ilustrada na Figura 5.1.

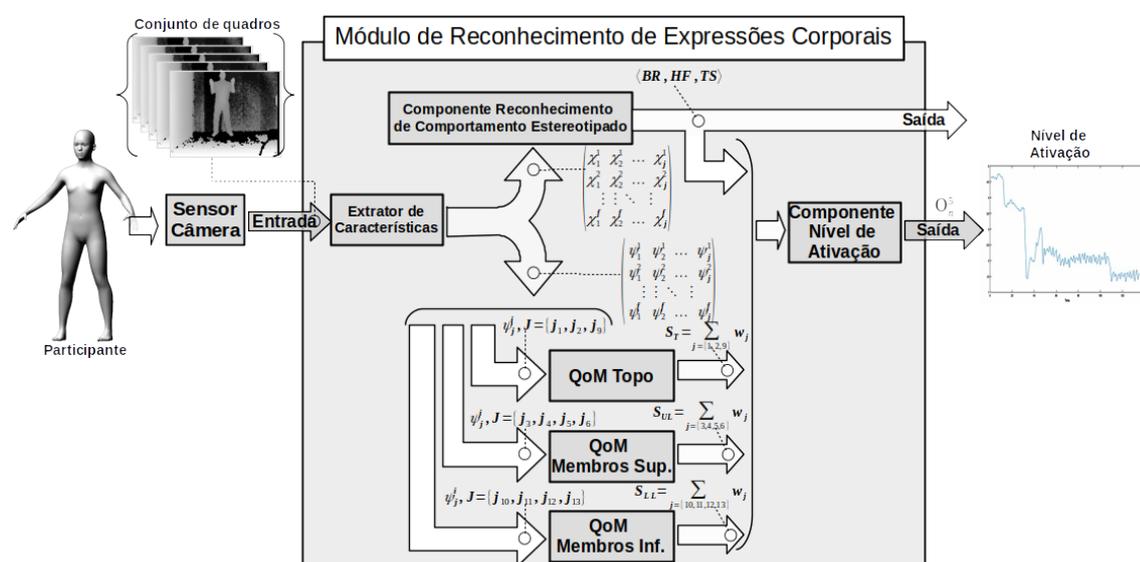


Figura 5.1: Arquitetura geral do Módulo de Reconhecimento de Expressões Corporais (MREC).

A captura dos quadros de imagens é realizada pelo dispositivo sensor câmera denominado Kinect[®]. Os recursos deste dispositivo são ilustrados na Figura 5.2. O Kinect[®] apresenta um conjunto de recursos, tais como um emissor de infravermelho, sensor de profundidade infravermelho, sensor de cor, um conjunto de microfones e um motor de inclinação vertical. Os três primeiros recursos são usados para obter quadros RGB com informações de profundidade (RGB-D) sobre um determinado cenário e elementos contidos nele tais como, pessoas e objetos. A taxa máxima de captura de quadros, tanto para cor como profundidade, é de 30 quadros/segundo e 640x480 *pixels* de resolução. Esta sequência de quadros RGB-D são os dados de

entrada do MREC.

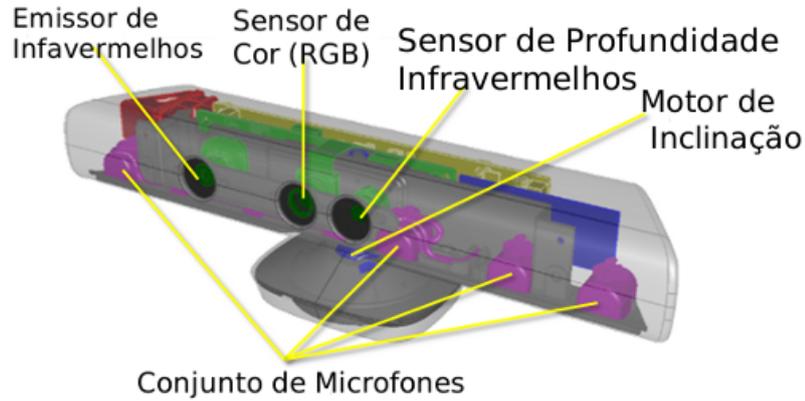


Figura 5.2: Recursos presentes no Kinect[®]: emissor de infravermelho, receptor de profundidade infravermelho, sensor RGB, conjunto de microfones e um motor de inclinação vertical.

Fonte: <https://developer.microsoft.com/pt-br/windows/kinect>

As características de interesse dos quadros RGB-D são obtidas pelo componente *Extrator de Características*. Este elemento é baseado nas bibliotecas OpenNI e NiTE (ver Subseção 5.2.1). As características de interesse são informações sobre posições e orientações definidas respectivamente pelas matrizes (5.1) e (5.2),

$$\begin{pmatrix} \chi_1^1 & \chi_2^1 & \cdots & \chi_j^1 \\ \chi_1^2 & \chi_2^2 & \cdots & \chi_j^2 \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1^f & \chi_2^f & \cdots & \chi_j^f \end{pmatrix}, \quad (5.1)$$

$$\begin{pmatrix} \psi_1^1 & \psi_2^1 & \cdots & \psi_j^1 \\ \psi_1^2 & \psi_2^2 & \cdots & \psi_j^2 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1^f & \psi_2^f & \cdots & \psi_j^f \end{pmatrix}, \quad (5.2)$$

onde j é uma determinada articulação e f é um determinado quadro do sensor câmara. A posição χ_j^f é dada em uma coordenada 3D, (x, y, z) . A orientação ψ_j^f segue a notação de quatérnios. Em um cenário 3D, a utilização de orientação seguindo a notação de quatérnios apresenta vantagem de evitar problemas de singularidade relacionados a operações de rotação (problema do Bloqueio do Cardan). A Figura 5.3 ilustra as 15 articulações consideradas pelas bibliotecas OpenNI/NiTE: cabeça (J_1), pescoço (J_2), ombros (J_3 e J_4), cotovelos (J_5 e J_6), mãos (J_7 e J_8), tronco (J_9), quadris (J_{10} e J_{11}), joelhos (J_{12} e J_{13}) e pés (J_{14} e J_{15}).

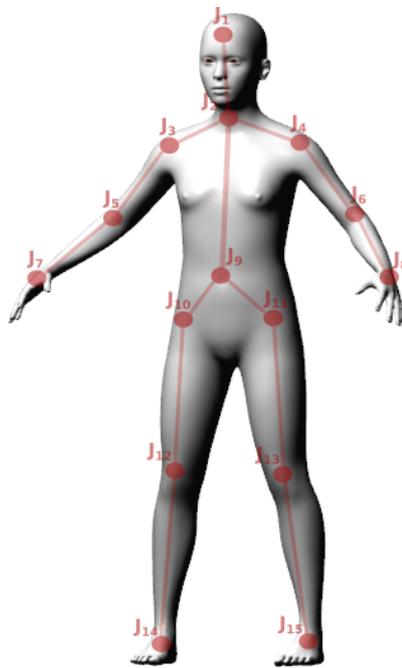


Figura 5.3: Articulações consideradas pelas bibliotecas OpenNI/NiTE.

O processo de classificação dos comportamentos estereotipados é realizado pelo *Componente de Reconhecimento de Comportamentos Estereotipados* (CRCE). As informações sobre o posicionamento das articulações, representadas pela Matriz (5.1), são as entradas deste componente. A Tabela 5.1 apresenta a estatística dos valores das coordenadas de todas as articulações no eixo referencial (Figura 5.4(a)) consideradas no processo de classificação. Assim, é possível observar que estas posições podem assumir uma ampla variedade de valores no cenário original, aumentando a complexidade do classificador. Assim, duas operações são realizadas para reduzir o

Tabela 5.1: Valores estatísticos das coordenadas das articulações no eixo referencial $x - y - z$ do Kinect[®].

Articulação	x			y			z		
	Mín.	Máx.	$\mu \pm \sigma$	Mín.	Máx.	$\mu \pm \sigma$	Mín.	Máx.	$\mu \pm \sigma$
Cabeça	-27,93	2258,92	1019,11 ± 1155,55	-29,00	2218,38	1001,34 ± 1135,26	-31,69	2207,98	1004,80 ± 1129,10
Pescoço	-36,74	2262,18	929,77 ± 1192,34	-35,48	2262,50	930,43 ± 1191,95	-32,79	2252,91	930,24 ± 1184,53
Ombro Dir.	119,12	2240,86	975,81 ± 1118,27	120,18	2242,58	976,70 ± 1118,85	122,37	2230,78	975,08 ± 1110,47
Ombro Esq.	-192,60	2283,50	883,72 ± 1269,35	-191,15	2282,43	884,16 ± 1268,02	-187,96	2275,04	885,39 ± 1261,59
Cotovelo Dir.	192,85	2326,51	962,74 ± 1184,34	194,83	2330,97	965,58 ± 1185,74	203,27	2320,69	966,24 ± 1176,11
Cotovelo Esq.	-256,35	2372,42	819,19 ± 1377,95	-255,28	2366,32	816,60 ± 1374,57	-250,60	2357,95	816,38 ± 1367,50
Tronco	-32,29	2254,16	848,43 ± 1230,31	-31,82	2253,99	848,46 ± 1230,10	-31,48	2246,43	851,23 ± 1222,38
Quadril Dir.	66,57	2233,22	794,99 ± 1245,57	66,93	2233,32	794,76 ± 1245,85	64,46	2226,46	799,56 ± 1235,92
Quadril Esq.	-122,22	2259,04	739,21 ± 1320,12	-123,23	2257,65	738,24 ± 1319,78	-124,78	2253,45	744,87 ± 1311,55
Joelho Dir.	-357,31	2235,29	659,85 ± 1383,53	-360,66	2236,25	659,18 ± 1385,23	-356,02	2239,89	662,20 ± 1385,44
Joelho Esq.	-379,64	2260,56	581,95 ± 1458,85	-374,89	2267,42	585,96 ± 1461,13	-377,98	2260,07	582,32 ± 1458,04

universo destes valores. Primeiro, a origem do eixo é deslocada para a articulação do tronco (Figura 5.4(b)). Em seguida, as coordenadas das articulações são normalizadas em cada eixo tridimensional (Figura 5.4(c)). A saída deste componente consiste em uma determinada classe de comportamento estereotipado. Esta saída alimenta o *Componente de Nível de Ativação* e também é a saída do MREC. As três classes de comportamentos estereotipados considerados são BR, HF e TS. O processo de classificação depende do classificador a ser determinado que é descrito no Anexo B e os experimentos realizados na Seção 5.3.

Um conjunto de qualificadores emocionais é definido pelo *Quantificador de Movimento (Quantify of Motion – QoM)* (Camurri *et al.*, 2003). Os quantificadores cinemáticos podem mensurar o nível de ativação de um indivíduo. Neste trabalho, o quantificador emocional cinemático considerado é o módulo da velocidade angular (w_j) de uma articulação (j). A orientação das articulações é representada pela Matriz (5.2). Cada componente QoM é relacionado a um grupo de articulações (J). Estes componentes retornam a soma de todas as articulações (S_Ω) que fazem parte,

$$S_\Omega(J) = \sum_{j=J} w_j, \quad (5.3)$$

onde J é o conjunto de todos os indexadores das articulações de um grupo Ω . A Figura 5.5 mostra três componentes propostos: (i) *QoM Topo* (cabeça, pescoço e tronco); (ii) *QoM Membros Superiores* (ombros e cotovelos); e (iii) *QoM Membros Inferiores* (quadril e joelhos). As mãos e os pé não pertence a nenhum dos grupos, pois as bibliotecas OpenNI e NITE não proveem dados de orientação destas articu-

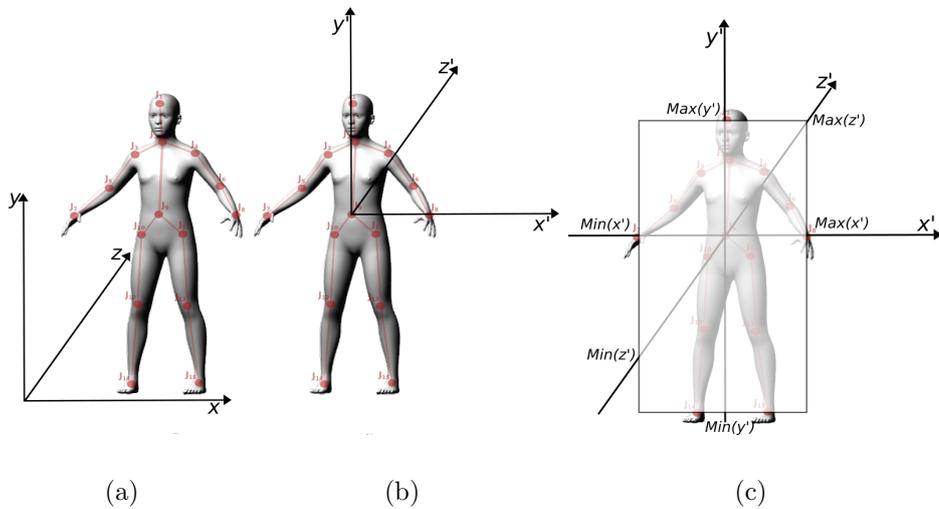


Figura 5.4: Operações para simplificar a entrada do classificador. (a) sistema de eixos referencial; (b) deslocamento da origem dos eixos para a articulação do tronco; (c) normalização das articulações.

lações. Cada componente QoM retorna um valor $S_c \in \mathfrak{R}$, onde c é um determinado componente: *QoM Topo*, *QoM Membros Superiores* e *QoM Membros Inferiores*. Assim, os valores destes componentes são S_T , S_{UL} e S_{LL} , respectivamente. Outros descritores presentes na literatura (Subseção 4.3.2 do Capítulo 4) não estão relacionados ao nível de ativação ou são aplicados no contexto dos movimentos corporais da dança (inclinação da cabeça).

O *Componente de Nível de Ativação* (CNA) é proposto neste trabalho e baseia-se em uma arquitetura do *Adaptive-Network-Based Fuzzy Inference System* (ANFIS) e tem como propósito inferir o nível de ativação de um indivíduo, representado por $O_n^5 \in \mathfrak{R}$ no intervalo $[0, 1]$ (ver equação B.34 do Apêndice B), onde valores graduais determinam o nível de ativação com valores mínimo e máximo 0 e 1, respectivamente. As entradas do CNA são o comportamento estereotipado reconhecido (BR, HF, TS), qualificadores cinemáticos determinados pelos componentes QoM (S_T , S_{UL} e S_{LL}).

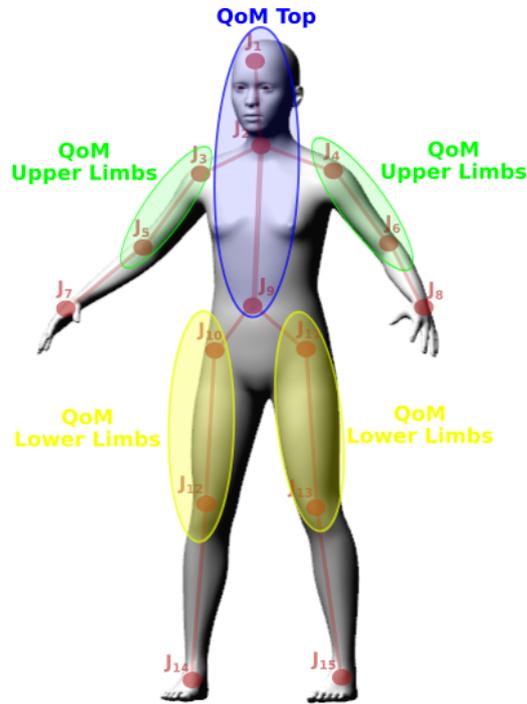


Figura 5.5: Definição dos grupos QoMs. QoM Topo [Azul] : cabeça (J_1), pescoço (J_2), tronco (J_9); QoM Membros Superiores [Verde]: ombros (J_3, J_4), cotovelos (J_5, J_6); QoM Membros Inferiores [Amarelo]: quadril (J_{10}, J_{11}), e joelhos (J_{12}, J_{13}).

5.2 Metodologia

Neste trabalho, os experimentos (treino/teste) foram realizados no mesmo dispositivo computacional de maneira *off-line*¹. Deste modo, as amostras gerais foram coletadas e depois utilizadas para os experimentos. Dois cenários foram determinados para cada comportamento estereotipado: níveis de ativação alto e baixo. O tipo de fonte de estímulo corporal destes cenários foi não-espontâneo (ver Seção 4.3 do Capítulo 4) por meio de 6 participantes voluntários não autistas, sendo 1 feminino e 5 masculinos com as idades entre 18 e 37 anos. Embora o uso de participantes não autistas apresente a desvantagem de depender da qualidade artística do participante, esta abordagem é justificada pela facilidade de se obter maior quantidade de amostras bem definidas. Além disto, há a barreira ética por envolver menores de idade em condição especial de saúde. Nesta situação, é necessário ser autorizado

¹As amostras estão disponíveis em <https://github.com/mcamada/samples-mrec>

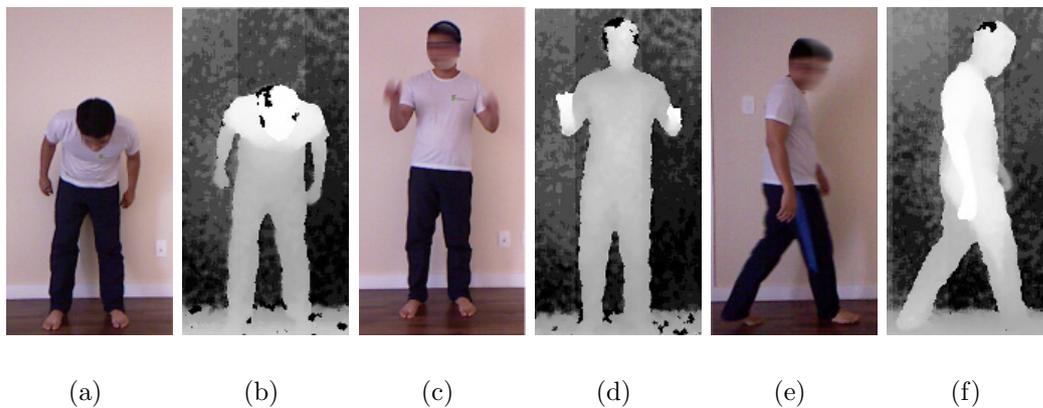


Figura 5.6: Comportamentos estereotipados típicos obtidos pela câmera Kinect[®] nos seus respectivos quadros de imagem e profundidade: BR (a) e (b); HF (c) e (d) e TS (e) e (f).

pelo Comitê de Ética.

A atuação dos participantes foi capturada por uma câmera Kinect[®], e seus dados articulares (posição e orientação) foram extraídos pelo componente *Extrator de Características* do MREC (ver Seção 5.1), e armazenados. A Figura 5.6 mostra os quadros RGB e de profundidade dos comportamentos estereotipados de um determinado participante.

Os participantes foram orientados sobre a caracterização dos comportamentos estereotipados antes da atuação. Durante a atuação, os participantes foram acompanhados por um supervisor. Os comportamentos inapropriados foram removidos manualmente, e os de interesse foram extraídos e segmentados.

Este trabalho selecionou os comportamentos estereotipados autistas mais comuns na literatura relacionada presente na Tabela 2.1 e nos trabalhos Kuhn (1999), Lovaas *et al.* (1971) e Watt *et al.* (2008). Estes comportamentos são listados e descritos como: (i) *Body Rocking* (BR) – movimento repetitivo para frente e para trás da parte superior do tronco (ver Figuras 5.6(a) e 5.6(b)); (ii) *Top Spinning* (TS) – andar ou correr em círculo (ver Figuras 5.6(e) e 5.6(f)); (iii) *Hand Flapping* (HF) – movimento de balançar para cima e para baixo (ver Figuras 5.6(c) e 5.6(d)); (iv) *Head Banging* (HB) – bater a cabeça no chão ou na paredes. O HB não foi

considerada neste trabalho, pois a sua trajetória é semelhante ao BR. Além destes comportamentos estereotipados, o Bater dos Dedos é tratado com menor frequência nos trabalhos (DiGennaro Reed *et al.*, 2012; Coronato *et al.*, 2014). O sensor câmera utilizado neste trabalho apresenta baixa resolução para detecção dos movimentos dos dedos. Portanto, devido a estes motivos, o comportamento estereotipados de Bater dos Dedos não foi considerado.

Para cada participante, o conjunto de amostras é constituído de 60 gestos dos comportamentos estereotipados (BR, HF, TS) e níveis de ativação (alto e baixo). Assim, para cada comportamento estereotipado de um determinado participante, são obtidos 120 amostras.

Os experimentos de treino e teste foram conduzidos em duas fases:

1. Análise de desempenho dos AAMs para o reconhecimento de comportamentos estereotipados (Seção 5.3);
2. Determinação do menor erro para inferir o nível de ativação (Seção 5.4);

Na primeira fase, os AAMs escolhidos foram por decisão de projeto motivados por serem algoritmos bem consolidados no processo de reconhecimento de gestos, como também por estar presentes nos trabalhos relacionados de reconhecimento de comportamentos estereotipados autistas (ver Tabela 2.1). Além disto, eles fazem parte de diferentes famílias de Algoritmos de Aprendizado de Máquina.

Os classificadores foram definidos para classificação binária. Assim, para cada rótulo (comportamento estereotipado), as amostras eram definidas para duas classes: verdadeiro ou falso. A primeira classe se uma determinada instância representa o rótulo considerado, e falso caso contrário.

Na segunda fase, a escolha do ANFIS se deve a necessidade de se investigar a natureza das características imprecisas ou nebulosas para determinação estado afetivo de indivíduos autistas. O rótulo afetivo escolhido para inferência da ferramenta ANFIS consiste no nível de ativação. O nível de ativação pode ser percebido através dos comportamentos estereotipados autistas (Troisi, 2002; Gaffary *et al.*, 2018). Deste modo, o nível de ativação alto associado aos comportamentos estereotipados pode ser considerado inapropriado para o processo terapêutico dos autistas, como

também pode estar relacionado ao nível de defesa do mesmo (Kuhn, 1999) ou mecanismo de auto-estimulação (Cunningham e Schreibman, 2008) ocasionados pela hipersensibilidade ambiental. Além disto, de acordo com Ekman (1992), as expressões corporais são capazes de melhor transmitir informações emocionais relacionadas a excitação e relaxamento.

5.2.1 Bibliotecas de software auxiliares para extração de características

O componente do MREC denominado Extrator de Características é constituído por duas bibliotecas de *software*: OpenNI e o NiTE. Estas dois *softwares* são mantidas pela empresa PrimeSense¹. O OpenNI é um arcabouço de *software* que visa ser uma Interface de Programação de Aplicações (*Application Programming Interface – API*) padrão para comunicação de sensores relacionados a Interação Naturais (*Natural Interaction – NI*), como de áudio e vídeo. O OpenNI é suportado em diferentes plataformas de Sistemas Operacionais e linguagens de programação (PrimeSense, 2011). A Figura 5.7 ilustra a arquitetura geral do OpenNI organizada em diferentes módulos.

O módulo de mais baixo nível lida diretamente com o dispositivo *NI*, que pode ser um sensor de áudio, cor ou profundidade. Este módulo apresenta os *softwares* (*drives*) necessários para lidar com as interfaces dos dispositivos. Acima deste módulo, existe um camada que encapsula as interfaces de comunicação com o dispositivo. O módulo Núcleo apresenta as operações básicas que gerenciam os recursos e falhas. A *API* apresenta um conjunto de ferramentas de mais alto nível para as aplicações superiores como análise corporal completa como informações sobre a articulação e centro de massa), rastreamento de mão, detecção de gestos, análise do cenário como separação do planos do cenário e coordenadas do planos (PrimeSense, 2011).

O algoritmo de reconhecimento de pose do OpenNI é baseado no trabalho de Shotton *et al.* (2011) em que é proposto um algoritmo de reconhecimento de pose humana em tempo real utilizando imagens de profundidade. Este algoritmo trata de

¹Em 2015, a PrimeSense foi vendida para Structure



Figura 5.7: Arquitetura das bibliotecas de software OpenNI (PrimeSense, 2011).

classificar cada *pixel* em partes do corpo, baseando em um base de dados de treino, o que permite estimar partes invariantes do corpo como forma e pose. O algoritmo utiliza Arvore de Decisão Aleatória para realizar a classificação. Após determinar os *pixels* pertencentes a cada região, é definido o centro de massa desta região, que representará o centro da articulação do segmento de um membro do corpo.

Outra biblioteca de *software* denominada NiTE é usando entre o OpenNI e MREC. Esta ferramenta apresenta algoritmos (Algoritmos NiTE) que oferecem funcionalidades de mais alto nível que as oferecidas pelo OpenNI, como rastreamento de movimentos de mãos e reconhecimento de gestos (PrimeSense, 2011). Neste trabalho, o NiTE é utilizado para obter informações sobre as articulações dos segmentos do corpo (posição e orientação).

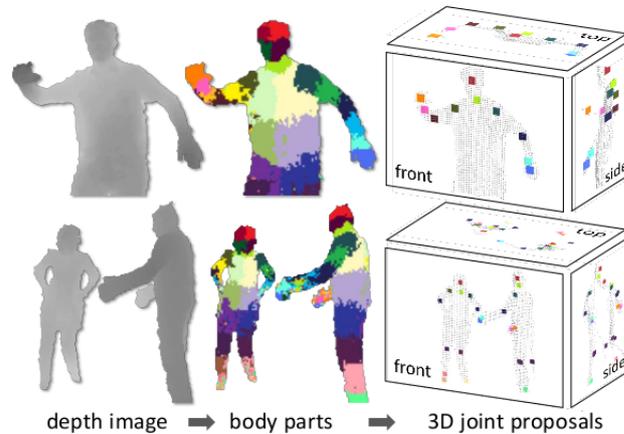


Figura 5.8: Definição das regiões das articulações dos segmentos dos membros do corpo (Shotton *et al.*, 2011).

5.2.2 Configuração dos experimentos

A primeira fase dos experimentos foi utilizada o *Waikato Environment for Knowledge Analysis* (WEKA). Esta ferramenta permite comparar o desempenho de determinados AAM em relação a diferentes critérios estatísticos armazenando em arquivos ou banco de dados. Nesta fase, foi realizado uma análise de desempenho do HMM, SVM, MLP e CNN para reconhecimento de comportamentos estereotipados.

Os experimentos para cada AAM foram organizados de duas formas: (i) análise de desempenho com amostras de treino/teste com o mesmo participante, e (ii) análise de desempenho com amostras de 4 participantes para o treino e amostras de 2 participantes para o teste. A primeira forma visa averiguar o desempenho dos algoritmos de classificação baseados na amostras de treino e teste providos pelo mesmo participante. Assim, cada algoritmos de classificação foi executado para cada participante.

A segunda forma deve determinar a capacidade de generalização dos algoritmos para classificação dos comportamentos estereotipados. As amostras dos Participantes 1, 2, 5, e 6 foram usadas para o treinamento, e as amostras dos Participantes 3 e 4 foram usados para o teste. Nesta última forma, os experimentos foram baseados no gênero feminino (Participante 4) para averiguar a influência do gênero.

Para as duas formas, foi utilizado a técnica de validação cruzada com *10-fold*. A validação cruzada é uma técnica empregada para avaliar a capacidade de generalização de modelos de AAM através do treino de diferentes modelos com subconjuntos de dados de entrada disponíveis. O processo da validação cruzada com *10-fold* consiste em treinar aleatoriamente e em separado com 10 conjuntos, tendo 9 conjuntos usados para o treinamento e 1 para validação.

Embora, a acurácia e a precisão sejam úteis para a análise de desempenho entre AAMs, elas não são suficientes como medida de desempenho aplicado em um ambiente clínico real. Em vista disto, este trabalho introduz um critério de desempenho denominado de Índice de Desempenho Temporal (IDT) definido como,

$$\text{IDT} = \frac{\mathcal{A} + \mathcal{P}}{\mathcal{T}}, \quad (5.4)$$

onde \mathcal{T} é o tempo de reconhecimento do comportamento estereotipado nos experimentos e seu valor mínimo é limitado a taxa dos quadros de um sensor Kinect[®] (ver Seção 5.1). Embora \mathcal{A} e \mathcal{P} sejam variáveis estatísticas importantes, somente elas não garantem um processo de interação. Deste modo, o tempo para reconhecimento do comportamento estereotipado é uma importante variável que contribui para a efetividade do processo interativo. Os trabalhos relacionados não consideram esta variável em suas análises de desempenho. O tempo de treino não foi considerado relevante para a análise dos AAM pois ele foi determinado *off-line*. Os valores do IDT são usados para análise de desempenho dos AAM (HMM, SVM, MLP e CNN) para classificar os comportamentos estereotipados (BR, HF, TS). Os resultados experimentais desta fase estão presentes na Seção 5.3.

A segunda fase foi realizada com o *software* MATLAB[®] com suporte ao ANFIS. O MATLAB apresenta um conjunto de ferramentas que facilita a análise do ANFIS dentre eles, gráficos e funções estatísticas nativos do próprio programa. Os experimentos com ANFIS foram divididos em duas fases: (i) desempenho de treino e (ii) desempenho de teste. A primeira fase visa determinar qual a estratégia de treino que apresenta o menor erro quadrático médio em relação ao número de interações de épocas. A segunda fase tem como objetivo avaliar a capacidade do ANFIS em diferenciar níveis de ativação a partir dos comportamentos estereotipados. Deste

modo é analisado a melhor estratégia de treino definida na fase anterior, baseando-se na distribuição normal dos resultados das amostras. A distribuição normal [ou gaussiana] descreve fenômenos naturais aleatórios, os quais geralmente interagem entre si. Ela é representada por uma curva gaussiana e definida por uma função denominada como função de densidade,

$$f(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}, \quad (5.5)$$

onde σ e μ são o desvio padrão e a média, respectivamente. Os resultados experimentais do processo de inferência do nível de ativação dos comportamentos estereotipados são apresentados e discutidos na Seção 5.4

5.3 Experimentos para Classificação dos Comportamentos Estereotipados

Os experimentos para classificação dos comportamentos estereotipados foram executados em relação aos seguintes AAMs: HMM, SVM, MLP e CNN. A definição dos parâmetros iniciais destes modelos foi uma decisão do projeto. A variação dos valores visam averiguar as melhores configurações destes modelos. A seguir, os resultados experimentais baseados nestes algoritmos são apresentados e discutidos.

5.3.1 Experimentos de classificação com HMM

Os HMMs aplicados neste trabalho foram modelados considerando um HMM para cada comportamento estereotipado. O maior valor da matriz de probabilidade (B) dentre os HMMs, determina o comportamento estereotipado reconhecido. O estado inicial (π) é definido aleatoriamente, e as matrizes de covariância (U) consideradas foram a diagonal e a esférica. Deste modo, a análise de desempenho do HMM foi realizada em relação ao tipo de matriz de covariância e a quantidade de estados (n_s). Para cada HMM, foram realizados 10 experimentos iniciando com 2 estados, aumentando em uma unidade de um experimento para outro. Uma breve

Tabela 5.2: Resultados experimentais do HMM com matriz de covariância diagonal variando o número de estados (n_s), baseando nas amostras de treino e teste do mesmo participante (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
1	2	98,33	100,00	0,03	5950	5	91,94	91,17	0,03	5493	3	96,11	91,76	0,03	5636		
2	3	83,61	95,75	0,03	5381	2	90,00	90,54	0,03	5416	3	81,67	81,30	0,03	4889		
3	7	98,61	99,09	0,03	5931	7	98,06	98,46	0,03	5896	6	98,33	95,69	0,03	5821		
4	2	93,33	93,92	0,03	5618	4	87,78	97,66	0,03	5563	6	85,83	96,26	0,03	5463		
5	2	90,56	89,78	0,03	5410	5	88,06	100,00	0,03	5642	4	85,83	84,37	0,03	5106		
6	2	97,50	95,43	0,03	5788	4	93,33	94,45	0,03	5633	3	96,94	94,07	0,03	5730		
Média					5680	Média					5607	Média					5441

Tabela 5.3: Resultados experimentais do HMM com matriz de covariância esférica variando o número de estados (n_s), baseando nas amostras de treino e teste do mesmo participante (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
1	5	92,22	96,72	0,03	5668	8	90,28	90,25	0,03	5416	5	94,17	91,70	0,03	5576		
2	2	89,17	92,16	0,03	5440	7	85,56	93,72	0,03	5378	2	88,06	78,69	0,03	5002		
3	8	95,83	95,89	0,03	5752	5	95,28	98,32	0,03	5808	3	95,83	91,40	0,03	5617		
4	5	83,89	95,31	0,03	5376	7	83,61	100,00	0,03	5508	2	93,89	87,55	0,03	5443		
5	7	82,22	93,54	0,03	5273	4	88,89	94,45	0,03	5500	3	89,72	81,41	0,03	5134		
6	7	91,11	98,89	0,03	5700	5	94,17	91,42	0,03	5568	3	98,06	94,83	0,03	5787		
Média					5535	Média					5530	Média					5427

descrição sobre o HMM é realizada na Seção B.1 do Apêndice B.

Os melhores resultados dos experimentos baseados nas amostras para treino e teste a partir de um mesmo participante usando HMMs diagonal e esférico são sumarizados nas Tabelas 5.2 e 5.3, respectivamente. A matriz de covariância completa não é avaliada neste trabalho, pois em experimentos preliminares esta não foi capaz de convergir o processo de classificação a partir das amostras utilizadas. Do mesmo modo, os melhores resultados baseados nas amostras dos Participantes 1, 2, 5 e 6 para o treinamento, e as amostras dos Participantes 3 e 4 para o teste usando HMMs diagonal e esférico estão presentes nas Tabelas 5.4 e 5.5, respectivamente.

Tabela 5.4: Resultados experimentais do HMM com matriz de covariância diagonal variando o número de estados (n_s), baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
3	2	83,61	67,63	0,0 $\bar{3}$	4537	2	86,67	86,00	0,0 $\bar{3}$	5180	2	89,17	75,47	0,0 $\bar{3}$	4939		
4	2	83,33	69,74	0,0 $\bar{3}$	4592	2	88,06	78,01	0,0 $\bar{3}$	4982	2	93,61	88,81	0,0 $\bar{3}$	5473		
Média					4565	Média					5081	Média					5206

Tabela 5.5: Resultados experimentais do HMM com matriz de covariância esférica variando o número de estados (n_s), baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_s	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
3	2	81,20	64,25	0,0 $\bar{3}$	4363	2	84,47	82,47	0,0 $\bar{3}$	5008	2	83,11	68,16	0,0 $\bar{3}$	4538		
4	2	80,65	64,20	0,0 $\bar{3}$	4346	2	76,84	65,77	0,0 $\bar{3}$	4278	2	93,19	85,71	0,0 $\bar{3}$	5367		
Média					4355	Média					4643	Média					4952

As Tabelas 5.2 e 5.3 apresentam altos valores de \mathcal{A} e \mathcal{P} , e \mathcal{T} com tempo mínimo (0.03s) para HMMs na classificação do BR, HF e TS com amostras de treino e teste de um mesmo participante. Entretanto, o HMM diagonal apresentou um IDT médio numericamente superior para cada comportamento estereotipado. O IDT médio dos participante com HMM diagonal para o reconhecimento do BR, HF e TS foram 5680%/s, 5607%/s e 5441%/s, respectivamente. Enquanto que o HMM esférico foram 5535%/s, 5530%/s e 5427%/s. O HMM diagonal também superou o HMM esférico em relação aos resultados dos participantes individualmente. Para o BR, o HMM diagonal com 2 estados obteve um IDT de 5950%/squ (98,33% \mathcal{A} ; 100% \mathcal{P} ; 0,0 $\bar{3}$ s \mathcal{T}) através do Participante 1. O HMM diagonal com 7 e 6 estados também obtiveram altos valores de IDT com o Participante 3, resultando em valores do IDT de 5896%/s (98,06% \mathcal{A} ; 98,46% \mathcal{P} ; 0,0 $\bar{3}$ s \mathcal{T}) e 5821%/s (98,33% \mathcal{A} ; 95,69% \mathcal{P} ; 0,0 $\bar{3}$ s \mathcal{T}), respectivamente.

As Tabelas 5.4 e 5.5 mostram que os HMMs baseados na arquitetura proposta neste trabalho tiveram uma diminuição do desempenho, representado pela variável

IDT, quando as amostras dos participantes de treino são diferentes dos participantes de teste. Entretanto, os HMMs com 2 estados foram capaz de manter \mathcal{T} com valor mínimo ($0,0\bar{3}s$). Neste cenário, o HMM diagonal teve melhor desempenho que o HMM esférico. O IDT médio do primeiro HMM para o BR, HF e TS foi $4355\%/s$, $4643\%/s$, e $4952\%/s$, respectivamente. Além disto, o HMM diagonal superou também o HMM esférico em relação aos resultados dos participantes individualmente. O HMM diagonal apresentou melhores desempenho para o BR e TS através do Participante 4, obtendo valores do IDT de $4592\%/s$ ($83,33\%\mathcal{A}$; $69,74\%\mathcal{P}$; $0,0\bar{3}s\mathcal{T}$) e $5473\%/s$ ($93,61\%\mathcal{A}$; $88,81\%\mathcal{P}$; $0,0\bar{3}s\mathcal{T}$), respectivamente. Para o HF, o HMM diagonal através do Participante 3 superou melhor desempenho que o HMM esférico, alcançando um IDT de $5180\%/s$ ($86,67\%\mathcal{A}$; $86,00\%\mathcal{P}$; $0,0\bar{3}s\mathcal{T}$).

5.3.2 Experimentos de classificação com SVM

A SVM do tipo C-SVC foi utilizado para a análise de desempenho em relação ao tipo de *kernel* (radial e polinomial) e o valor do γ . Para cada SVM, 10 experimentos foram executados, iniciando com $\gamma = 1$ e aumentando em uma unidade. Para as SVMs polinomial, o parâmetro grau (d) foi estabelecido com valor igual a 1. Uma breve descrição sobre a SVM é realizada na Seção B.2 do Apêndice B.

Os melhores resultados experimentais baseados nas amostras de treino e teste de um mesmo participante usando SVMs radial e polinomial são sumarizados nas Tabelas 5.6 e 5.7, respectivamente. Os melhores resultados experimentais baseados em amostras dos Participantes 1,2 e 6 para o treino e as amostras dos Participantes 3 e 4 para teste usando SVMs estão presentes nas Tabelas 5.8 e 5.9, respectivamente.

As Tabelas 5.6 e 5.7 revelam altos valores de \mathcal{A} e \mathcal{P} para as SVMs na classificação do BR, HF e TS com amostras de treino e teste de um mesmo participante. Entretanto, os valores do IDT foram baixos devido aos altos valores de \mathcal{T} . A SVM radial superou a SVM polinomial em relação ao valor médio de IDT para o BR ($2257\%/s$) e HF ($3342\%/s$). A SVM polinomial superou a SVM radial para o TS com um IDT de $2788\%/s$. Da mesma forma, os resultados individuais dos participantes com SVM radial mostram melhor desempenho para BR e HF, enquanto a SVM polinomial apresentou melhores resultados para o TS. A SVM radial com

Tabela 5.6: Resultados experimentais da SVM com *kernel* radial variando o valor de γ , baseando nas amostras de treino e teste do mesmo participante (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
1	10	97,82	95,53	0,05	3784	10	98,31	96,91	0,05	4338	10	97,19	98,45	0,05	3748		
2	7	98,35	97,90	0,09	2301	9	98,96	97,99	0,05	3854	8	97,58	97,54	0,10	1953		
3	4	96,42	95,00	0,06	3034	6	96,82	95,84	0,06	3195	3	95,54	96,23	0,06	3010		
4	6	94,84	92,45	0,16	1205	10	96,84	95,75	0,11	1682	4	96,61	97,99	0,10	1868		
5	4	93,99	91,09	0,19	996	10	98,66	98,09	0,08	2450	5	94,05	94,47	0,18	1063		
6	4	96,18	93,81	0,09	2219	10	99,23	99,23	0,04	4531	4	98,04	98,70	0,06	3108		
Média					2257	Média					3342	Média					2458

Tabela 5.7: Resultados experimentais da SVM com *kernel* polinomial variando o valor de γ , baseando nas amostras de treino e teste do mesmo participante (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
1	2	84,24	74,18	0,06	2483	6	90,26	84,03	0,04	3925	6	87,97	93,30	0,05	3576		
2	8	88,16	84,79	0,10	1711	6	84,77	71,85	0,11	1486	9	82,08	88,45	0,14	1235		
3	7	87,58	84,57	0,05	3450	5	82,16	70,08	0,07	2239	8	89,61	92,76	0,05	3905		
4	8	81,91	70,48	0,13	1201	10	79,82	74,13	0,14	1095	2	89,92	93,39	0,08	2217		
5	5	82,98	78,82	0,14	1180	10	86,64	77,87	0,13	1268	7	84,21	89,35	0,14	1261		
6	1	84,24	78,54	0,08	1931	10	91,16	85,30	0,05	3420	7	93,07	95,58	0,04	4535		
Média					1993	Média					2239	Média					2788

$\gamma = 10$ para a classificação do BR e HF resultou em um IDT de 3784%/s (97,82% \mathcal{A} ; 95,53% \mathcal{P} ; 0,05s \mathcal{T}) através do Participante 1, e 4531%/s (99,23% \mathcal{A} ; 99,23% \mathcal{P} ; 0,04s \mathcal{T}) através do Participante 6.

As Tabelas 5.8 e 5.9 mostram que as SVMs baseadas na arquitetura proposta neste trabalho tiveram baixo desempenho, quando consideradas as amostras dos participantes de treino diferente de teste. Os altos valores de \mathcal{T} e a diminuição no \mathcal{A} e \mathcal{P} diminuiram drasticamente os valores do IDT. Em relação ao IDT médio, a SVM radial apresentou melhor desempenho que o SVM polinomial apenas para o HF, sendo que o primeiro teve 47%/s, e o segundo 42%/s para este comportamento estereotipado. Para o TS, a SVM polinomial apresentou valores do IDT médio numericamente superior à SVM radial com 49%/s e 27%/s, respectivamente. Para

Tabela 5.8: Resultados experimentais da SVM com *kernel* radial variando o valor de γ , baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
3	6	82,02	76,77	5,86	27	10	85,50	80,74	3,31	50	5	86,67	79,08	5,66	29		
4	7	84,16	63,51	7,45	20	9	85,30	88,25	3,95	44	5	91,83	91,18	7,55	24		
				Média	23					Média	47					Média	27

Tabela 5.9: Resultados experimentais da SVM com *kernel* polinomial variando o valor de γ , baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	γ	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
3	4	84,88	54,33	4,78	29	8	72,31	82,24	3,30	47	10	84,47	92,99	3,05	58		
4	4	76,38	33,04	6,54	17	6	73,99	91,71	4,43	37	9	83,27	90,53	4,29	40		
				Média	23					Média	42					Média	49

o BR, ambas SVMs apresentaram valores 23%/s do IDT. Para os resultados dos participantes individuais, a SVM polinomial apresentou valores de IDT superiores à SVM radial nos comportamentos estereotipados BR e TS com $\gamma = 4$ e $\gamma = 10$, respectivamente. Os valores do IDT para estes comportamentos estereotipados foram 29%/s (84,88% \mathcal{A} ; 54,33% \mathcal{P} ; 4,78s \mathcal{T}) e 58%/s (84,47% \mathcal{A} ; 92,99% \mathcal{P} ; 3,05s \mathcal{T}), respectivamente. Para o HF, a SVM radial apresentou valor IDT médio superiores que a SVM polinomial com 50%/s (85,50% \mathcal{A} ; 80,74% \mathcal{P} ; 3,31s \mathcal{T}).

5.3.3 Experimentos de classificação com MLP

A análise de desempenho do MLP foi realizada em relação ao número de neurônios em uma única camada ($l = 1$). Para cada modelo de MLP, 20 experimentos foram executados iniciando com 1 neurônio e aumento em uma unidade. Uma descrição sobre o MLP é realizada na Seção B.3 do Apêndice B.

Os melhores resultados experimentais baseados em amostras de treino e de teste de um mesmo participante usando MLP são sumarizados na Tabela 5.10. Os melhores resultados dos experimentos baseados em amostras dos Participantes 1, 2,

Tabela 5.10: Resultados experimentais do MLP variando a quantidade de neurônios ($n_h^{[1]}$) baseando nas amostras de treino e teste do mesmo participante (Part).

Part	$n_h^{[1]}$	Body Rocking				$n_h^{[1]}$	Hand Flapping				$n_h^{[1]}$	Top Spinning			
		\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)
1	20	97,34	95,28	0,03	5779	19	98,27	96,83	0,03	5853	20	96,69	97,89	0,03	5838
2	15	97,50	97,28	0,03	5843	8	98,39	97,33	0,03	5872	16	96,71	96,28	0,03	5790
3	14	96,48	94,91	0,03	5742	19	97,24	95,72	0,03	5789	17	95,95	95,83	0,03	5753
4	18	95,41	92,85	0,03	5648	13	97,19	95,86	0,03	5792	20	97,39	97,42	0,03	5844
5	20	94,05	90,33	0,03	5531	16	98,17	97,36	0,03	5866	16	93,78	93,65	0,03	5623
6	19	97,82	96,55	0,03	5831	14	99,33	98,63	0,03	5939	17	97,99	98,15	0,03	5884
		Média					Média					Média			
		5729					5852					5789			

Tabela 5.11: Resultados experimentais do MLP a quantidade de neurônios ($n_h^{[1]}$), baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	$n_h^{[1]}$	Body Rocking				$n_h^{[1]}$	Hand Flapping				$n_h^{[1]}$	Top Spinning			
		\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)
3	7	81,11	74,67	0,03	4673	2	82,12	88,21	0,03	5110	1	86,68	91,50	0,03	5346
4	9	83,42	60,60	0,03	4321	9	83,28	87,22	0,03	5115	6	89,93	91,36	0,03	5439
		Média					Média					Média			
		4497					5113					5392			

5 e 6 como amostra de treino e os Participantes 3 e 4 como amostras testes usando MLP são sumarizados na Tabela 5.11.

A Tabela 5.10 mostra valores altos de \mathcal{A} e \mathcal{P} , e \mathcal{T} com tempo mínimo ($0,03s$) para o MLP na classificação do BR, HF e TS com amostras de teste e treino de um mesmo participante. O MLP apresentou valores altos para IDT médio com $5729\%/s$, $5852\%/s$, e $5789\%/s$ para classificação do BR, HF e TS, respectivamente. Para o HF, o MLP com 13 neurônios obteve um IDT de $5939\%/s$ ($99,33\%$ \mathcal{A} ; $98,63\%$ \mathcal{P} ; $0,03s$ \mathcal{T}). Para o TS, o MLP com 17 neurônios obteve um IDT de $5884\%/s$ ($97,99\%$ \mathcal{A} ; $98,15\%$ \mathcal{P} ; $0,03s$ \mathcal{T}). Estes valores de IDT foram obtidos através do Participante 6. Para o BR, o MLP com 15 neurônios apresentou um IDT de $5843\%/s$ ($97,50\%$ \mathcal{A} ; $97,28\%$ \mathcal{P} ; $0,03s$ \mathcal{T}) através do Participante 2.

A Tabela 5.11 mostra que o MLP definido neste trabalho apresentou um queda de desempenho quando as amostras dos participantes de treinamento são diferentes dos participantes de teste. O MLP apresentou uma diminuição do desempenho

na classificação do BR devido a diminuição dos valores de \mathcal{A} e \mathcal{P} . Entretanto, o MLP foi capaz de manter valores mínimos para o \mathcal{T} . Os valores do IDT médio para o BR, HF e TS foram 4497%/s, 5113%/s, e 5392%/s, respectivamente. Os resultados individuais dos participantes do MLP foram altos. Para o BR, o LP com 7 neurônios apresentou um IDT 4673%/s (81, 11% \mathcal{A} ; 74, 67% \mathcal{P} ; 0, 03s \mathcal{T}) através do Participante 3. Para o HF, o MLP com 9 neurônios apresentou um IDT de 5115%/s (83, 28% \mathcal{A} ; 87, 22% \mathcal{P} ; 0, 03s \mathcal{T}) através do Participante 4. Para o TS, o MLP com 6 neurônios apresentou um IDT de 5439%/s (89, 93% \mathcal{A} ; 91, 36% \mathcal{P} ; 0, 03s \mathcal{T}) através do Participante 4.

5.3.4 Experimentos de classificação com CNN

A arquitetura da CNN pode apresentar diferentes variáveis que podem influenciar no desempenho do problema de classificação. Deste modo, para restringir o conjunto destas variáveis avaliadas, os experimentos foram baseados na arquitetura proposta por (Rad *et al.*, 2018). Uma breve abordagem sobre a CNN é realizada na Seção B.4 do Apêndice B.

A análise de desempenho da CNN foi realizada em relação a duas variáveis: (i) número de filtros (n_k) e (ii) número de camadas (n_L). No primeiro cenário, 9 experimentos foram executados iniciando com 2 filtros e aumentando em uma unidade em uma CNN de 3 camadas. No segundo cenário, 4 experimentos foram realizados começando com 1 camada e aumentando em uma unidade em um CNN com 4 filtros em cada camada. Em ambos os cenários, as CNNs tiveram o tamanho (k) e preenchimento (ρ) do *kernel* iguais a 9 e 4, respectivamente. O número de filtros na última camada é igual ao dobro da penúltima.

Os melhores resultados experimentais baseados nas amostras de treino e teste com um mesmo participante em relação ao número de filtros e camadas do CNN são sumarizados nas Tabelas 5.12 e 5.13. Os melhores resultados experimentais baseados em amostras dos Participantes 1, 2, 5 e 6 para o treino e dos Participantes 3 e 4 como amostras de teste em relação ao número de filtros e camadas do CNN são sumarizados nas Tabelas 5.15 e 5.14, respectivamente.

Tabela 5.12: Resultados experimentais da CNN variando o número de filtros (n_k) baseando nas amostras de treino e teste do mesmo participante (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	n_k	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_k	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_k	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
1	5	95,97	93,05	0,11	1665	5	96,84	95,24	0,12	1647	5	94,50	95,79	0,12	1648		
2	2	94,09	94,00	0,14	1325	2	96,70	94,99	0,14	1332	3	93,22	92,16	0,14	1327		
3	7	95,00	91,97	0,11	1659	2	93,13	90,05	0,11	1636	2	91,32	91,02	0,11	1632		
4	3	89,92	83,17	0,14	1210	4	94,58	91,40	0,14	1318	3	94,73	94,17	0,14	1308		
5	2	87,41	80,15	0,15	1110	2	94,94	93,67	0,16	1193	3	89,73	89,53	0,15	1186		
6	7	95,31	93,47	0,13	1483	2	96,18	94,38	0,12	1571	2	93,71	93,42	0,12	1575		
Média					1409	Média					1450	Média					1446

Tabela 5.13: Resultados experimentais da CNN variando o número de camadas (n_L) baseando nas amostras de treino e teste do mesmo participante (Part).

Part	Body Rocking				Hand Flapping				Top Spinning								
	n_L	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_L	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_L	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
1	1	95,05	90,08	0,07	2776	1	95,94	93,75	0,06	3079	1	93,82	95,84	0,07	2887		
2	1	95,28	93,05	0,09	1997	1	97,65	96,78	0,08	2490	1	93,96	94,07	0,08	2225		
3	1	93,77	90,43	0,08	2329	1	93,72	89,22	0,06	3054	1	93,83	94,46	0,06	3411		
4	1	90,98	84,21	0,09	2004	1	93,70	90,24	0,08	2323	1	95,20	95,13	0,08	2358		
5	1	89,03	82,21	0,10	1787	1	96,35	94,41	0,09	2092	1	89,86	92,35	0,08	2407		
6	2	93,76	90,59	0,10	1807	1	96,80	94,82	0,07	2793	1	95,96	96,29	0,07	2659		
Média					2117	Média					2638	Média					2658

As Tabelas 5.12 e 5.13 mostram altos valores de \mathcal{A} e \mathcal{P} na CNN baseada na arquitetura deste trabalho. Contudo, os valores de \mathcal{T} foram altos, o que impactou na diminuição dos valores de IDT. As CNNs variando o número de camadas tiveram alto valor no IDT médio comparado com a CNN variando de filtros. Os valores do IDT médio do BR, HF e TS para a CNN variando o número de camadas foram 2117%/s, 2638%/s e 2658%/s, enquanto que a CNN variando o número de filtros foram 1409%/s, 1450%/s, e 1446%/s, respectivamente. De maneira similar, resultados de participantes individuais com a CNN variando o número de camadas foram superiores à CNN variando o número de filtros. As CNNs com uma única camada ($n_L = 1$) apresentaram melhor IDT médio para BR e HF através do Participante 1, com valores de 2776%/s (95,05% \mathcal{A} ; 90,08% \mathcal{P} ; 0,07s \mathcal{T}) e 3079%/s (95,94% \mathcal{A} ;

Tabela 5.14: Resultados experimentais da CNN variando o número de filtros (n_k), baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	Body Rocking					Hand Flapping					Top Spinning						
	n_k	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT	n_k	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT	n_k	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT		
3	4	82,03	71,40	1,01	152	5	81,13	88,98	1,02	167	2	84,16	77,20	1,07	151		
4	4	80,68	65,51	1,39	105	6	79,03	94,25	1,06	164	4	86,91	76,98	1,40	117		
Média					129	Média					166	Média					134

Tabela 5.15: Resultados experimentais da CNN variando o número de camadas (n_L), baseando nas amostras de treino e teste de diferentes participantes (Part).

Part	Body Rocking					Hand Flapping					Top Spinning						
	n_L	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_L	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)	n_L	\mathcal{A} (%)	\mathcal{P} (%)	\mathcal{T} (s)	IDT (%/s)		
3	1	78,83	64,53	0,63	229	1	81,69	82,07	0,56	293	1	81,00	70,36	0,63	241		
4	1	80,37	66,81	0,83	178	1	80,25	74,94	0,77	202	1	86,61	77,69	0,75	219		
Média					203	Média					248	Média					230

93,75% \mathcal{P} ; 0,06s \mathcal{T}), respectivamente. Para o TS, a CNN com os mesmo número de camadas que o anterior através do Participante 3 teve melhor IDT médio com 3411%/s (93,83% \mathcal{A} ; 94,46% \mathcal{P} ; 0,06s \mathcal{T}).

As Tabelas 5.14 e 5.15 revelam que as CNNs baseadas na arquitetura proposta neste trabalho apresentaram uma queda acentuada nos valores de \mathcal{A} e \mathcal{P} , quando as amostras dos participantes de treinamento são diferentes de teste. Além disto, os valores de \mathcal{P} foram altos, causando um IDT baixo. A CNN variando o número de camadas mostrou melhores valores para o IDT médio para o BR, HF e TS. Para estes comportamentos estereotipados, a primeira CNN teve valores do IDT médio de 203%/s, 248%/s e 230%/s, respectivamente, enquanto a segunda CNN teve IDT médio de 129%/s, 166%/s e 134%/s, respectivamente. Os resultados dos participantes individuais com CNN variando o número de camadas também foram melhores que a CNN variando o número de filtros. A primeira CNN com uma única camada apresentou um IDT de 229%/s (78,83% \mathcal{A} ; 64,53% \mathcal{P} ; 0,63s \mathcal{T}); 293%/s (81,69% \mathcal{A} ; 82,07% \mathcal{P} ; 0,56s \mathcal{T}); e 241%/s (81% \mathcal{A} ; 70,36% \mathcal{P} ; 0,63s \mathcal{T}) através do Participante 3 para BR, HF, e TS, respectivamente.

Tabela 5.16: Sumário dos melhores resultados experimentais dos AAM baseados nas amostras dos mesmos participantes para treino e teste.

AAM	BR	HF	TS
	IDT médio (%/s)	IDT médio (%/s)	IDT médio (%/s)
HMM diagonal	5680	5607	5441
HMM esférico	5535	5530	5427
SVM radial	2257	3342	2458
SVM polinomial	1993	2239	2788
MLP	5729	5852	5789
CNN variando a quantidade de filtros	1409	1450	1446
CNN variando a quantidade de camadas	2117	2638	2658

5.3.5 Discussão sobre os resultados experimentais dos classificadores

As Tabela 5.16 e 5.17 resumam os maiores valores de IDT médio dos AAM para as amostras de mesmo e diferente participantes, respectivamente. Os MLPs e HMMs foram capazes de manter um valor mínimo de \mathcal{T} (0,03s) para ambas formas de organização das amostras. Estes dois tipos de AAM obtiveram maiores valores de IDT médio que SVMs e CNNs.

Para estes AAMs, os valores de \mathcal{A} e \mathcal{P} diminuíram nas amostras de participantes de treino diferentes que os de teste. Estes resultados eram esperados pois há a influência da idiosincrasia entre as amostras dos participantes de treino e teste. De um modo geral, os valores dos IDT médio do MLP superaram numericamente os HMMs. O HMM diagonal apresentou um IDT médio (4565%/s) para o BR nas amostras de participantes de treino e teste.

As SVMs e CNNs tiveram valores de \mathcal{T} maiores que os MLP e HMMs, devido a isto os valores dos IDT médios foram baixos. Além disto, os dois primeiros AAMs

Tabela 5.17: Sumário dos melhores resultados experimentais dos AAM baseados nas amostras dos participantes de treino diferente de teste.

MLA	BR	HF	TS
	IDT médio	IDT médio	IDT médio
	(%/s)	(%/s)	(%/s)
HMM diagonal	4565	5081	5206
HMM esférico	4355	4643	4952
SVM radial	23	47	27
SVM polinomial	23	42	49
MLP	4497	5113	5392
CNN-filters	129	166	134
CNN-layers	203	248	230

tiveram uma queda acentuada dos valores de IDT médio para no cenário em que foram utilizadas amostras de participantes diferentes na etapa de treino e de teste, devido aos altos valores de \mathcal{A} e \mathcal{P} . Esta queda acentuada dos IDTs tanto das SVMs como das CNNs demonstram que elas não foram capazes de generalizar a classificação com as amostras de diferentes participantes, sendo assim, sensíveis a idiosincrasia dos participantes.

A partir dos resultados observados, as abordagens baseadas nos HMMs e MLP apresentaram melhor desempenho. Embora não concerne neste trabalho a implementação final do CREC, é importante destacar as orientações deste processo. Assim, neste passo seguinte há necessidade de escolher quais destes AAM devem ser utilizados para classificação dos comportamentos estereotipados. As abordagens observadas neste trabalho para serem aplicadas ao CRCE no processo classificação dos comportamentos estereotipados são:

1. Aplicação do HMM;
2. Aplicação o MLP;
3. Aplicar a combinação AAM usando estratégia de Aprendizado por Votos Majoritários (AVM);

As duas primeiras abordagens têm a desvantagem de considerar apenas o desempenho de um dos AAM. Portanto, a terceira abordagem pode ser mais apropriada, por possibilitar a agregação das características de Algoritmos de Aprendizado de Máquina analisados neste trabalho para o processo de classificação dos comportamentos estereotipados. O AVM combina classificadores que são treinados separadamente. Geralmente, cada classificador deve resultar em dois valores para uma determinada classe: aceitar ou rejeitar. Deste modo, quando a combinação de mais da metade destes valores, denominado de votos, dos classificadores significa que ocorreu uma aceitação da amostra em relação a uma determinada classe, caso contrário significa que tal amostra foi rejeitada (Catal e Nangir, 2017). A implementação do AVM não faz parte do escopo deste trabalho.

5.4 Experimentos de Inferência do Nível de Ativação

Os detalhes experimentais para inferência do nível de ativação são apresentados nesta seção. Neste trabalho, é utilizado um sistema baseado em regras nebulosas para inferir o nível de ativação dos comportamentos estereotipados típicos de autistas. Assim, uma análise de desempenho é realizada com o ANFIS. Esta análise é dividida em duas etapas: (i) desempenho de treino, e (ii) desempenho de teste. O ANFIS é descrito na Seção B.5 do Apêndice B.

5.4.1 Desempenho de treino do ANFIS

A primeira etapa visa determinar a abordagem de treinamento que apresenta o menor erro quadrático médio em relação ao número de épocas. Deste modo, os algoritmos apresentam menor erro, significa que seu treinamento foi capaz de discriminar as classes com uma determinada quantidade de épocas.

Da mesma maneira que foi definido nos experimentos de classificação na seção anterior, os valores do grupo de QoM dos Participantes 1, 2, 3 e 6 foram definidos como amostras de treino do ANFIS. Os algoritmos de treino do ANFIS considerados

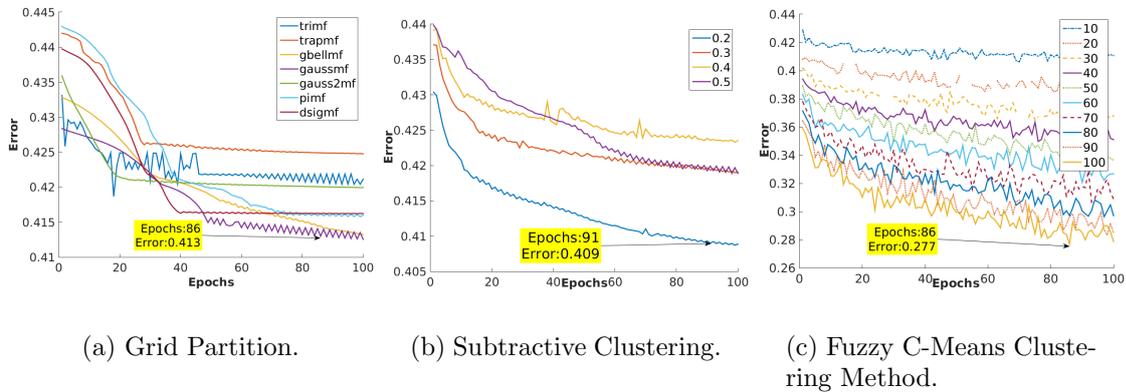


Figura 5.9: Erro quadrático médio do algoritmo de treino do ANFIS em relação ao número de épocas.

são: *Grid Partition* (GP), *Subtractive Clustering* (SB), e *Fuzzy C-Means Clustering Method* (FCM). Para o treinamento, os algoritmos citados assumiram valores máximos para épocas iguais a 100. Os resultados para GP, SC e FCM são ilustrados na Figura 5.9

Os resultados do treino do algoritmo GP são ilustrados na Figura 5.9(a). Para este algoritmo de treino, as funções de pertinências são definidas sobre os valores dos QoMs. Apenas um tipo de função de pertinência foi aplicada na entrada durante cada experimento. As funções pertinências consideradas foram: triangular (*trimf*), trapezoidal (*trapmf*), sino generalizado (*gbellmf*), gaussiana (*gaussmf*), gaussiana com parâmetros combinados (*gaussmf2*), curva baseada em *spline* (*pimf*), e sigmoidal (*dsigmf*). As funções pertinências gaussianas apresentaram menor erro quadrático médio de 0,413 a partir de 86 épocas.

Os resultados de treinamento do algoritmo SC são ilustrados na Figura 5.9(b). Os experimentos foram realizados com raio de influência dos grupos com os seguintes valores: 0,2, 0,3, 0,4, e 0,5. O raio de vizinhança igual a 0,2 apresentou o menor erro quadrático médio com valor de 0,409 a partir de 91 épocas.

Os resultados do treinamento do algoritmo FCM são ilustrados na Figura 5.9(c). Diferentes quantidades de grupos foram considerados nestes experimentos, iniciando de 10 grupos e finalizando com 100 (um passo de tamanho 10). A partir de quantidade superior a 40 grupos, o erro quadrático médio decaiu para abaixo de 0,4. O

menor valor de erro quadrático médio foi de 0,277 obtido com 100 grupos e 86 épocas. O FCM apresentou menor erro do quadrático médio em relação aos algoritmos GP e SC (ver Figura 5.9)).

5.4.2 Desempenho de teste do ANFIS

A segunda etapa dos experimentos visa analisar o desempenho do ANFIS para inferir o nível de ativação. O estudo é baseado nos parâmetros de Distribuição Normal [N (mean, variance)] das saídas do ANFIS para cada nível de ativação esperada dos comportamentos estereotipados. Os valores de referência para alto e baixo nível de ativação são 1 e 0, respectivamente.

Na Subseção 5.4.1, foi observado que o algoritmo de treinamento FCM com 86 épocas apresentou menor erro quadrático (0,277). Os resultados dos testes apresentados nesta seção são baseados no modelo de treinamento do FCM – ANFIS definidos na subseção anterior. A Figura 5.10 ilustra os resultados dos testes para os Participantes 3 e 4.

Na Figura 5.10(a), pode-se observar que os níveis de ativação baixo e alto do comportamento estereotipado BR para o Participante 3 são descritos pela distribuição normal $N_L(\mu_L = 0,58, \sigma_L = 0,17)$ e $N_H(\mu_H = 0,71, \sigma_H = 0,24)$, respectivamente. Para o Participante 4 (ver Figura 5.10(b)), a distribuição normal de baixo e alto nível de ativação são $N_L(\mu_L = 0,58, \sigma_L = 0,11)$ e $N_H(\mu_H = 0,82, \sigma_H = 0,13)$, respectivamente.

Na Figura 5.10(c), observa-se que os níveis de ativação baixo e alto do comportamento estereotipado HF para o Participante 3 são descritos pela distribuição normal $N_L(\mu_L = 0,45, \sigma_L = 0,10)$ e $N_H(\mu_H = 0,69, \sigma_H = 0,20)$, respectivamente. Para o Participante 4 (ver Figura 5.10(d)), a distribuição normal de baixo e alto nível de ativação são $N_L(\mu_L = 0,39, \sigma_L = 0,07)$ e $N_H(\mu_H = 0,6, \sigma_H = 0,14)$, respectivamente.

Na Figura 5.10(e), é possível observar que os níveis de ativação baixo e alto do comportamento estereotipado TS para o Participante 3 são descritos pela distribuição normal $N_L(\mu_L = 0,44, \sigma_L = 0,11)$ e $N_H(\mu_H = 0,67, \sigma_H = 0,21)$, respectiva-

mente. Para o Participante 4 (ver Figura 5.10(f)), a distribuição normal de baixo e alto nível de ativação são $N_L(\mu_L = 0,41, \sigma_L = 0,13)$ e $N_H(\mu_H = 0,65, \sigma_H = 0,23)$, respectivamente.

5.4.3 Discussão sobre os resultados experimentais do ANFIS

Os experimentos realizados com ANFIS visaram determinar a melhor configuração para inferir o nível de ativação dos comportamentos estereotipados. Estes experimentos foram organizados em duas etapas.

A primeira etapa visa determinar o algoritmo de treinamento do ANFIS (*Grid Partition* – GP, *Subtractive Clustering* – SC e *Fuzzy C-Means Clustering Method* – FCM) que apresenta o menor erro quadrático médio. O FCM apresentou o menor erro com 0,277 com 86 épocas.

Na segunda etapa, o modelo do ANFIS que apresentou melhor desempenho para o treino (FCM) foi aplicado para inferir o nível de ativação cada um dos comportamentos estereotipados. As amostras para os testes são os valores dos grupos de QoM dos Participantes 3 e 4, sendo que o primeiro é do gênero masculino e o segundo feminino. O uso de amostras femininas visam verificar a capacidade dos modelos de generalizar independentemente do gênero.

Para a análise de desempenho de teste do ANFIS, a distância entre os valores médios de alto e baixo nível de ativação de cada comportamento estereotipado determina a capacidade de diferenciar os níveis de ativação (alto e baixo), seguindo a notação de regras nebulosas. Distâncias maiores dos valores médios das saídas do ANFIS significa maior capacidade de diferenciação dos níveis de ativação. Por outro lado, menor distância entre estes valores significa, que o ANFIS apresenta menor capacidade de diferenciar o níveis de ativação.

Para o comportamento estereotipado BR, as diferenças dos níveis de ativação para os Participantes 3 e 4 foram 0,13 e 0,24, respectivamente. Para o comportamento estereotipado HF, as diferenças dos níveis de ativação para estes participantes foram 0,24 e 0,21, respectivamente. Para o comportamento estereotipado TS, as

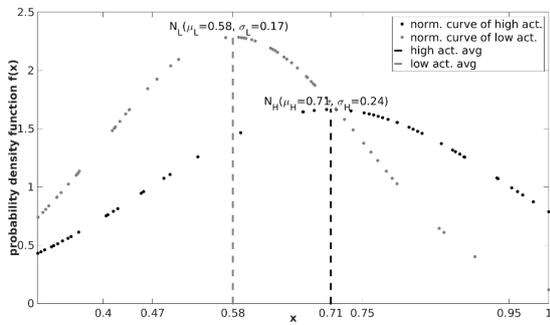
diferenças dos níveis de ativação para os referidos participantes foram 0, 23 e 0, 24, respectivamente. A partir destes resultados, é possível observar que o FCN ANFIS foi capaz de diferenciar os comportamentos estereotipados BR, HF e TS com as amostras do Participante 4 (para BR e TS) e 3 (para HF).

5.5 Considerações Finais

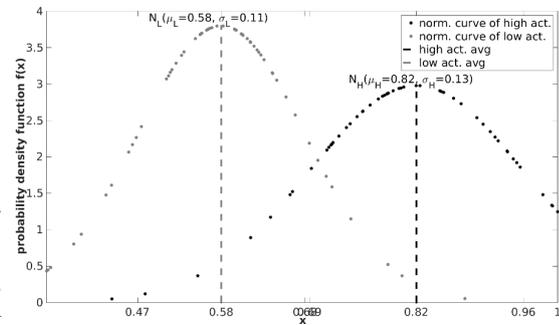
Neste capítulo, foram descritos estudos sobre o desempenho das abordagens para reconhecimento do comportamento estereotipado e inferência do nível de ativação destes comportamentos. Estes estudos foram dividido em duas partes: (i) classificação de comportamentos estereotipados, e (ii) inferência do nível de ativação.

Na primeira parte, observou-se que tanto o HMM como MLP apresentaram desempenho superior em relação aos demais AAM (SVM e CNN). Para auxiliar esta análise, foi utilizado um variável denominada de Índice de Desempenho Temporal (IDT). Esta variável é definida pelas variáveis estatísticas precisão, acurácia e tempo de teste (ver eq (5.4)). Em relação aos trabalhos relacionados (ver Tabela 2.1), considerando as variáveis acurácia e precisão, os experimentos com HMM e MLP baseadas na arquitetura proposta neste trabalho apresentaram um desempenho superior.

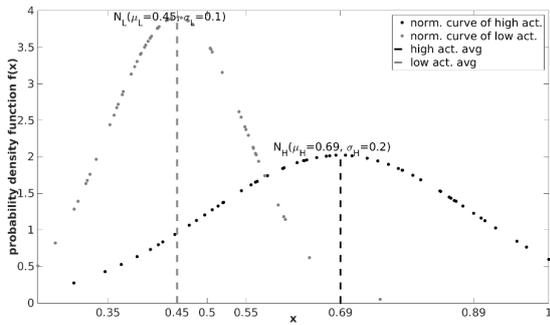
Na segunda parte, o processo de inferência do nível de ativação foi baseado em modelos do ANFIS. Observou-se que o FCM apresentou melhor desempenho para o treinamento. Além disto, na fase de teste, o ANFIS baseado no FCM, foi capaz de diferenciar adequadamente os níveis de ativação dos comportamentos estereotipados.



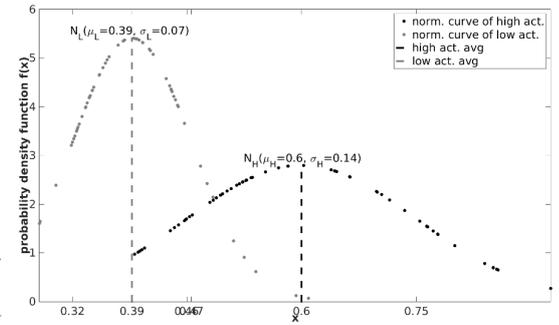
(a) Body Rocking do Participante 3.



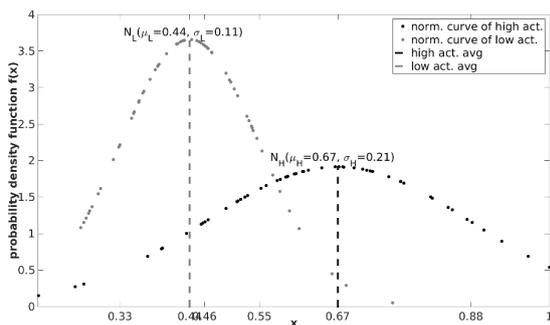
(b) Body Rocking do Participante 4.



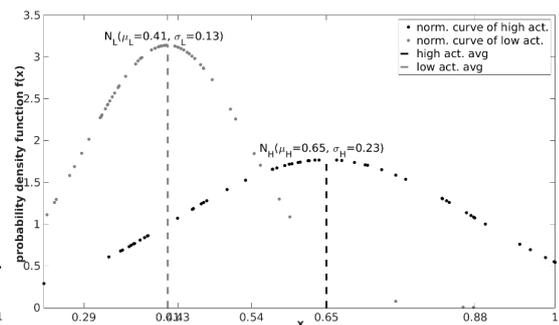
(c) Hand Flapping do Participante 3.



(d) Hand Flapping do Participante 4.



(e) Top Spinning do Participante 3.



(f) Top Spinning do Participante 4.

Figura 5.10: Teste do modelo FCM ANFIS para os Participantes 3 e 4.

Capítulo 6

Conclusão

Este trabalho faz parte do projeto HiBot. O HiBot é uma arquitetura robótica em desenvolvimento para experimentos de Interação Humano Robô (IHR). Especificamente, este trabalho trata de um dos módulos afetivos denominado Módulo de Reconhecimento de Expressões Corporais (MREC) (ver Figura 1.1).

O MREC deve ser capaz de reconhecer o estado afetivo (nível de ativação) de indivíduos autistas a partir dos seus comportamentos estereotipados. A detecção e quantificação destes fenômenos podem ser importantes para o auxílio no processo terapêutico de indivíduos autistas.

Um estudo analítico das ferramentas de inteligência computacional torna-se importante para determinação da ferramenta mais adequada para o aplicação no ambiente clínico real. Dessarte, experimentos foram realizados baseando-se no uso de estímulos corporais não-espontâneos através do uso de 6 participantes voluntários não autistas. Assim, estes experimentos foram divididos em duas etapas visando avaliar o desempenho de algoritmos de inteligência computacional de maneira isolada.

A primeira etapa teve como objetivo avaliar os desempenhos dos algoritmos de HMM, SVM, MLP e CNN para classificação dos comportamentos estereotipados (BR, HF e TS). Além da precisão e acurácia, a análise de desempenho neste trabalho considerou também o tempo de teste para classificação. A relação destas três variáveis definida pela eq. 5.4 é denominada Índice de Desempenho Temporal (IDT).

Através do IDT, é possível determinar os melhores resultados considerando também a aplicação no processo interativo em ambiente clínico real através do tempo de classificação.

Os experimentos dessa etapa foram baseados em duas estratégias de organização de amostras de treino e de teste. A primeira estratégia consiste em avaliar o desempenho dos AAMs em relação a amostra de cada participante. A segunda estratégia visa utilizar um grupo de participantes para o treino e outro grupo para o teste. Os resultados de ambas abordagens demonstram que as ferramentas HMM e MLP apresentaram altos IDT. O alto desempenho destas duas ferramentas se deve aos altos valores de acurácia e precisão e baixo tempo de classificação em relação a arquitetura proposta neste trabalho. Além disto, os resultados da segunda estratégia demonstram que tais algoritmos foram capazes de generalizar em relação a amostras de diferentes do treino. Em ambas estratégias, demonstram que os valores de IDT do SVM e CNN foram baixos. Estes valores foram influenciados principalmente pelo tempo longo para classificação, como também acurácia e precisão relativamente baixos. Do mesmo modo, foi possível perceber a partir da segunda estratégia que estes dois tipos de AAM não foram capazes de generalizar o processo de classificação com participantes diferentes do treino.

Na segunda etapa, realizou-se experimentos de algoritmos do ANFIS para inferir o nível de ativação a partir de determinados comportamentos estereotipados. O estudo de desempenho foi feito em relação aos métodos *Grid Partition* (GP), *Subtractive Clustering* (SB), e *Fuzzy C-Means Clustering Method* (FCM). As amostras dos experimentos nessa etapa foram organizadas dividindo os dados provenientes de um grupo de participantes para o treino, e de outro de participantes para o teste.

No treino, o FCM apresentou menor erro (ver Figura 5.9). Deste modo, para o teste foi utilizado o FCM com os parâmetros definidos na etapa de treino. Durante a fase de teste, pode-se observar que o FCM foi capaz de diferenciar os níveis de ativação (alto e baixo) dos comportamentos estereotipados considerados neste trabalho (BR, HF e TS).

6.1 Trabalhos Futuros

Este trabalho realizou o estudo analítico dos algoritmos de inteligência computacional. Etapas seguintes provenientes deste trabalho são previstas. As principais etapas são:

1. Estudo de mapas de cores alternativas ao RGB;
2. Combinação de AAM para classificação de comportamentos estereotipados;
3. Aplicação da abordagem Detecção de Novidades;
4. Validação com autistas reais.

A primeira etapa consiste verificar se a atribuição de mapas de cores (modelos de cores ou espaço de cores) alternativas pode influenciar no processamento das imagens captadas pelo sensor câmera. Atualmente, é utilizado o espaço de cores RGB.

A segunda etapa visa determinar estudar uma estratégia de combinação de classificadores para classificação de comportamentos estereotipados. A estratégia por votos majoritários (Hassan e Abdel-Qader, 2015) tem apresentado resultados promissores na combinação de classificadores. Um desafio presente nesta abordagem consiste na combinação de classificadores capazes maximizar a generalização desta combinação. Uma determinada classe w_j será escolhida de acordo com a regra do voto majoritário, se

$$\sum_{i=1}^N d_{i,j} = \max_{j=1}^M \sum_{i=1}^N d_{i,j}, \quad (6.1)$$

onde N é tamanho de um conjunto de classificadores, M é o número de classes e $d_{imj} \in \{0, 1\}$ é a decisão do i -ésimo classificador para j -ésima classe.

A terceira etapa consiste em implementar a abordagem de Detecção de Novidades (Ouafae *et al.*, 2020). A Detecção de Novidades tem ganhado atenção dos pesquisadores em domínio de aplicações que envolve banco de dados grandes que

foram adquiridos de sistemas críticos. Na condição do entendimento parcial de complexos sistemas modernos, surge uma grande quantidade de possíveis “anormais”, alguns dos quais podem não ser conhecidos a priori, o que torna os esquemas de classificação multi-classe convencionais inadequados. Assim, a Detecção de Novidade basicamente consiste em aprender uma descrição de normalidade através de instancias positivas. Deste modo, a aplicação desta abordagem visa tratar os movimentos corporais que não são os comportamentos estereotipados considerados neste trabalho.

A quarta etapa prevista é a validação do processo interativo baseado em protocolos sociais do HiBot com autistas reais. Nesta etapa, há necessidade de elaborar um plano de interação deste dispositivo com autistas. Para isto, é necessário definir o espectro do autismo que concernirá para estas avaliações. Ademais, é necessário submeter esse plano, para avaliação e autorização prévia do Comitê de Ética em Pesquisa (CEP).

Referências Bibliográficas

- Ahmed, F., A. S. M. H. Bari e M. L. Gavrilova (2020). **Emotion recognition from body movement**. *IEEE Access* **8**, 11761–11781.
- Albinali, F., M. S. Goodwin e S. S. Intille (2009). **Recognizing stereotypical motor movements in the laboratory and classroom**. Em: *Proceedings of the 11th international conference on Ubiquitous computing - Ubicomp '09*. ACM Press. p. 71.
- APA, American Psychiatric Association (2013). ***Diagnostic and statistical manual of mental disorders***. 5 ed.. American Psychiatric Association. Washington, DC, USA.
- Atkinson, A. P., M. L. Tunstall e W. H. Dittrich (2007). **Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures**. *Cognition* **104**(1), 59–72.
- Benmouiza, Khalil e Ali Cheknane (2019). **Clustered anfis network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting**. *Theoretical and Applied Climatology* **137**(1-2), 31–43.
- Bernhardt, D. e P. Robinson (2009). **Detecting emotions from connected action sequences**. *Visual Informatics: Bridging Research and Practice Lecture Notes in Computer Science* **5857**, 1–11.
- Black, K. R., R. A. Stevenson, M. Segers, B. L. Ncube, S. Z. Sun, A. Philipp-Muller, J. M. Bebko, M. D. Barense e S. Ferber (2017). **Linking anxiety and insis-**

- tence on sameness in autistic children: The role of sensory hypersensitivity.** *Journal of Autism and Developmental Disorders* **47**(8), 2459–2470.
- Boruah, D., K. K. Sarma e A. K. Talukdar (2015). **Different face regions detection based facial expression recognition.** 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN) pp. 459–464.
- Breazeal, C. (2004). **Function meets style: Insights from emotion theory applied to hri.** *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* **34**(2), 187–194.
- Camurri, A., I. Lagerlöf e G. Volpe (2003). **Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques.** *International Journal of Human Computer Studies* **59**(1-2), 213–225.
- Castellano, G., I. Leite, A. Pereira, C. Martinho, A. Paiva e P. W. McOwan (2010). **Affect recognition for interactive companions: challenges and design in real world scenarios.** *Journal on Multimodal User Interfaces* **3**(1 – 2), 89–98.
- Catal, C. e M. Nangir (2017). **A sentiment classification model based on multiple classifiers.** *Applied Soft Computing* **50**, 135–141.
- Cheok, M. J., Z. Omar e M. H. Jaward (2019). **A review of hand gesture and sign language recognition techniques.** *International Journal of Machine Learning and Cybernetics* **10**(1), 131–153.
- Colby, B. N., A. Ortony, G. L. Clore e A. Collins (1989). *The Cognitive Structure of Emotions*. Vol. 18. Cambridge University Press.
- Coronato, A., G. De Pietro e G. Paragliola (2014). **A situation-aware system for the detection of motion disorders of patients with autism spectrum disorders.** *Expert Systems with Applications* **41**(17), 7868–7877.
- Cortes, C. e V. Vapnik (1995). **Support-vector networks.** *Machine Learning* **20**(3), 273–297.

- Cunningham, A. B. e L. Schreibman (2008). **Stereotypy in autism: The importance of function**. *Research in Autism Spectrum Disorders* **2**(3), 469–479.
- den Stock, J. V., R. Righart e B. Gelder (2007). **Body expressions influence recognition of emotions in the face and voice**. *Emotion* **7**(3), 487–494.
- DiGennaro Reed, F. D., J. M. Hirst e S. R. Hyman (2012). **Assessment and treatment of stereotypic behavior in children with autism and other developmental disabilities: A thirty year review**. *Research in Autism Spectrum Disorders* **6**(1), 422–430.
- Ekman, P. (1992). **Are there basic emotions?**. *Psychological Review* **99**(3), 550–553.
- Ekman, P. e W. V. Friesen (1974). **Detecting deception from the body or face**. *Journal of Personality and Social Psychology* **29**(3), 288–298.
- Farhan, H. R., M. H. Al-Muifraje e T. R. Saeed (2016). **Using only two states of discrete hmm for high-speed face recognition**. Em: *2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*. number 5. IEEE. pp. 1–5.
- Flowers, J., J. Lantz, T. Hamlin e R. J. Simeonsson (2020). **Associated factors of self-injury among adolescents with autism spectrum disorder in a community and residential treatment setting**. *Journal of Autism and Developmental Disorders*.
- Foundation, BeagleBoard.org (2020). **What is beagleboard-xm?**. <https://beagleboard.org/beagleboard-xm>. Acessado Junho 23, 2020.
- Gaffary, Y., J. C. Martin e M. Ammi (2018). **Haptic expression and perception of spontaneous stress**. *IEEE Transactions on Affective Computing* **X**(X), 1–1.
- Galagedarage Don, M. e F. Khan (2019). **Dynamic process fault detection and diagnosis based on a combined approach of hidden markov and bayesian network model**. *Chemical Engineering Science* **201**, 82–96.

- Giese, C. L. R., L. Omlor e A. Martin (2008). **Lateral asymmetry of bodily emotion expression**. *Current Biology* **18**, 329–330.
- Glowinski, D., N. Dael, A. Camurri, G. Volpe, M. Mortillaro e K. Scherer (2011). **Toward a minimal representation of affective gestures**. *IEEE Transactions on Affective Computing* **2**(2), 106–118.
- Gonçalves, N., J. L. Rodrigues, S. Costa e F. Soaresa (2012). **Automatic detection of stereotyped hand flapping movements: Two different approaches**. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* pp. 392–397.
- Goodwin, M. S., S. S. Intille, F. Albinali e W. F. Velicer (2011). **Automated detection of stereotypical motor movements**. *Journal of Autism and Developmental Disorders* **41**(6), 770–782.
- Gross, M. M., E. A. Crane e B. L. Fredrickson (2010). **Methodology for assessing bodily expression of emotion**. *Journal of Nonverbal Behavior* **34**(4), 223–248.
- Hartmann, M., A. Mavrolampados, E. Allingham, E. Carlson, B. Burger e P. Toiviainen (2019). **Kinematics of perceived dyadic coordination in dance**. *Scientific Reports* **9**(1), 15594.
- Hassan, M. F. e I. Abdel-Qader (2015). **Performance analysis of majority vote combiner for multiple classifier systems**. Em: *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. IEEE. pp. 89–95.
- Hooker, J. L., D. Dow, L. Morgan, C. Schatschneider e A. M. Wetherby (2019). **Psychometric analysis of the repetitive behavior scale revised using confirmatory factor analysis in children with autism**. *Autism Research* **12**(9), 1399–1410.
- Hudson, M. R., E. Sokolenko and T. J. O'Brien e N. C. Jones (2020). **NMDA receptors on parvalbumin-positive interneurons and pyramidal neu-**

- rons both contribute to MK-801 induced gamma oscillatory disturbances: Complex relationships with behaviour. *Neurobiology of Disease* **134**, 104625.
- Hueber, T. e G. Bailly (2016). **Statistical conversion of silent articulation into audible speech using full-covariance HMM**. *Computer Speech & Language* **36**, 274–293.
- Imeh-Nathaniel, A., N. Rincon, V. B. Orfanakos, L. Leanne, L. Wormack, E. Richardson, R. Huber e T. I. Nathaniel (2017). **Effects of chronic cocaine, morphine and methamphetamine on the mobility, immobility and stereotyped behaviors in crayfish**. *Behavioural Brain Research* **332**(April), 120–125.
- Izard, C. E. (1994). **Innate and universal facial expressions: Evidence from developmental and cross-cultural research**. *Psychological Bulletin* **115**(2), 288–299.
- Jang, J. S. R. (1993). **Anfis: Adaptive-network-based fuzzy inference system**. *IEEE Transactions on Systems, Man and Cybernetics* **23**(3), 665–685.
- Johnsen, T. K., M. S. Marley e V. C. Gulick (2020). **A Multilayer Perceptron for Obtaining Quick Parameter Estimations of Cool Exoplanets from Geometric Albedo Spectra**. *Publications of the Astronomical Society of the Pacific* **132**(1010), 044502.
- Juang, B. H. e L. R. Rabiner (1991). **Hidden markov models for speech recognition**. *Technometrics* **33**, 251–272.
- Kamal, M. S., L. Chowdhur, M. I. Khan, A. S. Ashour, J. M. R. S. Tavares e N. Dey (2017). **Hidden markov model and chapman kolmogrov for protein structures prediction from images**. *Computational Biology and Chemistry* **68**, 231–244.
- Kapoor, A. e A. Singhal (2017). **A comparative study of k-means, k-means++ and fuzzy c-means clustering algorithms**. Em: *2017 3rd Internatio-*

- nal Conference on Computational Intelligence & Communication Technology (CICT)*. IEEE. pp. 1–6.
- Karg, M., A. A. Samadani, R. Rob, K. Kolja, J. Hoey e D. Kulic (2013). **Body movements for affective expression: A survey of automatic recognition and generation**. *IEEE Transactions on Affective Computing* **4**(4), 341–359.
- Kirsty, R., H. Andrew, M. C. Meriel e H. Catherine (2015). **Cognitive differences in horses performing locomotor versus oral stereotypic behaviour**. *Applied Animal Behaviour Science* **168**, 37–44.
- Kleinsmith, A. e N. Bianchi-Berthouze (2013). **Affective body expression perception and recognition: A survey**. *IEEE Transactions on Affective Computing* **4**(1), 15–33.
- Kounine, L. (2017). **Emotions, Mind, and Body on Trial: A Cross-Cultural Perspective**. *Journal of Social History* **51**(2), 219–230.
- Kuhn, J. (1999). **Stereotypic behavior as a defense mechanism in autism**. *Harvard Brain Special* **6**(1), 11–15.
- Kulic, D. e E. A. Croft (2007). **Affective state estimation for human-robot interaction**. *IEEE Transactions on Robotics* **23**(5), 991–1000.
- Kushki, A., A. Khan, J. Brian e E. Anagnostou (2014). **A Kalman filtering framework for physiological detection of anxiety in children with autism spectrum disorder**. *IEEE Transactions on Biomedical Engineering* **62**(3), 990–1000.
- Kyriakopoulos, M. (2020). **Autism spectrum disorders and psychosis**. Em: *Autism 360°*. pp. 161–173. Elsevier.
- Laban, R. e F. C. Lawrence (1974). ***Effort: Economy of Human Movement***. 2 ed.. Macdonald & Evans.
- Lassalle, A., D. Pigat, H. O’Reilly, S. Berggen, S. Fridenson-Hayo, S. Tal, S. Elfström, A. Råde, O. Golan, S. Bölte, S. Baron-Cohen e D. Lundqvist (2019). **The eu-emotion voice database**. *Behavior Research Methods* **51**(2), 493–506.

- LeCun, Y. e Y. Bengio (1998). **Convolutional networks for images, speech, and time series**. Em: *The Handbook of Brain Theory and Neural Networks* (Michael A. Arbib, Ed.). Cap. Convolutio, p. 43. MIT Press.
- Levy, S. E., D. S. Mandell e R. T. Schultz (2009). **Autism**. *The Lancet* **374**(9701), 1627–1638.
- Liang, Y., C. Li, Y. Guan e Y. Hu (2016). **Gait recognition based on the golden ratio**. *EURASIP Journal on Image and Video Processing* **2016**(1), 22.
- Lima, R. C., M. Cristina, V. Couto, F. P. Solis, B. Diniz, C. De Oliveira, P. Gabriel e G. Delgado (2017). **Atenção psicossocial a crianças e adolescentes com autismo nos CAPSi da região metropolitana do rio de janeiro 1 2**. pp. 196–207.
- Liu, C., K. Conn, N. Sarkar e W. Stone (2007). **Affect recognition in robot assisted rehabilitation of children with autism spectrum**. Em: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. number April. pp. 10–14.
- Loomes, R., L. Hull e W. P. L. Mandy (2017). **What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis**. *Journal of the American Academy of Child & Adolescent Psychiatry* **56**(6), 466–474.
- Lovaas, O. Ivar, Alan Litrownik e Ronald Mann (1971). **Response latencies to auditory stimuli in autistic children engaged in self-stimulatory behavior**. *Behaviour Research and Therapy* **9**(1), 39–49.
- Mamdani, E.H. e S. Assilian (1975). **An experiment in linguistic synthesis with a fuzzy logic controller**. *International Journal of Man-Machine Studies* **7**(1), 1–13.
- Mammone, A., M. Turchi e N. Cristianini (2009). **Support vector machines**. *Wiley Interdisciplinary Reviews: Computational Statistics* **1**(3), 283–289.

- Martínez-González, A. E. e J. A. Piqueras (2018). **Validation of the repetitive behavior scale-revised in spanish-speakers participants with autism spectrum disorder**. *Journal of Autism and Developmental Disorders* **48**(1), 198–208.
- Mason, G. J. (1991). **Stereotypies: A critical review**. *Animal Behaviour* **41**(6), 1015–1037.
- Meeren, H. K. M., C. C. R. J. van Heijnsbergen e B. Gelder (2005). **Rapid perceptual integration of facial expression and emotional body language**. *Proceedings of the National Academy of Sciences of the United States of America* **102**(45), 16518–16523.
- Meijer, M. (1989). **The contribution of general features of body movement to the attribution of emotions**. *Journal of Nonverbal Behavior* **13**(4), 247–268.
- Moore, Richard S., Francis Cabana e K. A I Nekaris (2015). **Factors influencing stereotypic behaviours of animals rescued from asian animal markets:a slow loris case study**. *Applied Animal Behaviour Science* **166**(1), 131–136.
- Moss, R., C. Moss, N. Jo, L. Jo e C. Oliver (2016). **Persistence of self-injurious behaviour in autism spectrum disorder over 3 years: A prospective cohort study of risk markers**. *Journal of Neurodevelopmental Disorders* **8**(1), 1–12.
- Novoa, J., J. Fredes, V. Poblete e N. B. Yoma (2018). **Uncertainty weighting and propagation in dnn–hmm-based speech recognition**. *Computer Speech & Language* **47**, 30–46.
- O’Dwyer, J., R. Flynn e N. Murray (2017). **Continuous affect prediction using eye gaze**. Em: *2017 28th Irish Signals and Systems Conference (ISSC)*. IEEE. pp. 1–6.

- Ohnishi, T., H. Matsuda, T. Hashimoto, T. Kunihiro, M. Nishikawa, T. Uema e M. Sasaki (2000). **Abnormal regional cerebral blood flow in childhood autism**. *Brain : a journal of neurology* **123 (Pt 9)**, 1838–1844.
- Ortony, A. (2009). **Affect and emotions in intelligent agents: Why and how?**. Em: *Affective Information Processing* (J. Tao e T. Tan, Eds.). Cap. 2, pp. 11–21. Springer London.
- Othman, M. e A. Wahab (2010). **Affective face processing analysis in autism using electroencephalogram**. Em: *Proceeding of the 3rd International Conference on Information and Communication Technology for the Moslem World (ICT4M) 2010*. IEEE. pp. E23–E27.
- Ouafae, B., L. Oumaima, R. Mariam e L. Abdelouahid (2020). **Novelty detection review state of art and discussion of new innovations in the main application domains**. Em: *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. IEEE. pp. 1–7.
- Paiva Junior, F. e S. Ribeiro (2010). **Pesquisa do CDC revela número alto de prevalência de autismo nos EUA em crianças de oito anos, além de grande aumento em relação a pesquisa anterior**. *Revista Autismo*.
- Paula, C. S., C. S. Duarte e I. A. S. Bordin (2007). **Prevalence of mental health problems in children and adolescents from the outskirts of São Paulo city: treatment needs and service capacity evaluation**. *Revista Brasileira de Psiquiatria* **29(1)**, 11–17.
- Péter, Z., M. E. Oliphant e T. V. Fernandez (2017). **Motor stereotypies: A pathophysiological review**. *Frontiers in Neuroscience* **11(MAR)**, 1–6.
- Polanco, A., M. Díez-León e G. Mason (2018). **Stereotypic behaviours are heterogeneous in their triggers and treatments in the american mink, neovison vison, a model carnivore**. *Animal Behaviour* **141**, 105–114.

- Poyo Solanas, M., M. J. Vaessen e B. Gelder (2020). **The role of computational and subjective features in emotional body expressions**. Scientific Reports **10**(1), 6202.
- PrimeSense (2011). **OpenNI: User guide**. https://github.com/OpenNI/OpenNI/blob/master/Documentation/OpenNI{_}UserGuide.pdf. Acessado Junho 26, 2020.
- Rabiner, L. R. (1989). **A tutorial on hidden markov models and selected applications in speech recognition**. IEEE **77**(2), 257–286.
- Rad, N. M., A. Bizzego, S. M. Kia, G. Jurman, P. Venuti e C. Furlanello (2015). **Convolutional neural network for stereotypical motor movement detection in autism**. pp. 1–9.
- Rad, N. M., S. M. Seyed, C. Zarbo, T. Laarhoven, G. Jurman, P. Venuti, E. Marchiori e C. Furlanello (2018). **Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders**. Signal Processing **144**, 180–191.
- Ramchoun, H., M. Amine, J. Idrissi, Y. Ghanou e M. Ettaouil (2016). **Multilayer perceptron: Architecture optimization and training**. International Journal of Interactive Multimedia and Artificial Intelligence **4**(1), 26.
- Razali, N. e A. Wahab (2011). **2d affective space model (asm) for detecting autistic children**. pp. 536–541.
- Rodrigues, J. L., N. Gonçalves, S. Costa e F. Soares (2013). **Stereotyped movement recognition in children with ASD**. Sensors and Actuators, A: Physical **202**, 162–169.
- Roether, C. L., L. Omlor, A. Christensen e M. A. Giese (2009). **Critical features for the perception of emotion from gait**. Journal of Vision **9**(6), 1–32.
- Rosenblatt, Frank (2005). **The perceptron: A probabilistic model for information storage and organization in the brain**. Psychological Review **65**(6), 1–23.

- Russell, J. A. (2003). **Core affect and the psychological construction of emotion**. *Psychological Review* **110**(1), 145–172.
- Sadouk, L., T. Gadi e E. H. Essoufi (2018). **A novel deep learning approach for recognizing stereotypical motor movements within and across subjects on the autism spectrum disorder**. *Computational Intelligence and Neuroscience* **2018**, 1–16.
- Sarabadani, S., L. C. Schudlo, A. Samadani e A. Kushki (2018). **Physiological detection of affective states in children with autism spectrum disorder**. *IEEE Transactions on Affective Computing* pp. 1–1.
- Savva, N., A. Scarinzi e N. Bianchi-Berthouze (2012). **Continuous recognition of player’s affective body expression as dynamic quality of aesthetic experience**. *IEEE Transactions on Computational Intelligence and AI in Games* **4**(3), 199–212.
- Scholkopf, B., A. J. Smola, R. C. Williamson e P. L. Bartlett (2000). **New support vector algorithms**. *Neural computation* **12**(5), 1207–1245.
- Schopler, E., M. E. V. Bourgondien, G. J. Wellman e S. R. Love (2010). ***Childhood Autism Rating Scale (CARS2)***. 2 ed.. Western Psychological Services. Torrance,CA.
- Schouwstra, S. J. e J. Hoogstraten (1995). **Head position and spinal position as determinants of perceived emotional state**. *Perceptual and Motor Skills* **81**(2), 673–674.
- Scott, J. (1988). **Psychological models of depression**. *Current Opinion in Psychiatry* **1**(6), 719–724.
- Shotton, J., A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman e A. Blake (2011). **Real-time human pose recognition in parts from single depth images**. Em: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. pp. 1297–1304.

- Shu, L., J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu e X. Yang (2018). **A Review of Emotion Recognition Using Physiological Signals**. *Sensors* **18**(7), 2074.
- Silva, P. R. e N. Bianchi-Berthouze (2004). **Modeling human affective postures: an information theoretic characterization of posture features**. *Computer Animation and Virtual Worlds* **15**(3 – 4), 269–276.
- Singh, J. P., S. Jain, S. Arora e U. P. Singh (2018). **Vision-based gait recognition: A survey**. *IEEE Access* **6**, 70497–70527.
- Spunt, R. P., E. Ellsworth e R. Adolphs (2017). **The neural basis of understanding the expression of the emotions in man and animals**. *Social Cognitive and Affective Neuroscience* **12**(1), 95–105.
- Sugeno, M. (1985). *Industrial Applications of Fuzzy Control*. Elsevier Science Inc.. New York, NY, USA.
- Thakur, N. e C. Y. Han (2018). **A complex activity based emotion recognition algorithm for affect aware systems**. Em: *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018*. Vol. 2018-January. IEEE. pp. 748–753.
- Troisi, A. (2002). **Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects**. *Stress* **5**(1), 47–54.
- Tse, C. Y. A., C. L. Pang e P. H. Lee (2018). **Choosing an appropriate physical exercise to reduce stereotypic behavior in children with autism spectrum disorders: A non-randomized crossover study**. *Journal of Autism and Developmental Disorders* **48**(5), 1666–1672.
- Wallbott, H. G. (1998). **Bodily expression of emotion**. *European Journal of Social Psychology* **28**(6), 879–896.
- Wang, L. X. (1997). *A Course in Fuzzy Systems and Control*. PTR, Prentice Hall.

- Wang, Tim T., Louis Wessels, Gazi Hussain e Steve Merten (2017). **Discriminative thresholds in facial asymmetry: A review of the literature**. *Aesthetic Surgery Journal* **37**(4), 375–385.
- Watt, N., A. M. Wetherby, A. Barber e L. Morgan (2008). **Repetitive and stereotyped behaviours in children with autism spectrum disorders in the second year of life**. *J Autism Dev Disord* **38**(8), 1518–1533.
- Wei, M., B. Bai, A. H. Sung, Q. Liu, J. Wang e M. E. Cather (2007). **Predicting injection profiles using ANFIS**. *Information Sciences* **177**, 4445–4461.
- WHO, World Health Organization (2020*a*). **Autism spectrum disorders**. <http://www.who.int/mediacentre/factsheets/autism-spectrum-disorders/en/>. Acesso em 26 de junho de 2020.
- WHO, World Health Organization (2020*b*). **International statistical classification of diseases and related health problems: Tenth revision – ICD – 10**. Relatório técnico. World Health Organization. Geneva.
- WHO, World Health Organization (2020*c*). **International classification of functioning, disability and health – ICF**. Relatório técnico. World Health Organization. Geneva.
- Yacoub, S., S. Simske, X. Lin e J. Burns (2003). **Recognition of emotions in interactive voice response systems**. Relatório técnico. Hewlett-Packard. HP Laboratories Palo Alto.
- Zhao, W., R. Chellappa, P. J. Phillips e A. Rosenfeld (2003). **Face recognition: A literature survey**. *ACM Comput. Surv.* **35**(4), 399–458.

Apêndice A

Grupos Associados TEA de acordo com CID

A.1 Grupos Associados TEA de acordo com CID

O TEA é tratado no *Capítulo V – Transtornos Mentais e Comportamentais (F00–F99)* – na subparte de *Transtornos do Desenvolvimento Psicológico (F80–F89)* e *Transtornos Invasivos do Desenvolvimento* da Classificação Internacional de Doenças - 10ª revisão (CID-10). Assim, o TEA é classificado como:

- Autismo infantil (*F84.0*);
- Autismo atípico (*F84.1*);
- Síndrome de Rett (*F84.2*);
- Transtorno desintegrativo da infância (*F84.3*);
- Transtorno hiperativo associado a retardo mental e movimentos estereotipados (*F84.4*);
- Síndrome de Asperger (*F84.5*);
- Outros transtornos globais do desenvolvimento (*F84.8*);

- Transtorno invasivo do desenvolvimento, não especificado (*F84.9*).

Além desses, o autismo também está relacionado aos grupos (WHO, 2020*b*):

- Transtornos específicos do desenvolvimento da fala e da linguagem (*F80*);
- Distúrbios específicos de desenvolvimento das habilidades escolares (*F81*);
- Transtorno específico do desenvolvimento da função motora (*F82*).

A.2 Áreas do CIF Associados ao TEA

As áreas associadas ao TEA de acordo com o Classificação Internacional de Funcionalidades (CIF) são:

- Aprendizagem e aplicação do conhecimento;
- Tarefas e demandas gerais;
- Comunicação;
- Mobilidade;
- Cuidado pessoal;
- Vida doméstica;
- Relações e interações interpessoais;
- Áreas principais da vida;
- Vida comunitária, social e cívica.

A.3 Áreas do CARS2 Associados ao TEA

Childhood Autism Rating Scale (CARS2). Esta escala é organizada e 15 categorias:

- Relação com pessoas;
- Imitação;
- Respostas emocionais;
- Uso do corpo;
- Uso de objeto;
- Adaptação a mudanças;
- Respostas visuais;
- Respostas audição;
- Resposta e uso do paladar, cheiro e toque;
- Medo ou nervosismo;
- Comunicação verbal;
- Comunicação não verbal;
- Nível de atividade;
- Nível e consciência de resposta intelectual;
- Impressão geral.

Apêndice B

Ferramentas de Inteligência Computacional

As ferramentas de Inteligência Computacional utilizadas neste trabalho são descritas a seguir. Os Modelos Ocultos de Markov é abordado na Seção B.1. A Máquina de Vetores de Suporte é tratada na Seção B.2. A abordagem neural Peceptron Multicamadas é descrita na Seção B.3. Mais uma abordagem neural, Redes Neurais Convolucionais, é descrita na Seção B.4. Finalmente, o Sistema de Inferência Neuro-Fuzzy Adaptativo é tratado na Seção B.5.

B.1 Modelos Ocultos de Markov

Modelos Ocultos de Markov (*Hidden Markov Models – HMM*) são constituídos por uma cadeia markoviana dividida em duas sub-cadeias estocásticas: uma cadeia formada por símbolos observáveis e outra com símbolos não observáveis (ou oculta). Os processos estocásticos observáveis podem produzir observações (símbolos) discretos (HMM discreto) ou contínuos (HMM contínuo) (Rabiner, 1989).

A Figura B.1 ilustra os principais elementos de um HMM.

Assim, os principais elementos que caracterizam um HMM são,

- Conjunto finito de estados $S = \{S^{[1]}, S^{[2]}, \dots, S^{[n_S]}\}$ de tamanho n_S ;

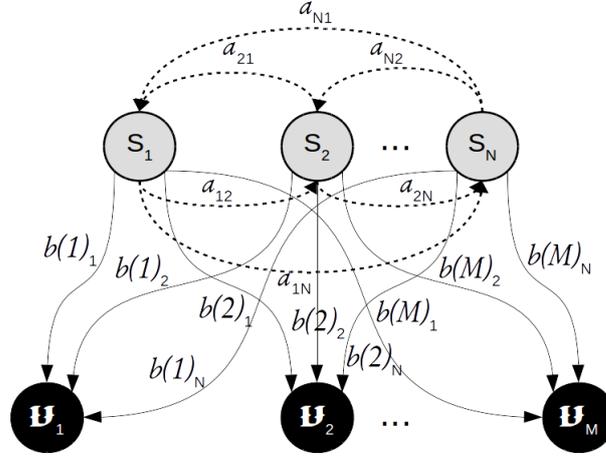


Figura B.1: Caracterização da arquitetura geral de um Modelo Oculto de Markov.

- Distribuição da probabilidade de transição dos estados $A = \{a_{ij}\}$, onde $a_{ij} = P[q_t = S_j | q_{t-1} = S_i]$, $1 \leq i, j \leq N$ no instante t ;
- Conjunto discreto de tamanho M de símbolos observáveis possíveis $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M\}$;
- Para o HMM discreto, distribuição de probabilidade de transição de símbolos de observação no estado j , $B = \{b_j(f)\}$, onde

$$b_j(f) = P[v_f \text{ em } t | q_t = S_j], \quad 1 \leq j \leq N, \quad (\text{B.1})$$

$$1 \leq f \leq M. \quad (\text{B.2})$$

- Distribuição do estado inicial $\pi = \{\pi_i\}$, onde $\pi_i = P(q_i = S_i)$, $1 \leq i \leq N$ no instante t .

A caracterização anterior do HMM deve apresentar restrições na forma do modelo da Função de Densidade de Probabilidade (FDP) para aplicação para o HMM contínuo, assegurando que os parâmetros da FDP possam ser re-estimados de

maneira consistentes. Para isto uma mistura finita é definida na forma,

$$b_j(O) = \sum_{m=1}^M c_{jm} R[O, \boldsymbol{\mu}_{jm}, U_{jm}], \quad 1 \leq j \leq N, \quad (\text{B.3})$$

onde O é o vetor inicialmente modelado, c_{jm} é o coeficiente de mistura para o m -ésima mistura no estado j e R é, normalmente, uma Função Densidade de Probabilidade (FDP) Gaussiana com vetor de média $\boldsymbol{\mu}_{jm}$ e matriz de covariância U_{jm} para a m -ésima componente de mistura no estado j . O ganho da mistura c_{jm} satisfaz a seguinte restrição estocástica,

$$\sum_{m=1}^M c_{jm} = 1, \quad 1 \leq j \leq N \quad (\text{B.4})$$

$$c_{jm} \geq 0, \quad 1 \leq j \leq N, 1 \leq m \leq M. \quad (\text{B.5})$$

Assim, a FDP deve ser normalizada,

$$\int_{-\infty}^{\infty} b_j(x) dx = 1, \quad 1 \leq j \leq N. \quad (\text{B.6})$$

A FDP da eq. (B.3) pode ser aproximada de maneira arbitrária da função de densidade contínua. Assim, pode-se aplicar em diferentes tipos de problemas. A re-estimação dos coeficientes de densidade de mistura c_{jm} , m_{jk} e U_{jk} é realizada em (Rabiner, 1989, eq. 52, 53, 54).

Deste modo, uma notação [compacta] possível para o HMM contínuo é definida como,

$$\Lambda = (A, c, \boldsymbol{\mu}, U, \pi), \quad (\text{B.7})$$

onde A é uma matriz de probabilidade de transição, c é um conjunto de coeficiente (pesos para cada Gaussiana), $\boldsymbol{\mu}$ representa as médias de cada mistura Gaussiana, U representa a matriz de covariância da Gaussiana, e π é distribuição do estado inicial.

Assim, a partir desta notação, uma sequência de observações, $O_1 O_2 \cdots O_T$ através dos valores de N , M , A , B e π é gerada de acordo com os seguintes passos:

1. Escolha um estado inicial, $q_1 = S_i$, de acordo com distribuição do estado inicial π ;
2. Atribua $t = 1$;
3. Escolha $O_t = \mathbf{v}_k$ de acordo com b_{i_t} ;
4. Transite para um novo estado $q_{t+1} = S_j$ de acordo com a_{ij} ;
5. Atribua $t = t + 1$; se $t < T$ retorne para o passo 3, caso contrário termine a operação.

O HMM tem sido aplicado em diferentes tipos de problema, mas sendo o problema de reconhecimento de fala a sua aplicação clássica (Juang e Rabiner, 1991; Noivoa *et al.*, 2018; Hueber e Bailly, 2016) devido ao seu desempenho satisfatório. Além deste tipo de aplicação, o HMM tem sido utilizado também para o reconhecimento de face (Farhan *et al.*, 2016) ou expressões faciais (Boruah *et al.*, 2015), na computação biológica através do auxílio na modelagem de proteínas (Kamal *et al.*, 2017), como também em sistemas dinâmicos no processo de detecção de falhas (Galagedarage Don e Khan, 2019).

O processo de reconhecimento de gestos envolvidos no estado afetivo de um indivíduo apresenta uma maior complexidade em relação a Linguagem de Sinais, pois além das informações explícitas do movimento dos gestos, deve considerar também a contextualização determinada através de meta-dados sugestivos dos próprio gesto. Bernhardt e Robinson (2009) utilizam o HMM para classificar ações do cotidiano, como caminhar, levantar, bater da porta e movimento de jogar e a partir disto, determina o estado afetivo do indivíduo. Silva e Bianchi-Berthouze (2004) propuseram a aplicação do HMM para mapear uma sequência de posturas de uma coreografia relacionando a uma probabilidade de cada estado emocional. Embora Kulic e Croft (2007) não utilizem gestos no processo de reconhecimento do estado afetivo do indivíduo, seu trabalho é interessante por utilizar o HMM para estimar o estado afetivo de um indivíduo durante o processo de IHR, utilizando dados de sinais biológicos como nível de condutividade da pele, batimentos cardíacos e nível de atividade muscular.

B.2 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte (*Support Vector Machine – SVM*) é uma ferramenta aplicada para o reconhecimento de padrões complexos, regressão ou extração de características. A SVM aplicada para o reconhecimento de padrões é uma máquina de aprendizado supervisionado que trata, originalmente, de problemas de classificação em dois grupos, mas que pode ser estendida para uma quantidade superior atualmente. A SVM é capaz de gerar funções de mapeamento entrada-saída a partir de um conjunto de dados rotulados de treino (Cortes e Vapnik, 1995).

O objetivo da SVM para classificação consiste em pesquisar eficientemente por um “bom” (em relação alguma medida de desempenho de generalização) hiperplano de separação em um espaço característico de alta dimensão. O hiperplano “bom” é definido como um com margem máxima (margem rígida) de separação entre duas classes (Cortes e Vapnik, 1995).

A SVM de margem rígida é aplicado em amostras de treino linearmente separáveis $\mathfrak{S} = \{(x_1, y_1), \dots, (x_l, y_l)\} \subseteq (X \times Y)^l$ onde $X \subseteq \mathfrak{R}^n$ representa o espaço de entrada, $Y = \{-1, +1\}$ é o domínio de saída para uma classificação binária, l o total de pontos de dados, e x_i e y_i são seus rótulos. Esta SVM é baseada em classe de hiperplano,

$$\langle w, x \rangle + b = 0, \quad (\text{B.8})$$

onde $w, x \in \mathfrak{R}^n, b \in \mathfrak{R}$, correspondendo a função de decisão

$$f(x) = \text{sign}(\langle w, x \rangle + b) \quad (\text{B.9})$$

onde o ponto x se encontra no hiperplano satisfazendo a condição $\langle w, x \rangle + b = 0$, com w definindo uma direção perpendicular ao hiperplano, na variação do valor de b movendo-se para o hiperplano paralelo a ele mesmo (Cortes e Vapnik, 1995).

Deste modo o hiperplano ótimo é definido por um com margem máxima de separação entre duas classes. Ele pode ser unicamente construído solucionando um problema de otimização quadrada restrita, com solução w tendo uma expansão $w = \sum_i v_i x_i$ em termos de subconjunto do padrão de treino (denominado de vetores

de suporte) que estão na margem (ver Figura B.2) (Cortes e Vapnik, 1995; Mammone *et al.*, 2009).

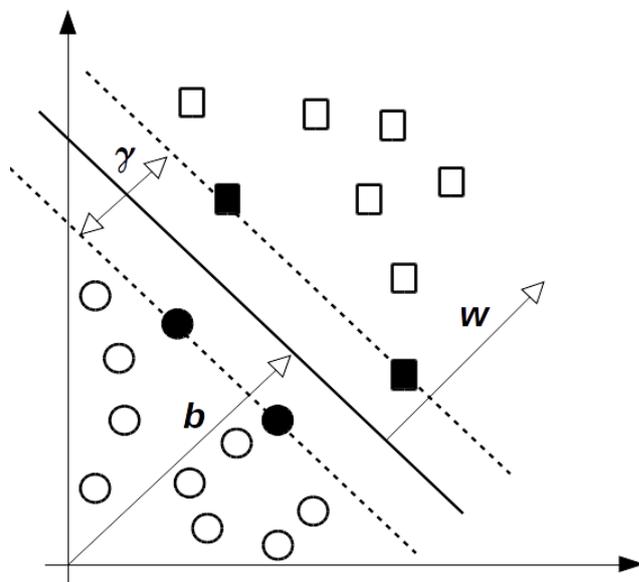


Figura B.2: Uma separação do hiperplano para dados de duas dimensões, onde os quadrados e circunferências representam pontos de dados nas classes -1 e +1, respectivamente. Os elementos nas linhas tracejadas pertencem ao vetor de suporte.

Há dois algoritmos de classificação para a SVM: (i) *C-Support Vector Classifier* (C-SVC) (Cortes e Vapnik, 1995), e (ii) *v-Support Vector Classification* (v-SVC) (Scholkopf *et al.*, 2000). O C-SVC é algoritmo clássico proposto por Cortes e Vapnik (1995).

No mundo real, os dados das amostras normalmente não pode ser separáveis linearmente. Neste sentido, há a necessidade de se definir funções mais complexas para lidar com dados reais. Para isto, o produto interno $\langle x_i, x_j \rangle$ deve ser representado de melhor forma através do mapeamento dos pontos dos dados em um espaço alternativo superior de características, deste modo, substituindo,

$$\langle x_i, x_j \rangle \rightarrow \langle \phi(x_i), \phi(x_j) \rangle. \quad (\text{B.10})$$

A forma da função $\phi(x)$ é definida por uma função de núcleo (*kernel*). Deste

modo, a representação dos dados torna-se,

$$x = (x_1, \dots, x_n) \rightarrow \phi(x) = (\phi_1(x), \dots, \phi_N(x)). \quad (\text{B.11})$$

Assim, o novo espaço de características é definido como,

$$F = \{\phi(x) : x \in X\}, \quad (\text{B.12})$$

onde $\phi : X \rightarrow F \subseteq \mathfrak{R}^N$ é o mapeamento embutido e x é um vetor contendo os valores característicos.

Uma função *kernel* é uma função \mathfrak{K} que para todos $x, z \in X$ satisfaz,

$$\mathfrak{K}(x, z) = \langle \phi(x), \phi(z) \rangle, \quad (\text{B.13})$$

onde ϕ é um mapeamento $\phi : x \rightarrow \phi(z) \in F$.

A dimensão de espaço de características não afeta a computação, mas operações requeridas para computar o produto interno na avaliação da função *kernel* não são proporcionais ao número de características, mas ao número de pontos de dados. Assim, as únicas informações obtidas das amostras de treino são a sua matriz de *kernel*, definida pela matriz quadrática $K \in \mathfrak{R}^{l \times l}$, tal que $K_{ij} = k(x_i, x_j)$ para um conjunto de vetores $\{x_1, \dots, x_l\} \subseteq X$ e alguma função *kernel* k .

A seguir os tipos básicos de funções *kernels*:

$$\mathfrak{K}(x_i, x_j) = x_i^T x_j, \quad (\text{B.14})$$

$$\mathfrak{K}(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0, \quad (\text{B.15})$$

$$\mathfrak{K}(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0, \quad (\text{B.16})$$

$$\mathfrak{K}(x_i, x_j) = \tanh(\gamma x_i^T x_j + r), \quad (\text{B.17})$$

onde as equações (B.14), (B.15), (B.16), e (B.17) são linear, polinomial, Função Básica Radial (*Radial Basis Function - RBF*) e sigmoïdal, respectivamente, e γ , r , d são os parâmetros dos *kernels*.

B.3 Perceptron Multicamadas

O Perceptron Multicamadas (*Multilayer Perceptron* – MLP) é um modelo de redes neurais supervisionado baseado na estrutura de conhecimento do cérebro humano (Ramchoun *et al.*, 2016). Um Perceptron consiste em um sistema nervoso artificial proposto por Rosenblatt (2005). O processo de treinamento é feito com algoritmo de *back-propagation*.

O MLP pode ser organizado em unidades denominado de camadas. Deste modo, o MLP apresenta no mínimo três camadas, consistindo de uma cada de entrada, uma [ou mais] camada oculta e camada de saída. Há os modelos degenerados em que apresentam apenas uma camada de entrada e uma de saída (Johnsen *et al.*, 2020). Além disto, é comum a omissão das camadas de entrada e de saída definindo apenas as camadas ocultas. Cada camada é constituída por nodos denominados de neurônios. Cada neurônio é conectada em uma única direção (*feed-forward*) com todos outros neurônios da camada subsequente (ver Figura B.3).

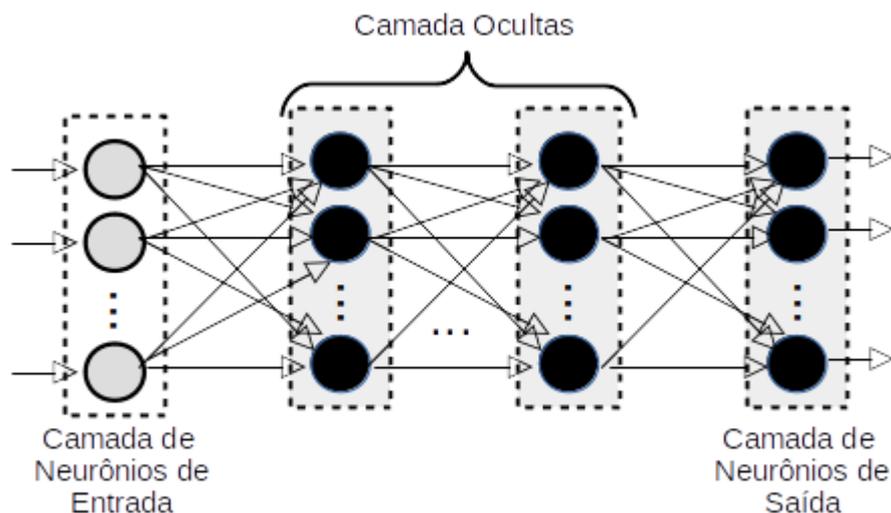


Figura B.3: Modelo da arquitetura do MLP constituído por neurônios organizados em camadas. Há três tipos de camadas: entrada, ocultas e de saída.

Os neurônios da camada de entrada são especiais, pois tem como objetivo apenas distribuir cada uma das entradas da rede para todos os neurônios da ca-

mada seguinte. Estes neurônios apresentam funções de ativação linear sem limiar. Os neurônios da camada de saída são associados a limiares além de pesos. Já os neurônios das camadas ocultas apresentam funções de ativação não linear, tal como função sigmóide (a mais popular),

$$\mathcal{S}(x) = \frac{1}{1 + e^{ax}}, \quad (\text{B.18})$$

onde a é o parâmetro de inclinação da função sigmóide.

A saída de cada um desses neurônios possui funções de ativações lineares. Deste modo, cada sinal que alimenta um neurônio em uma camada subsequente tem a entrada original multiplicada por um peso com um limiar adicionado e é passada através de uma função de ativação que pode ser linear ou não (no caso das camadas ocultas) (Johnsen *et al.*, 2020).

No MLP, o processo de aprendizado é iterativo com ajustes aplicados aos seus pesos sinápticos e limiares. O desempenho do MLP é mensurado pelo cálculo do Erro Quadrático Médio entre o padrão e a saída desejada (Johnsen *et al.*, 2020).

A principal vantagem do MLP é que ele aprende a partir de exemplos e é capaz de formar bordas não lineares entre classes. A relação entre os dados de entrada e o tipos de saída é encontrada através da adaptação dos seus pesos. Assim, não há a necessidade um domínio de conhecimento profundo sobre problema pelo usuário (Ramchoun *et al.*, 2016).

B.4 Redes Neurais Convolucionais

Redes Neurais Convolucionais (*Convolutional Neural Network – CNN*) são uma classe de Algoritmo de Aprendizado Neural Profundo proposto por LeCun e Bengio (1998). A CNN foi desenvolvida para reconhecer padrões em imagens visando uma maior dependência ao aprendizado automático ao invés de heurísticas projetadas manualmente.

O desempenho dos Algoritmos de Aprendizado de Máquina tradicionais são dependentes da habilidade do projetista de criar uma técnicas de extração de ca-

racterísticas apropriada para um determinado fenômeno alvo. No problema de reconhecimento de imagens, a CNN é capaz de determinar características de alto nível mesmo quando há algum grau de deslocamento (*shift*), mudança de escala, e distorção. A arquitetura da CNN é baseada na combinação de campos receptivos locais, pesos compartilhados (ou replicação de pesos) e subamostragem espacial ou temporal.

Uma vez as características tenham sido detectadas, sua exata localização torna-se menos importante (LeCun e Bengio, 1998). Um benefício imediato do uso de características de alto-nível é a redução de números de parâmetros para reconhecer padrões complexos em comparação com Redes Neurais Artificiais (Artificial Neural Networks - ANN) clássicas.

Uma CNN é caracterizada por apresentar múltiplas camadas ocultas. Existem 3 tipos básicos de camadas: convolucional, subamostragem (*pooling*) e camada totalmente conexa (ver Figura B.4).

A entrada da CNN pode ser uma imagem de dimensões com largura L_E e altura A_E e profundidade P_E . Em uma imagem colorida de canais RGB, pode-se considerar $P_E = 3$ (em imagens tons de cinza ou preto e branco o $P_E = 1$).

A camada convolucional visa reduzir o tamanho das entradas sem perder as características chaves da entrada através filtros de *kernel*. Após o filtro passar por toda a imagem, é forma o mapa de características de dimensão com largura L_i e altura A_i . Um filtro pode apresentar duas variáveis com propósito diferentes: passo (*stride*) e o preenchimento (*padding*). O tamanho da saída (O) após aplicação deste filtro é definido pela seguinte fórmula,

$$O = 1 + \frac{\mathcal{N} + 2\rho - \mathcal{K}}{\varsigma}, \quad (\text{B.19})$$

onde \mathcal{N} é o tamanho de entrada, \mathcal{K} , ς e ρ são os tamanhos do filtro, do passo (*stride*) e do espaçamento nas margens da matriz (*padding*), respectivamente.

A variável *stride* tem como objetivo diminuir o tamanho da saída resultante do filtro. A variável *padding* visa determinar a quantidade de informações que podem ser perdidas nas bordas das matrizes convolucionais.

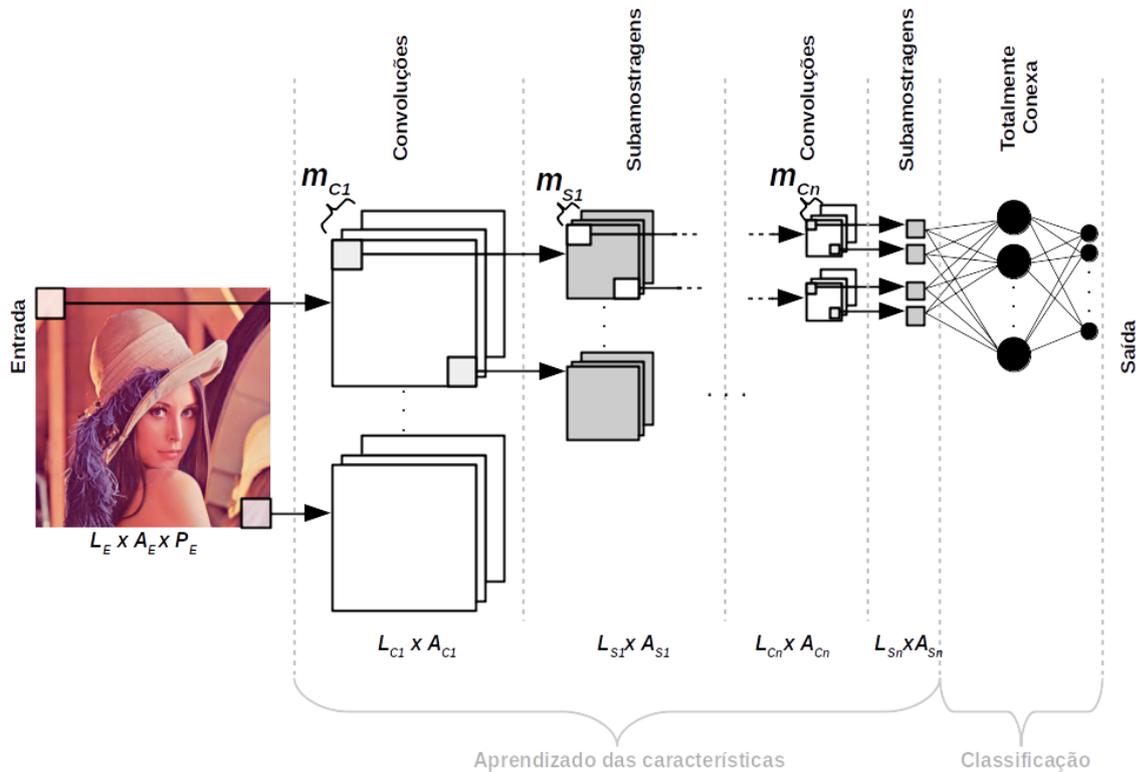


Figura B.4: Modelo geral da arquitetura da CNN.

Após determinação dos mapas de características após a convolução, é aplicada uma função de ativação neste elemento. A função de ativação tem como objetivo tornar as características não-lineares, deste modo, se tornando uma rede mais genérica. Diversas funções de ativação podem ser aplicadas, entretanto, a mais comum é a *Rectified Linear Unit* (ReLU) definida pelas equações B.20 e B.21.

$$ReLU(x) = \max(0, x) \quad (B.20)$$

$$\frac{d}{dx} ReLU(x) = 1 \quad \text{se } x > 0; 0 \quad \text{caso contrário} \quad (B.21)$$

A camada de amostragem *pooling* também visa reduzir o tamanho espacial das características convolucionais, além de extrair características dominantes. Esta camada pode ser de dois tipos: máxima e média. O primeiro tipo retorna valores

máximos a partir da subamostragem da imagem coberta por um filtro, suprimindo ruídos. O segundo tipo retorna a média de todos os valores de uma subamostragem das imagens coberta por um filtro.

Finalmente, a camada totalmente conectada é similar a uma rede neural tradicional. Cada nodo nesta camada é diretamente conecta a todos os nós da camada anterior e a classe de saída. Esta camada aprende combinações não-lineares das características de alto-nível representados pela camada convolucional. Deste modo, a camada totalmente conexa é um aprendizado de funções não lineares neste espaço. As camadas convolucionais e totalmente conexas são parametrizadas. Entretanto, as camadas de *polling* não são parametrizadas.

B.5 Sistemas de Inferência Neuro-Fuzzy Adaptativo

O Sistema de Inferência Neuro-Fuzzy Adaptativo (*Adaptive-Network-Based Fuzzy Inference System* – ANFIS) consiste em um sistema de inferência do tipo TSK (Subseção B.5.1). A utilização do ANFIS torna-se atrativa pois esta ferramenta consegue combinar regras de aprendizado das Redes Neurais Artificiais com a transparência linguística da teoria da lógica difusa.

B.5.1 Sistemas de Inferências Nebulosos

Os Sistemas Inteligentes têm sido aplicados para lidar com fenômenos sobre domínio incerto do mundo real. Uma classe desses sistemas são os Sistemas Baseados em Regras Nebulosas (ou Difusas), também conhecidos como Sistemas de Inferência Nebulosos (*Fuzzy Inference Systems* – FIS) (Wang, 1997).

De acordo com Wang (1997), os Sistemas de Inferência Nebulosos são baseados em conhecimento, sendo este representado por um conjunto regras (nebulosas) do tipo *SE-ENTÃO*. Estas regras são compostas por palavras que são caracterizadas por funções de pertinências. Wang (1997) classifica tais sistemas em três tipos:

- **Puro** – Representa o modelo genérico de FIS caracterizado com entrada e saída baseadas em linguagens naturais e um conjunto de regras como o Exemplo B.22. Quando a máquina de inferência é realimentada (como a seta pontilhada no Item A da Figura B.5), este sistema especialista é denominado também como *Sistema Dinâmico Nebuloso*.

$$\begin{aligned} & \mathbf{SE} \text{ velocidade está baixa e nível do combustível está alto,} \\ & \mathbf{ENTÃO} \text{ aumente a aceleração.} \end{aligned} \tag{B.22}$$

Como este modelo utiliza apenas palavras da linguagem natural, ele não é muito útil para área da engenharia, principalmente relacionado a Controle, em que há necessidade de valores em termos numéricos.

- **Takagi-Sugeno-Kang (TSK)** – Modelo proposto por Sugeno (1985) em que as entradas e saídas são valores reais. O Sistema Nebuloso neste modelo é uma média ponderada de um conjunto de regras que são descritas combinando a linguagem natural com simples fórmulas matemáticas como é representada pelo exemplo a seguir,

$$\begin{aligned} & \mathbf{SE} \text{ } v \text{ está baixa e } c \text{ está alto,} \\ & \mathbf{ENTÃO} \text{ } a = \gamma v, \end{aligned} \tag{B.23}$$

onde γ é uma constante.

A sua aplicação na engenharia torna-se mais atraente, pois suas entradas podem lidar com informações diretas de sensores ($x \in \mathfrak{R}$) e sua saída apresenta diretamente valores ($y \in \mathfrak{R}$) para os atuadores. A principal desvantagem do modelo TSK é que a parte seguida do *ENTÃO* é constituída por elementos matemáticos, o que dificulta a representação do conhecimento humano de maneira natural. Além disto, o Sistema de Inferência Nebuloso torna-se menos versátil devido a estes elementos matemáticos. Este modelo é representado pelo Item B da Figura B.5.

- **Sistema com *Fuzzifier* e *Defuzzifier* (ou Mamdani)** – Modelo de Sistema de Inferência Nebuloso proposto por Mamdani e Assilian (1975) para ser

aplicado na lógica de controle. Neste modelo, as entradas do sistema utilizam um método que transforma uma variável real ($x \in \mathfrak{R}$) em conjuntos nebulosos, denominado de *Fuzzifier*, e na saída emprega um método que transforma um conjunto nebuloso em variáveis com valores reais ($y \in \mathfrak{R}$), conhecido como *Defuzzifier*, como é ilustrado pelo Item C da Figura B.5. Deste modo, este modelo consegue representar de maneira natural o conhecimento humano ao mesmo tempo em que atende as necessidades relacionadas à área de Controle através dos valores reais de entrada e saída.

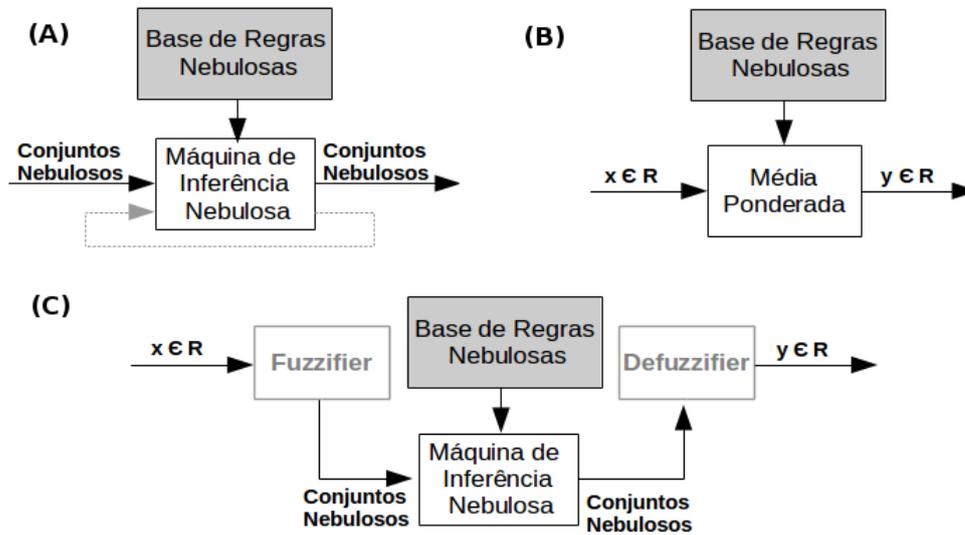


Figura B.5: Tipos de Sistemas Nebulosos de acordo com Wang (1997): (A) Puro, (B) TSK e (C) Mamdani.

Funções pertinência

Na Teoria dos Conjuntos Clássico, a Função Pertinência (μ_c) determina se um certo elemento (x) pertence ou não a um conjunto (C), ou seja,

$$\mu_c(x) = \begin{cases} 1, & x \in C \\ 0, & x \notin C \end{cases} \quad (\text{B.24})$$

A Função de Pertinência nos Conjuntos Nebulosos determina o quanto um elemento pertence ao um conjunto nebuloso. Assim, seja F um conjunto nebuloso tal que $F \subset \mathbb{U}$, onde \mathbb{U} é o conjunto universo de discurso. Então F é caracterizado pela *Função Pertinência* (μ_F), tal que,

$$\mu_F : \mathbb{U} \rightarrow [0, 1] \quad (\text{B.25})$$

Assim, F é representado pela Equação B.26, constituído por um conjunto de pares ordenados de valores da variável u e das suas funções de pertinência $\mu_F(u)$ (Wang, 1997).

$$F = \{(u, \mu_F(u)) | u \in \mathbb{U}\} \quad (\text{B.26})$$

A representação da coleção de todos os pontos de $u \in \mathbb{U}$ de F vai depender se \mathbb{U} é contínuo (Equação B.27) ou é discreto (Equação B.28).

$$F = \int_{\mathbb{U}} \frac{\mu_F(x)}{x} \quad (\text{B.27})$$

$$F = \sum_U \frac{\mu_F(x)}{x} \quad (\text{B.28})$$

Variáveis linguísticas

A utilização de palavras da linguagem natural associadas às variáveis quantitativas servem para caracterizar um determinado conjunto nebuloso. Estas palavras são denominadas de *Variáveis Linguísticas*. Nos Sistemas Nebulosos, as *Variáveis Linguísticas* são elementos importantes para representação do conhecimento humano (Wang, 1997).

Uma *Variável Linguística* é caracterizada por quatro parâmetros ($N, \mathcal{V}, \mathbb{U}, R$). O primeiro parâmetro, N , é nome da variável linguística (p. ex.: velocidade, temperatura). O conjunto de valores linguísticos possíveis (p. ex.: alta, média, baixa, quente, frio, morno) é representada por \mathcal{V} . O domínio físico da variável linguis-

tica N é determinado por \mathbb{U} . Deste modo, \mathbb{U} assume os valores quantitativos (p. ex.: $[0 \text{ km/h}, 120 \text{ km/h}]$, $[0^\circ\text{C}, 250^\circ\text{C}]$). A regra semântica que relaciona cada valor linguístico em V com o conjunto nebuloso \mathbb{U} é representada pelo parâmetro R (Wang, 1997).

Regras nebulosas

O conhecimento humano é representado nos Sistemas de Inferências Nebulosas através de um conjunto de Regras *SE-ENTÃO* Nebulosas. Estas regras são nebulosas, pois são constituídas por *Variáveis Linguísticas*, as quais são caracterizadas por *Funções de Pertinência* (Wang, 1997). Com o objetivo de simplificar o entendimento do modelo geral do ANFIS, é considerado apenas duas entradas s_i e s_j . Assim, sejam duas regras do tipo Takagi-Sugeno,

Regra 1 :

Se $s^{(i)}$ *é* $A^{(1)}$ *e* $s^{(j)}$ *é* $B^{(1)}$,

Então $f^{(1)} = p^{(1)}s^{(i)} + q^{(1)}s^{(j)} + r^{(1)}$,

Regra 2 :

Se $s^{(i)}$ *é* $A^{(2)}$ *e* $s^{(j)}$ *é* $B^{(2)}$,

Então $f^{(2)} = p^{(2)}s^{(i)} + q^{(2)}s^{(j)} + r^{(2)}$,

onde $A^{(1)}$, $A^{(2)}$, $B^{(1)}$, e $B^{(2)}$ são rótulos linguísticos associados a uma função de pertinência.

Arquitetura do ANFIS

A arquitetura do ANFIS é organizada em cinco camadas. Cada Camada l contém um conjunto de nodos n descrito por uma função pertinência e saída $O_h^{[l]}$.

A primeira camada, denominada de *fuzzification*, converte as entradas $s^{(i)}$ e $s^{(j)}$ dentro de um conjunto nebuloso pelas médias das funções pertinência (μ). A

saída da primeira camada é definida pela equação (Jang, 1993, eq. (17)),

$$O_n^1 = \mu_{A^{(n)}}(s^{(i)}) \quad (\text{B.29})$$

$$O_n^1 = \mu_{B^{(n)}}(s^{(j)}), \quad (\text{B.30})$$

onde $A^{(n)}$ e $B^{(n)}$ são variáveis linguísticas associadas a uma função pertinência das entradas $s^{(i)}$ e $s^{(j)}$, e μ é a função pertinência.

A segunda camada realiza o processo de inferência em que cada nodo multiplica o sinal de entrada e envia o produto para saída (Jang, 1993, eq. (20)),

$$O_n^2 = w^{(n)} = \mu_{A^{(n)}}(s^{(i)}) \times \mu_{B^{(n)}}(s^{(j)}), \quad n = 1, 2, \quad (\text{B.31})$$

onde cada nodo de saída representa a força de disparo da regra.

A terceira camada realiza o processo de normalização. O n -ésimo nodo calcula a relação da n -ésima força de disparo da regra (Jang, 1993, eq. (21)):

$$O_n^3 = \bar{w} = \frac{w^{(n)}}{w^{(1)} + w^{(2)}} \quad n = 1, 2, \quad (\text{B.32})$$

A quarta camada realiza o processo de interpolação. Todos os nodos n nesta camada é um nó ajustável com a função de nó definido da seguinte forma (Jang, 1993, eq. (22)):

$$O_n^4 = \bar{w}^{(n)} f^{(n)} = \bar{w}^{(n)} (p^{(n)} s^{(i)} + q^{(n)} s^{(j)} + r^{(n)}), \quad (\text{B.33})$$

onde $\bar{w}^{(n)}$ é a saída da dessa camada, e $p^{(n)}$, $q^{(n)}$, $r^{(n)}$ são o conjunto de parâmetros. Os parâmetros nesta camada serão referidos como parâmetros consequentes.

A quinta camada, denominada de camada de saída, processa a saída geral pela equação (Jang, 1993, eq. (23)),

$$O_n^5 = \text{overall output} = \sum_n \bar{w}^{(n)} f^{(n)} = \frac{\sum_n w^{(n)} f^{(n)}}{\sum_n w^{(n)}}. \quad (\text{B.34})$$

B.5.2 Estratégia de Treinamento do ANFIS

Existem diferentes estratégias, tanto para extrair regras ou para definir funções pertinências no processo de treinamento do ANFIS. Assim, os principais métodos de treinamento não supervisionado do ANFIS são: (i) Partição de Grade (*Grid Partition – GP*), Agrupamento Subtrativo (*Subtractive Clustering – SC*) e *Fuzzy C-Means (FCM)*.

O *Grid Partition* é um método de treinamento que particiona o espaço de dados dentro de sub-espacos retangulares usando eixos paralelos baseando-se em um número pré-definido de funções pertinências em cada dimensão. Quando associado ao ANFIS, o método dos mínimos quadrados pode ser usado para estimar tanto os parâmetros de premissas como consequentes do sistema nebuloso baseando-se na partição e funções pertinências. Uma desvantagem do GP aplicado ao ANFIS é o crescimento exponencial da quantidade de regras como aumento de entradas, necessitando de uma grande quantidade de recursos computacionais (Wei *et al.*, 2007).

O *Subtractive Clustering* é um método de agrupamento nebuloso interativo considerado rápido e capaz de estimar tanto o número de grupos como a localização dos centros deles. O algoritmo SC define o raio de vizinhança dos centros dos grupos. Caso o valor d deste raio seja pequeno os resultados podem negligenciar o efeito dos pontos de dados da vizinhança. Enquanto, valores muito grandes resultam que todos os pontos dos dados estarão dentro do alcance do raio, criando apenas um grupo (Benmouiza e Cheknane, 2019).

O *Fuzzy C-Means (FCM)* é uma outra técnica de agrupamento no qual cada ponto de dados pertence a um ou mais grupos. Cada ponto de dado é atribuído a um grupo por uma função pertinência. Esta atribuição é otimizada pelo processo interativo (Kapoor e Singhal, 2017).