

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

RECONHECIMENTO DE ATIVIDADES HUMANAS USANDO MEDIDAS
ESTATÍSTICAS DOS SENSORES INERCIAIS DOS SMARTPHONES

HENDRIO LUIS DE SOUZA BRAGANÇA

RECONHECIMENTO DE ATIVIDADES HUMANAS USANDO MEDIDAS
ESTATÍSTICAS DOS SENSORES INERCIAIS DOS SMARTPHONES

Dissertação apresentada ao Programa
de Pós-Graduação em Informática do
Instituto de Computação da
Universidade Federal do Amazonas
como requisito para a obtenção do grau
de Mestre em informática

ORIENTADOR: PROF. DR. EDUARDO JAMES PEREIRA SOUTO

CO-ORIENTADOR: PROF. DR. JUAN GABRIEL COLONNA

Manaus-AM

Maio de 2019

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

B813r Bragança, Hendrio Luis de Souza
Reconhecimento de atividades humanas usando medidas
estatísticas dos sensores inerciais dos smartphones / Hendrio Luis
de Souza Bragança. 2019
153 f.: il. color; 31 cm.

Orientador: Eduardo James Pereira Souto
Coorientador: Juan Gabriel Colonna
Dissertação (Mestrado em Informática) - Universidade Federal do
Amazonas.

1. Reconhecimento de atividades humanas. 2. Representação
simbólica. 3. Teoria da informação. 4. Sensores inerciais. 5.
Smartphone. I. Souto, Eduardo James Pereira II. Universidade
Federal do Amazonas III. Título



PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
INSTITUTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



UFAM

FOLHA DE APROVAÇÃO

"Reconhecimento de Atividades Humanas Usando Medidas Estatísticas dos Sensores Inerciais dos Smartphones"

HENDRIO LUIS DE SOUZA BRAGANÇA

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

Prof. Eduardo James Pereira Souto - PRESIDENTE

Prof. Juan Gabriel Colonna - MEMBRO INTERNO

Prof. Raimundo da Silva Barreto - MEMBRO INTERNO

Prof. Osvaldo Anibal Rosso - MEMBRO EXTERNO

Manaus, 01 de Março de 2019

Agradecimentos

Agradeço à minha mãe pela dedicação, por sempre ter me incentivado e apoiado em todos os momentos da minha vida e que sempre se esforçou para que eu chegasse até aqui.

Agradeço ao meu orientador professor Dr. Eduardo Souto pelo companheirismo nesta jornada, que se mostrou paciente e dedicado no seu trabalho e me ajudou em momentos chaves do meu projeto. Agradeço ao meu coorientador professor Dr. Juan Colonna por sua contribuição em um momento chave deste projeto.

Agradeço a todos os meus amigos pelas palavras de conforto e pelo apoio dado durante minha jornada acadêmica e pessoal. Agradeço por entenderem a ausência em alguns momentos e por estarem presentes sempre que possível. Agradecimento especial a meu primo Wesllen pelo companheirismo profissional e acadêmico em todos os anos. Agradeço aos amigos dos laboratórios da UFAM pelo companheirismo e contribuição durante o projeto e por compartilharem a vida acadêmica da forma mais agradável.

Por fim, agradeço a CAPES pelo suporte com a bolsa de mestrado, à UFAM e a todos os professores que de alguma forma estiveram presentes na minha jornada acadêmica.

“Se vi mais longe foi por estar sobre os ombros de gigantes.”

(Isaac Newton)

Resumo

O reconhecimento de atividades humanas (RAH) tem emergido como uma tecnologia revolucionária para monitorar o estilo de vida das pessoas. Com a popularização dos smartphones essa tarefa tem sido possível através dos sensores embarcados (e.g. os sensores inerciais acelerômetro e giroscópio) que permitem o reconhecimento de diferentes tipos de atividades usando técnicas de aprendizagem de máquina.

Os dados coletados pelos sensores são tratados com séries temporais e devem passar por diversas etapas até que o reconhecimento de atividades seja concluído. Assim, os dados devem ser processados, segmentados e transformados em um conjunto de características que representam as atividades desenvolvidas pelos usuários. Tais características são tradicionalmente expressas a partir de medidas matemáticas como, por exemplo, média, variância e desvio padrão, extraídas dos sinais dos sensores.

Essa forma de reconhecer atividades que compõe o modo mais convencional, tem obtido êxito e levado a grandes avanços neste domínio nos últimos anos. Entretanto, o emprego dessa abordagem requer que cálculos sejam definidos manualmente com o apoio de um especialista no domínio. À medida que novas atividades surgem, os modelos de classificação gerados perdem desempenho, requerendo que o especialista gere novas características para representar bem as novas atividades. Assim, a adoção dessa abordagem tem afetado a capacidade de generalização dos modelos de reconhecimento com o tempo.

Por esse motivo, as pesquisas mais recentes têm direcionado seus esforços em soluções que aprendem os padrões implícitos dos sinais (*feature learning*) e realizam o processo de extração de características de forma automática. Exemplos desses algoritmos são as redes neurais profundas e os algoritmos de representação simbólica. Em particular, os algoritmos de representação simbólica extraem características discretas automaticamente dos dados. Nesse contexto, este trabalho apresenta o método HAR-SR (do inglês, *Human Activity Recognition based on Symbolic Representation*), que corresponde a uma nova abordagem para problemas envolvendo classificação de atividades humanas. O método HAR-SR realiza extração automática de características no domínio discreto utilizando quantificadores estatísticos como novas características que representam atividades. Esses quantificadores associam valores às séries temporais conforme sua natureza determinística ou estocástica. Além disso, o HAR-SR utiliza estratégias de fusão de dados para combinar os sinais dos sensores e realiza uma redução de dimensionalidade nos dados. Os resultados desta pesquisa mostram que o HAR-SR apresenta um desempenho similar aos trabalhos estado-da-arte no reconhecimento de

atividades. O diferencial do método proposto, além da boa capacidade de generalização, está relacionado ao custo computacional, que é menor no processo de extração de características, visto que utiliza um número muito menor de características para gerar o modelo de classificação. Os resultados de avaliação usando três bases de dados reais (SHOAIB, UCI, WISDM) em seis cenários mostram que é possível classificar atividades com 93% de precisão, e na média de todos os cenários, apresenta 81% de precisão utilizando a validação cruzada por indivíduo.

Palavras-chave: Reconhecimento de atividades Humanas, representação simbólica, teoria da informação, sensores inerciais, smartphone.

Abstract

Human Activity Recognition (HAR) using different types of sensors has emerged as a revolutionary technology for monitoring people's lifestyles. With the popularization of smartphones, this task has been possible through embedded sensors (e.g. the accelerometer and gyroscope of inertial sensors) that allow recognize different types of activities using machine learning techniques.

The data collected by the sensors are treated as time series and must go through several stages until the activity recognition is completed. Thus, the data must be processed, segmented and transformed into a set of features that represent the activities performed by the users. These features are traditionally expressed from mathematical measures such as mean, variance and standard deviation extracted from the sensor signals.

This way of recognizing activities composes the most conventional way, has been successful and has made great strides in this field in recent years. However, using this approach requires that calculations be manually defined with the support of a domain expert. As new activities arise, the generated classification models lose performance, requiring the specialist to generate new measures to represent the new activities well. Thus, the adoption of this approach has affected the generalization capacity of recognition models over time.

For this reason, more recent research has focused its efforts on solutions that learn the feature patterns and perform the feature extraction process automatically. Examples of such algorithms are deep neural networks and symbolic representation algorithms. In particular, symbolic representation algorithms automatically extract discrete data characteristics. In this context, this work presents the HAR-SR (Human Activity Recognition based on Symbolic Representation) method, which corresponds to a new approach to solve problems involving classification of human activities. The HAR-SR method performs automatic extraction of characteristics in the discrete domain using statistical quantifiers as new characteristics that represent activities. These quantifiers associate values to the time series according to their deterministic or stochastic nature. In addition, the HAR-SR uses data fusion strategies to combine the signals from the sensors and perform a dimensionality reduction in the data. The results of this research show that HAR-SR performs similarly to state-of-the-art researches in activity recognition. The differential of the proposed method, besides the good generalization capacity, is related to the computational cost, which is smaller in the process of feature extraction, since it uses a smaller number of characteristics to generate the classification model. Evaluation results using three real databases (SHOAIB, UCI, WISDM) in five

scenarios show that it is possible to classify activities with 93% accuracy, and on average for all scenarios, presents 81% accuracy using leave-one-subject-out cross-validation.

Keywords: Human activity recognition, symbolic representation, information theory, inertial sensors, smartphone.

Lista de Figuras

Figura 1-1: Exemplo da extração manual e automática de características para um sistema de reconhecimento de atividades humanas.....	27
Figura 1-2: Processo de discretização de uma série temporal para obtenção de uma distribuição de frequência de símbolos.	28
Figura 1-3: Visão geral do método HAR-SR para reconhecer atividades humanas.	30
Figura 2-1: Metodologia utilizada nos trabalhos RAH: sensoriamento, pré-processamento, extração de características e classificação. Adaptado (Li, et al., 2018).....	34
Figura 2-2: Exemplo de dados do sensor de acelerômetro para as atividades andar e correr.....	36
Figura 2-3: A) Representação de cada um dos eixos do acelerômetro relativos ao usuário. B) Principais posições do smartphone em um usuário comum. Adaptado de Shoaib, et al. (2014).	36
Figura 2-4: Aplicação da técnica de segmentação simples e sem sobreposição de janelas na série temporal T . Logo, abaixo, a técnica de segmentação com sobreposição de janelas de 75% aplicada na série T	38
Figura 2-5: Exemplo da fusão dos sinais de dois sensores x_1 e $x_2 \in \mathbb{R}^3$, em um único sinal $r_1 \in \mathbb{R}$	40
Figura 2-6: Categorização baseada na natureza das características extraídas dos sinais dos sensores inerciais proposta por (Figo, et al., 2010).....	42
Figura 2-7: A- representação das instâncias rotuladas em um plano. A instância preta não está rotulada. B- cálculo da distância entre a instância preta e todas as outras instâncias.....	47
Figura 3-1. A nova categorização proposta separa os trabalhos em abordagens baseadas na extração manual e automática de características.	51
Figura 3-2. Uma série temporal com 50 amostras pode ser representada pela sequência “ abc ” através da utilização de um algoritmo de discretização.	53
Figura 3-3: Aplicação da técnica de janela deslizante em uma série temporal. Cada segmento é transformado em uma sequência de símbolos. O exemplo mostra quatro segmentos e suas respectivas representações simbólicas.....	54

Figura 3-4: Histograma de frequência de símbolos, que contém 16 símbolos no intervalos aa – dd. Note que a frequência é normalizada no intervalo [0, 1].....	55
Figura 3-5: A aplicação do PAA em uma série temporal (N=160) resultando um novo sinal (M=8). O novo sinal é composto por todos os valores t_i , que são calculados em cada intervalo.	57
Figura 3-6. A tabela de pesquisa mostra os pontos de interrupções que dividem a distribuição gaussiana em um número arbitrário de regiões equiprováveis.	57
Figura 3-7. Uma série temporal é discretizada da seguinte forma: i) obter a representação em PAA; ii) utilizar os pontos de interrupção predeterminados para mapear os valores do PAA em símbolos do SAX. Na figura acima, $N = 160$, $M = 8$ e $a = 3$, a série temporal é mapeada para a palavra 'caabccb'.....	58
Figura 3-8: Processo de aproximação de uma série temporal T através da Transformada de Fourier.	59
Figura 3-9: Processo de discretização utilizando a tabela de pesquisa MCB.....	60
Figura 3-10: Exemplo da representação de uma série temporal em uma sequencia de símbolos.	60
Figura 3-11. Transformação de uma série temporal T em um histograma de frequência de símbolos utilizando o modelo BOP. A série temporal é segmentada através da técnica de janela deslizante, e então discretizada por algum algoritmo de representação simbólica. O resultado é histograma de frequência de símbolos.	61
Figura 3-12: Histograma de frequência de símbolos obtido através do BOP. $\omega = 3$, $\alpha = 3$	62
Figura 4-1: Visão geral do método HAR-SR.....	71
Figura 4-2: Processo de segmentação aplicada aos sinais multidimensionais dos sensores de movimento.....	73
Figura 4-3: A: série temporal multidimensional. B: método da magnitude aplicado à série temporal da Figura A. C: método do PCA aplicado à série temporal da Figura A. D: método da concatenação aplicado à série temporal da Figura A.	74
Figura 4-4: Histograma de frequência de símbolos obtido através do método BOSS para uma série temporal.	75
Figura 4-5: Exemplo do <i>Bag-Of-Patterns</i> (BOP) para uma série temporal. Cada linha representa uma palavra do método de representação simbólica e	

cada coluna representa um segmento da série (ou o histograma gerado a partir desse segmento).	76
Figura 4-6: A entropia mínima pode ser representado pela distribuição à esquerda, onde $H_P = H_{min} = 0$. A distribuição uniforme P_e , à direita na figura, chamada também de histograma de uniforme (H_REF), representa uma série aleatória na qual todas as ocorrências possuem o mesmo valor de probabilidades (e.g. um ruído aleatório). No exemplo existem quatro possibilidades de símbolos, cada um com probabilidade $p = 1/4$	79
Figura 4-7: As distribuições P e Q são diferentes, mas apresentam a mesma entropia $H = 0.875$ (valor da entropia normalizada)	79
Figura 4-8: Exemplo do plano Entropia-Complexidade para duas classes A e B.	81
Figura 4-9: Exemplo do plano Entropia-Complexidade para três classes. Note que as classes B e C se sobrepõem em alguns pontos.	81
Figura 4-10: Processo de geração dos histogramas de referência PA e PB para cada classe A e B.	82
Figura 4-11: histograma de referência PA e PB para a classe A e classe B em um espaço de 9 símbolos.	83
Figura 4-12: Plano $H \times C_i$, onde C_i são as medidas de complexidade que utilizam os histogramas de referência de duas classes. Os valores são $4 = P_A$, $6 = P_B$	84
Figura 4-13: O processo de extração de características adotado pelo método HAR-SR.	86
Figura 4-14: Processo de extração de características de um novo sinal de atividade.	87
Figura 5-1: A: validação cruzada utilizando dados do mesmo indivíduo. B: Validação cruzada mista com dados de diferentes indivíduos.	92
Figura 5-2: Validação cruzada <i>leave-one—subject-out</i> (LOSO).	93
Figura 5-3: A métrica de validação utilizada na geração dos histogramas de referência das atividades é o <i>leave one out</i>	94
Figura 5-4. Em modelos personalizados, é utilizada a validação cruzada tradicional ou mista para geração dos histogramas de referência das atividades.	94
Figura 5-5. A matriz mostra a diagonal principal representada pelos valores $\{d_1, d_2, \dots, d_n\}$	95
Figura 5-6. Média de desempenho dos algoritmos de representação simbólica com variação nos parâmetros ω e α para cada posição da base de dados SHOAIB.	99

Figura 5-7: Comparação entre os métodos de fusão de dados aplicados a cada conjunto de sensores. A base de dados é a SHOAIIB na posição da cintura.	101
Figura 5-8: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAIIB com o smartphone localizado no bolso esquerdo.	101
Figura 5-9: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAIIB com o smartphone localizado no bolso direito.....	101
Figura 5-10: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAIIB com o smartphone localizado braço.	102
Figura 5-11: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAIIB com o smartphone localizado no pulso dos indivíduos.	102
Figura 5-12: Desempenho de classificação para cada grupo de sensores em cada posição da base SHOAIIB. o método de discretização é o SAX.	104
Figura 5-13: Desempenho de classificação para cada grupo de sensores em cada posição da base SHOAIIB. o método de discretização é o SFA.....	104
Figura 5-14: Desempenho de classificação para cada grupo de sensores em cada posição da base SHOAIIB. o método de discretização é o SAX.	104
Figura 5-15: Desempenho dos métodos de discretização SAX e SFA em cada base de dados no cenário D1.....	106
Figura 5-16: Comparação entre o método TF e o método proposto HAR-SR. Apenas o sensor do acelerômetro foi utilizado.....	109
Figura 5-17: Comparação entre o método BOSS – VS e o método proposto HAR-SR. Apenas o sensor do acelerômetro foi utilizado.....	109
Figura 5-18 Comparação entre o método SAX – VSM e o método proposto HAR-SR. Apenas o sensor do acelerômetro foi utilizado.....	110
Figura 5-19: Média do desempenho de classificação dos métodos utilizados no cenário E1. Os parâmetros utilizados podem ser encontrados na Tabela 5-13.....	110
Figura 5-20: Matriz de confusão da base de dados SHOAIIB utilizando dados do sensor de acelerômetro na posição da cintura. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, C = SAX, ω = 6, α =6). ...	112
Figura 5-21: Matriz de confusão da base de dados SHOAIIB utilizando dados do sensor de acelerômetro com o smartphone localizado na posição do bolso	

esquerdo. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, $C = SAX$, $\omega = 6$, $\alpha = 6$).....	112
Figura 5-22: Matriz de confusão da base de dados SHOAIIB utilizando dados do sensor de acelerômetro com o smartphone localizado na posição do bolso direito. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, $C = SAX$, $\omega = 6$, $\alpha = 6$).....	113
Figura 5-23: Matriz de confusão da base de dados SHOAIIB utilizando dados do sensor de acelerômetro com o smartphone localizado no braço. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, $C = SAX$, $\omega = 6$, $\alpha = 6$).....	114
Figura 5-24: Matriz de confusão da base de dados SHOAIIB utilizando dados do sensor de acelerômetro com o smartphone localizado no pulso do indivíduo. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, $C = SAX$, $\omega = 6$, $\alpha = 6$).....	114
Figura 5-25: Matriz de confusão da base de dados SHOAIIB utilizando dados do sensor de acelerômetro com dados de todas as posições. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, $C = SAX$, $\omega = 6$, $\alpha = 6$).....	115
Figura 5-26: Diferença entre a média dos resultados obtidos entre o conjunto de treino (CV) e teste (LOSO) do método TF.	116
Figura 5-27: Diferença entre a média dos resultados obtidos entre o conjunto de treino (CV) e teste (LOSO) do método HAR-SR.	117
Figura 5-28: Diferença entre a média dos resultados do modelo personalizado utilizando validação cruzada tradicional para diferentes classificadores. O método **HAR-SR utiliza histogramas de referência do mesmo usuário.	117
Figura 6-1: A figura mostra diversos exemplos de séries temporais e seu comportamento no plano Entropia-Complexidade.	119
Figura 6-2: Exemplo do plano Entropia-Complexidade de um indivíduo para seis atividades.....	120
Figura 6-3. histograma de referência PR e PP para a classe de atividade ‘correr’ e ‘parado’ respectivamente, em um espaço de 256 símbolos.	121
Figura 6-4: Plano Entropia – Complexidadei, onde C_i são as medidas de complexidade que utilizam os histogramas de referência de cada atividade. Os valores são 1 = PD, 2 = PR, 3 = PS, 4 = PP, 5 = PU, 6 = PW.....	122

Figura 6-5: Diferenças entre as atividades “correndo” e ”sentando” geradas no Plano Entropia-Complexidade para diferentes histogramas de referência.	122
Figura 6-6: Comparação entre dois indivíduos através do plano Entropia – Complexidadei, onde i utiliza o histograma PR.	123
Figura 6-7: Comparação entre dois indivíduos que estão executando a atividade ‘correr’. Histogramas utilizados: 1 = PD, 2 = PR, 3 = PS, 4 = PP, 5 = PU, 6 = PW.	123
Figura 6-8: O gráfico de densidade é utilizado para comparar as diferenças entre dois indivíduos. No exemplo, dois indivíduos executam a atividade “correr”.....	124
Figura 6-9: Plano Entropia-Complexidade para os métodos de fusão de dados Magnitude, PCA e concatenação simples.	125
Figura 6-10: Plano de complexidade da atividade “sentando” da base de dados WISDM com os algoritmos de representação simbólica SAX e SFA, variando-se os parâmetros de tamanho de palavra e tamanho do alfabeto.	126
Figura 7-1: O método de classificação pode reconhecer a posição do smartphone e a partir disso, selecionar os sensores apropriados para o reconhecimento de atividades.	136
Figura 7-2: Uma nova metodologia pode realizar a extração de características a partir de cada sinal do sensor para gerar o conjunto de treinamento.....	136

Lista de Tabelas

Tabela 2-1: Principais características do domínio do tempo utilizadas pelas abordagens de extração manual de características. Todas as características da tabela são utilizadas nesta pesquisa.	43
Tabela 2-2: Principais características do domínio da frequência utilizadas pelas abordagens de extração manual de características. Todas as características desta tabela são utilizadas nesta pesquisa.	44
Tabela 2-3: Exemplo do vetor de características de uma base de dados rotulada.....	46
Tabela 2-4: métricas de distância que podem ser utilizadas no algoritmo KNN.	48
Tabela 3-1: Sumário das notações usadas no SAX.....	56
Tabela 3-2: Resumo dos trabalhos de reconhecimento de atividades em smartphones utilizando sensores de movimento que utilizam abordagens baseadas na extração manual de características.....	65
Tabela 3-3: Resumo dos trabalhos de reconhecimento de atividades em smartphones utilizando sensores de movimento que utilizam abordagens baseadas na extração automática de características.....	67
Tabela 4-1: A Tabela mostra a relação dos histogramas de referência utilizados em um problema de classificação com duas classes.	83
Tabela 4-2: cálculo da divergência com base no histograma de referência PA e PB	83
Tabela 4-3: cálculo da medida de complexidade com base no histograma de referência PA e PB	84
Tabela 4-4: tabela de valores obtidos através do cálculo da entropia, divergência e complexidade para um problema com duas classes.	84
Tabela 5-1: Sumário das bases de dados selecionadas para esta pesquisa.	91
Tabela 5-2: Terminologia utilizada na matriz de confusão	95
Tabela 5-3. Os quatro números fundamentais para estimar as medidas estatísticas de performance de um classificador.....	95
Tabela 5-4: Matriz de confusão para duas classes.....	95
Tabela 5-5: Sumário das métricas de avaliação utilizadas nesta pesquisa.	96
Tabela 5-6: Variação dos parâmetros do tamanho das palavra e tamanho do alfabeto	98
Tabela 5-7: A tabela apresenta os parâmetros utilizados no cenário A1.....	98

Tabela 5-8: Os métodos utilizados na fusão de dados são a Magnitude, o PCA e a concatenação. Os parâmetros utilizados são: $C = \text{SAX}$, $T=128$, $w = 50100.T$ (50% janela deslizante), $\omega = 6$, $\alpha = 6$	100
Tabela 5-9: Agrupamento do conjunto de sensores para realização dos testes do cenário C1.....	100
Tabela 5-10: Posições em que os dados do smartphone foram coletados.....	103
Tabela 5-11: Tabela de parâmetros utilizada na avaliação dos algoritmos de representação simbólica.....	106
Tabela 5-12: Sumário das características utilizadas pelo método TF.....	107
Tabela 5-13: Parâmetros utilizados em cada um dos trabalhos avaliados no cenário E1. *TF é um método de classificação que utiliza características do domínio do tempo e frequência.	108
Tabela 5-14: Parâmetros utilizados na geração dos resultados apresentados na matriz de confusão no cenário E1.....	111
Tabela 6-1: Tabela de cada histograma de referência utilizado.	121

Sumário

Agradecimentos	vii
Resumo	xi
Abstract	xii
Lista de Figuras	xv
Lista de Tabelas	xxi
1 Introdução	26
1.1. Objetivos	29
1.2. Abordagem proposta para o reconhecimento de atividade humanas	29
1.3. Contribuições	31
1.4. Organização da Dissertação	32
2 Reconhecimento de atividades humanas em smartphones	33
2.1. Aquisição de dados	35
2.2. Pré-processamento e segmentação de sinais	37
2.3. Fusão e agregação de dados	39
2.4. Extração de características dos sinais	42
2.5. Algoritmos de aprendizagem supervisionada	45
2.6. Considerações sobre o capítulo	48
3 As diferentes abordagens para reconhecimento de atividades humanas	50
3.1. Extração automática de características	51
3.2. Por que utilizar métodos de discretização?	53
3.3. Discretização de séries temporais	55
3.3.1. O Algoritmo SAX	56
3.3.2. O Algoritmo SFA	58
3.3.3. Extração de subestruturas de séries temporais	60
3.4. Trabalhos Relacionados	63
3.4.1. Abordagens baseadas na extração manual de características	63
3.4.2. Abordagens Baseadas na extração automática de características	66

3.4.3.	Discussão sobre os trabalhos relacionados	68
4	Reconhecimento de atividades baseado em representação simbólica	71
4.1.	Visão Geral	71
4.2.	Aquisição, segmentação e fusão dos dados	73
4.3.	Representação simbólica dos sinais dos sensores	75
4.4.	Extração de características utilizando medidas estatísticas.....	77
4.5.	Método de classificação do HAR-SR	85
4.6.	Considerações Finais do Capítulo.....	88
5	Experimentos e resultados.....	89
5.1.	Protocolo Experimental.....	89
5.1.1.	Conjunto de dados.....	89
5.1.2.	Métricas de validação	91
5.1.3.	Méridas de Avaliação	94
5.1.4.	Cenários de avaliação	96
5.2.	Cenário A1: avaliação de parâmetros.....	98
5.3.	Cenário B1: fusão de dados.....	99
5.4.	Cenário C1: avalia a seleção de sensores por posição	103
5.5.	Cenário D1: Avaliação dos algoritmos de representação simbólica.....	105
5.6.	Cenário E1: comparação do HAR-SR com outros trabalhos	107
5.6.1.	Interpretação da matriz de confusão.....	111
5.7.	Cenário F1: modelo generalizado vs modelo personalizado.....	116
5.8.	Considerações do capítulo	118
6	Análise de atividades utilizando o plano Entropia-Complexidade	119
6.1.	Considerações do capítulo	127
7	Conclusões e direções futuras	128
7.1.	Considerações sobre a extração automática de características	128
7.2.	Considerações sobre a seleção de sensores e fusão de dados	130
7.3.	Considerações sobre os métodos de representação simbólica	131
7.4.	Considerações sobre o HAR-SR.....	132
7.5.	Considerações sobre o plano Entropia-Complexidade.....	134

7.6. Dificuldades encontradas.....	135
7.7. Direções futuras	135
8 Referências	138
Apêndice A	144

CAPÍTULO

1

Introdução

O reconhecimento de atividades humanas tem emergido como uma área de pesquisa revolucionária para monitoramento autônomo e em tempo real de diversos tipos de comportamentos humanos (Nweke, et al., 2018). Um fato comum é que as pessoas expressam suas rotinas por meio de atividades que elas executam em situações particulares. Assim, o reconhecimento de atividades físicas como andar, correr, deitar, ou andar de bicicleta é útil para prover respostas sobre determinadas rotinas (Lara & Labrador, 2013). Entender esse contexto possibilita que as pessoas prevejam situações inesperadas ou melhorem situações do seu cotidiano (Dernbach, et al., 2012) (Shoaib, et al., 2015) (Banovic, et al., 2016). Por essa razão, as pesquisas na área de reconhecimento de atividades humanas que tem ganhado muita atenção devido à alta demanda de aplicações em diferentes domínios como detecção de comportamento, cuidados com idosos, reabilitação, entretenimento e vigilância (Nweke, et al., 2018).

Recentemente, o smartphone tem sido utilizado como a ferramenta mais adequadas para monitorar automaticamente as atividades do dia-a-dia sem impor inconvenientes aos usuários. Esses dispositivos são equipados com uma variedade de sensores embarcados, tais como acelerômetro, giroscópio, microfone, câmera, GPS (localização), entre outros (Harari, et al., 2016).

Entender como os indivíduos se comportam por meio da análise de dados coletados pelos sensores de *smartphones* é um desafio da área de pesquisa conhecida como RAH - Reconhecimento de Atividades Humanas (do inglês, *Human Activity Recognition*), que tem como objetivo reconhecer ações a partir de uma série de observações, tanto dos agentes envolvidos (usuários) quanto das condições do ambiente (Gu, et al., 2016).

A literatura tem proposto diferentes estratégias que utilizam algoritmos de inteligência artificial. Na forma mais convencional, a maior parte dos trabalhos se concentra na extração manual de características dos sinais dos sensores (do inglês, *handcraft features* - *HF*) (Anguita, et al., 2013) (Shoaib, et al., 2015) (Li, et al., 2018). O processo de extração manual de características precisa de um especialista para transformar os dados dos sensores em uma representação de alto nível antes de serem inseridos no algoritmo de aprendizagem e essa transformação é decidida ou executada por um especialista a partir de seu conhecimento na área.

Essas características são obtidas por cálculos estatísticos simples, como média, mediana e desvio padrão, que, em geral, fornecem um modelo de classificação de atividades com boa precisão. As características estatísticas são simples de entender e possuem baixa complexidade computacional quando comparadas à outras abordagens, como por exemplo as baseadas em redes neurais profundas. (Figo, et al., 2010) (Nweke, et al., 2018) (Ravi, et al., 2016). A desvantagem das abordagens manuais são que os recursos criados ou selecionados manualmente consomem tempo, são específicos do domínio e por isso exigem conhecimento especializado, além de serem dependentes da tarefa ou aplicação em questão e não podem ser transferida para atividades ou padrões similares (Ronao & Cho, 2016).

A necessidade de técnicas mais robustas para extrair características de sinais de forma automática tem sido o foco de muitos trabalhos. O processo de extração automática de características pode acontecer sem a intervenção de um especialista. Neste contexto, as redes neurais profundas (do inglês, *Deep Neural Network* - *DNN*) têm sido recentemente utilizadas no contexto de RAH (Nweke, et al., 2018) (Ravi, et al., 2017). A Figura 1-1 mostra um exemplo da extração manual e automática de características para um sistema de reconhecimento de atividades humanas.

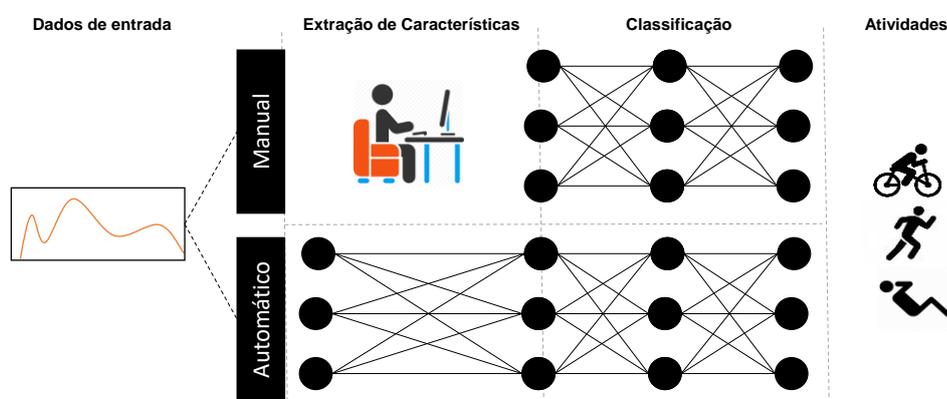


Figura 1-1: Exemplo da extração manual e automática de características para um sistema de reconhecimento de atividades humanas.

No entanto, por mais que essa categoria de métodos seja promissora, também existem diversos desafios a serem superados (Nweke, et al., 2018) (Raví, et al., 2017), como:

1. As redes neurais profundas apresentam custo computacional elevado quando comparado às abordagens convencionais, o que os torna impróprios para criação de aplicações RAH em tempo real implementadas em dispositivos móveis com baixo poder computacional (Ravi, et al., 2016) (Nweke, et al., 2018).
2. Para treinar uma rede neural profunda é necessário um conjunto massivo de dados para que o modelo seja efetivo e confiável. Caso contrário, o modelo pode até ser capaz de aprender com uma pequena quantidade de dados, porém, ele não será genérico o suficiente para reconhecer novas entradas que ainda não foram observadas (Nweke, et al., 2018).
3. Em muitas aplicações, os dados brutos não podem ser inseridos diretamente na rede, sendo necessárias algumas etapas adicionais como pré-processamento, normalização e mudanças no domínio da entrada.
4. Os modelos DNN, de modo geral, não são interpretáveis e funcionam como uma caixa preta, no qual não é possível explicar o porquê de bons resultados serem obtidos. Em aplicações do mundo real seria interessante que as interpretações desses modelos mostrassem diferenças entre indivíduos ou até mesmo anomalias de comportamento quando comparados com outros grupos.

Por outro lado, os algoritmos de discretização (ou representação simbólica) apresentam soluções simples e com baixo custo computacional. Esses algoritmos têm mostrado bons resultados quando aplicados no domínio de sinais (Bagnall, et al., 2016). O processo de discretização consiste em transformar uma série temporal, representada por valores reais, em um conjunto de símbolos pertencentes ao domínio discreto, dispostos em um espaço muito menor que o original, conforme mostra a Figura 1-2.

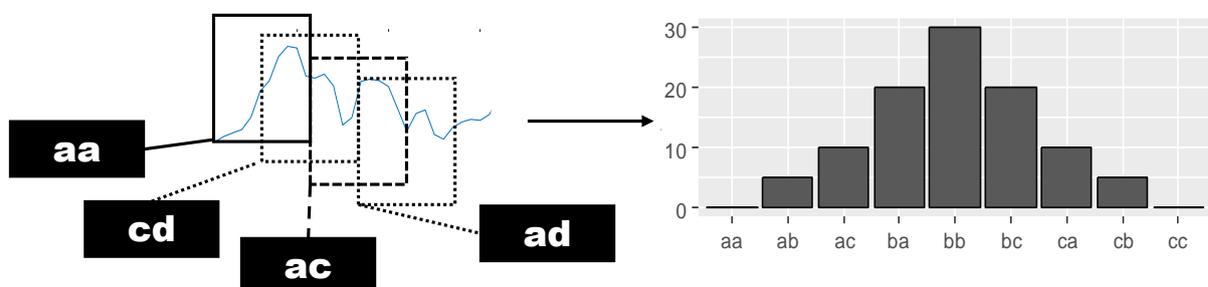


Figura 1-2: Processo de discretização de uma série temporal para obtenção de uma distribuição de frequência de símbolos.

A representação desses dados no domínio discreto combinados com algoritmos de mineração de dados (Lin, et al., 2003) (Manning, et al., 2008) e da teoria da informação (Martin, et al., 2006) (Rosso, et al., 2010) (Antonelli, et al., 2018) apresentam bom desempenho na tarefa de classificação de séries temporais, domínio onde a área RAH baseada em sensores inerciais está inserida (Schäfer, 2016).

1.1. Objetivos

Este trabalho tem como objetivo aprimorar o processo de reconhecimento de atividades humanas por meio do uso da medida de complexidade estatística. O método HAR-SR (do inglês, *Human Activity Recognition based on Symbolic Representation*) aprende as características implícitas dos sinais dos sensores e às apresenta no domínio discreto, na forma de uma distribuição de probabilidades. A partir dessa distribuição, esta pesquisa apresenta um novo conjunto de características para representar atividades humanas. As novas características são medidas estatísticas, provenientes da teoria da informação, que serão utilizadas na criação do modelo de classificação de atividades. Um algoritmo de aprendizagem supervisionada é utilizado para criar um modelo que finaliza a construção do método proposto.

Para atingir esse objetivo, pretende-se alcançar os seguintes objetivos específicos:

- Avaliar os métodos de fusão de dados para transformar as séries multidimensionais dos sinais dos sensores em séries unidimensionais.
- Propor um método de extração de características no domínio discreto.
- Avaliar a capacidade de generalização e especialização do método utilizando a métrica de validação cruzado por indivíduo.
- Criar uma forma de visualização de padrões no domínio RAH através do plano Entropia-Complexidade utilizando o conjunto de características obtidos.

1.2. Abordagem proposta para o reconhecimento de atividade humanas

A nova abordagem de extração automática de características proposta nesta pesquisa combina os métodos de representação simbólica e quantificadores estatísticos da teoria da informação. Os quantificadores estatísticos atribuem valores às séries temporais

conforme sua natureza determinística ou estocástica. Essa combinação se mostra promissora para classificar problemas relacionados não somente às atividades humanas, mas também a problemas de classificação de séries temporais de modo geral.

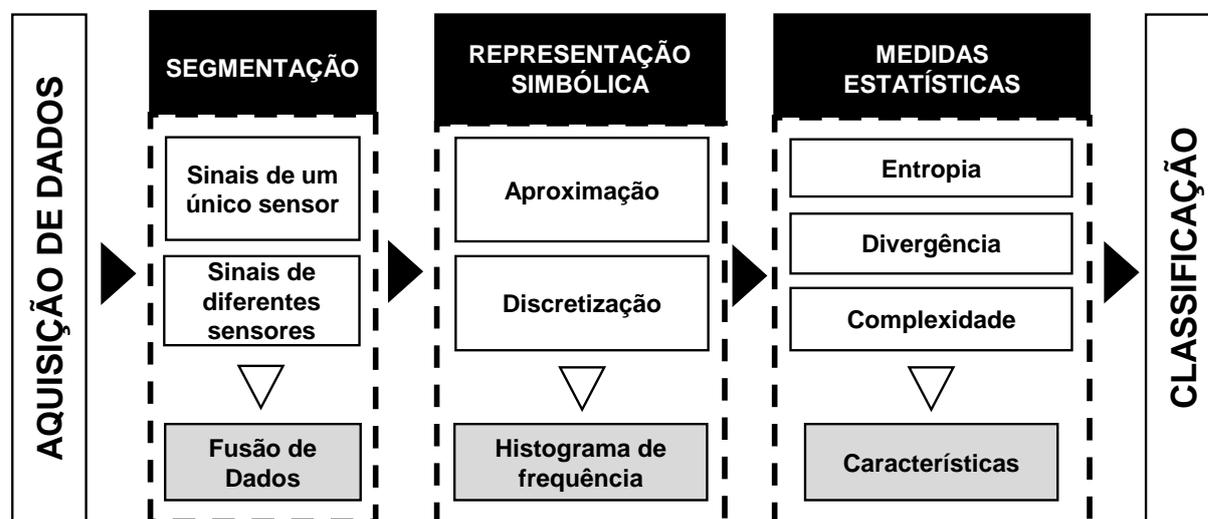


Figura 1-3: Visão geral do método HAR-SR para reconhecer atividades humanas.

As medidas estatísticas que utilizadas neste trabalho são a Entropia de Shannon, a Divergência Jensen-Shannon e a Medida de Complexidade Estatística (Riedl & Wessel, 2013) (Rosso, et al., 2007) (Antonelli, et al., 2018). A partir dessas características são derivadas novas características que utilizam as informações da distribuição de frequências de cada atividade. Os resultados obtidos têm mostrado que as novas características conseguem melhorar a representação das atividades e possuem um alto poder de generalização quando aplicadas em problemas que envolvem classificação.

Outro diferencial desta pesquisa é apresentado através do plano Entropia-Complexidade, que combina a entropia de Shannon e a medida de complexidade estatística. Esse plano permite que os diferentes comportamentos de atividades sejam interpretados visualmente. O plano Entropia-Complexidade tem se mostrado uma ferramenta poderosa de visualização de padrões. Até o presente momento, nenhum trabalho utilizou as medidas estatísticas como características para o problema RAH.

Uma das questões levantadas neste trabalho é sobre a influência das características no desempenho dos algoritmos de classificação. Outro tópico de estudo analisa como os sinais de diferentes tipos de sensores podem influenciar na inferência de uma atividade, bem como a aplicação de técnicas de fusão de dados nesses sinais (Gu, et al., 2015) (Shoaib, et al., 2014). Essas questões devem mostrar porque os métodos que extraem muitas características tendem a ser computacionalmente mais caros. Isso pode tornar o processo inviável, principalmente devido às altas taxas de amostragem dos sensores comumente adotadas pelas aplicações.

Os resultados da avaliação usando três bases de dados reais (SHOAIB, UCI-HAR e WISDM) em seis cenários mostram que é possível classificar atividades com 93% de precisão, e na média de todos os cenários, apresenta 81% de precisão utilizando a validação cruzada por indivíduo. Além disso, o processo de extração de características para gerar o modelo de classificação de atividades possui baixo custo computacional, sendo inferior à ordem quadrática. Esse processo resulta em um conjunto menor de características quando comparado com os modelos tradicionais (Anguita, et al., 2013) e com os principais algoritmos do domínio discreto (Senin & Malinchik, 2013) (Schäfer, 2016).

1.3. Contribuições

As principais contribuições deste trabalho são:

- Um algoritmo de extração automática de características a partir das séries temporais dos sensores inerciais presentes nos smartphones. Para o processo de classificação de atividades, uma nova representação é obtida a partir das medidas estatísticas, advindas da teoria da informação, como entropia de Shannon, divergência Jensen-Shannon e medida complexidade estatística. A partir dessas três medidas, um novo conjunto de características é obtido com base nas informações de cada classe de atividade. A quantidade de características necessárias para geração do modelo de classificação de atividades é menor se comparado com outros trabalhos da literatura.
- Uma nova forma de representar atividades e seus padrões através de um plano Entropia-Complexidade. A partir desse plano é possível analisar o comportamento de indivíduos e determinar seus padrões de forma visual. Os valores apresentados no plano possuem interpretações físicas e estão relacionadas com as características dos sinais dos sensores inerciais aos quais estão associadas, onde o determinismo e o estocástico são os dois extremos. Essa forma de representação é uma inovação na área RAH, não sendo apresentada em nenhuma pesquisa estudada até o momento.
- Utilização do plano Entropia-Complexidade, que utiliza um novo conjunto de características com informações de todas as classes do problema. O novo plano utiliza distribuições acerca das diferentes atividades para separá-las, permitindo assim, que seus padrões sejam analisados. Esse novo plano é inédito na literatura.

1.4. Organização da Dissertação

O restante desta dissertação está organizado como segue:

O Capítulo 2 apresenta os principais conceitos utilizados na área de RAH, incluindo as atividades executadas nos smartphones, tipos de sensores, aquisição de dados, pré-processamento, segmentação de sinais, técnicas de fusão de dados, metodologias para extração de características e algoritmos de aprendizagem de máquina clássicos.

O Capítulo 3 introduz o domínio discreto dos sinais e os motivos pelo qual é utilizado nesta pesquisa. Esse capítulo também apresenta a metodologia de RAH abordada no Capítulo 2, porém focada em diferentes estratégias, como extração automática de características e representação dos dados no domínio discreto. Para finalizar, são apresentados os trabalhos da literatura para reconhecer as atividades humanas que utilizam smartphones. Os trabalhos relacionados a esta pesquisa estão organizados conforme a natureza de suas características, que podem ser extração de características manual e automática.

O Capítulo 4 descreve em detalhes o método proposto, chamado HAR-SR, utilizado no reconhecimento de atividades humanas. A arquitetura do novo método possui três fases principais: segmentação do sinal dos sensores, discretização dos dados e extração de características. O método HAR-SR apresenta uma estratégia que ainda não foi aplicada no domínio de RAH como, por exemplo, a utilização de quantificadores estatísticos como novas características utilizadas no processo de classificação.

O Capítulo 5 apresenta o protocolo experimental, os resultados e comparações obtidos pelo método HAR-SR. Nesse capítulo são descritas as bases de dados utilizadas, a metodologia de validação, as métricas de avaliação e os resultados obtidos nos seis cenários avaliados.

O Capítulo 6 mostra uma análise exploratória do reconhecimento de atividades humanas através do plano de Entropia-Complexidade obtido através do cálculo dos quantificadores estatísticos. Esse plano permite que os padrões de atividades sejam identificados visualmente e permite que diferentes grupos de atividades ou indivíduos sejam estudados individualmente.

O Capítulo 7 apresenta uma discussão sobre os pontos positivos e negativos encontrados no decorrer da pesquisa, mostrando as conclusões acerca dos resultados obtidos pelo método HAR-SR. Por fim, são apontadas futuras direções para o método HAR-SR.

Reconhecimento de atividades humanas em smartphones

O reconhecimento de atividades humanas (RAH) é fundamental para o desenvolvimento de soluções que capturam o comportamento e a rotina humana (Lara & Labrador, 2013). Nos últimos anos o smartphone tem empenhado um importante papel no RAH devido à ubiquidade desses dispositivos e por estar sempre na posse dos usuários. Estudos mostram que até 2020, mais de 2.5 bilhões de pessoas irão utilizar smartphones¹ com alto poder de processamento, memória e acesso à internet. Além disso, esses dispositivos também possuem sensores embarcados, como acelerômetro, giroscópio, magnetômetro, temperatura, pressão e luminosidade. Por essa razão, os smartphones tem se tornado uma ferramenta necessária e a mais adequada para obter informações importantes sobre os hábitos de seus portadores, tais como o registro da rotina social, padrões de mobilidade e atividades do dia-a-dia (Miller, 2012).

Os dados obtidos através dos sensores embarcados podem prover informações contínuas das interações sociais, tais como chamadas de voz ou texto ou duração das conversas; registram os padrões de mobilidade para determinar quais períodos de tempo que pessoas passam em determinados locais, ou a frequência com que esses locais são visitados; e registram atividades diárias, como atividades físicas e padrões de sono, diretamente ligadas à saúde e a qualidade de vida (Harari, et al., 2016). Isso possibilita o desenvolvimento de aplicações de sensoriamento em diversos domínios, como monitoramento da saúde, condições de tráfego, comércio, monitoramento ambiental (Colonna, et al., 2014) e reconhecimento de diversas atividades (Khan, et al., 2013).

¹ <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>

O reconhecimento de atividades inicia com a coleta de dados através de uma variedade de sensores. Neste trabalho, por exemplo, serão utilizados apenas os sensores de movimento para reconhecer atividades físicas, como andar, correr, sentar, entre outras. Na literatura, os principais sensores de movimento são o acelerômetro e o giroscópio. O RAH é baseado no pressuposto de que movimentos específicos do corpo possuem padrões de características específicas nos sinais dos sensores. Por exemplo, a atividade física “andar” possui um padrão de sinal característico diferente do padrão de sinal para a atividade “deitar”. Dessa forma, reconhecer atividades físicas se resume na tarefa de encontrar padrões nos sinais dos sensores e associá-los às atividades executadas.

Os trabalhos em RAH costumam dividir o processo de classificação em cinco etapas principais (Figura 2-1): sensoriamento, pré-processamento, extração de características e treinamento ou classificação (Shoaib, et al., 2015) (Li, et al., 2018). A escolha das técnicas utilizadas em cada etapa deve levar em consideração às limitações dos recursos computacionais do smartphone (processamento, bateria e memória). Por isso, é fundamental a escolha correta dos sensores a serem utilizados, do método de extração de característica e do algoritmo para a classificação de atividades. O uso inadequado dos recursos pode significar a rejeição da aplicação por seus usuários (Pires, et al., 2016).

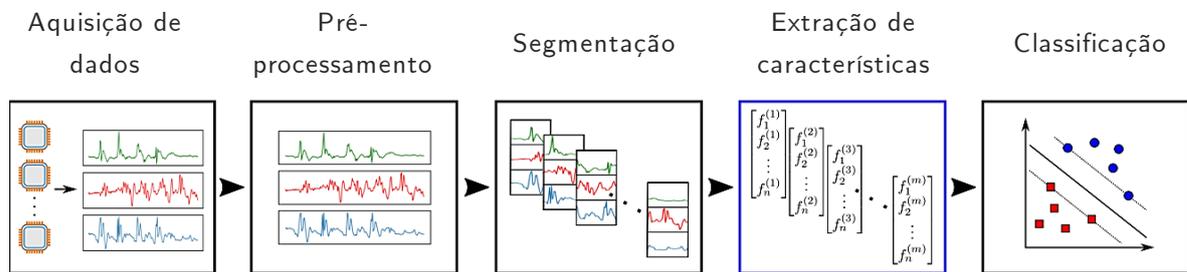


Figura 2-1: Metodologia utilizada nos trabalhos RAH: sensoriamento, pré-processamento, extração de características e classificação. Adaptado (Li, et al., 2018)

Na fase de aquisição de dados, a literatura mostra que os sensores de movimento (e.g. acelerômetro e giroscópio) têm sido a escolha mais popular para reconhecer atividades físicas humanas (Shoaib, et al., 2016) (Khan, et al., 2013) (Incel, et al., 2013). A escolha desses sensores está relacionada à capacidade de extrair padrões de movimento dos sinais e a quantidade de recursos computacionais que eles consomem. Outra vantagem é que os sinais coletados desses sensores podem ser representados na forma de séries temporais. Em resumo, uma série temporal é uma sequência de pontos ordenados no tempo. Dessa forma, todo conhecimento aplicado no domínio de séries temporais pode ser utilizado no contexto RAH.

Na fase de pré-processamento e segmentação, os estudos focam na escolha do melhor segmento do sinal que represente uma atividade. A escolha do tamanho do segmento impacta diretamente no tempo de resposta da aplicação, ou seja, quanto menor é o segmento, mais rápida é a resposta e vice-versa. Além disso, o tamanho do segmento também influencia na quantidade de dados necessário para representar as atividades e na complexidade dos modelos (Banos, et al., 2014).

A etapa de extração de características consiste em extrair representações de dados de alto nível. Essas representações, chamadas de características, são fundamentais no reconhecimento ou distinção de diferentes tipos de atividades. As características são utilizadas por algoritmos de aprendizagem de máquina para gerar um modelo de classificação (Duda, et al., 2000) (Cichosz, 2014) (Witten, et al., 2016). Por fim, o modelo de classificação é utilizado para reconhecer as atividades dos usuários. Assim, o problema RAH consiste em um problema de classificação de séries temporais, cuja solução está em encontrar uma métrica para extração de características relevantes dessas séries para geração de um modelo de classificação.

Diversos trabalhos da literatura mostram que o desempenho de um sistema de reconhecimento de atividades depende de cada um dos processos mostrados (Banos, et al., 2014) (Shoaib, et al., 2015) (Li, et al., 2018). No entanto, o foco principal desses trabalhos está na fase de extração de características, pois a qualidade de classificação de atividades depende da qualidade das características obtidas. Nas próximas seções cada fase da metodologia será detalhada.

2.1. Aquisição de dados

A coleta de dados no domínio RAH é realizada comumente através dos sensores de movimento presentes no smartphone (Lara & Labrador, 2013) (Khan, et al., 2013) (Incel, et al., 2013) (Shoaib, et al., 2016). Dentre os sensores, o acelerômetro é o principal sensor utilizado na literatura, seguido pelo giroscópio e magnetômetro. Esses sensores servem para monitorar o movimento do smartphone, como inclinação, vibração, rotação e oscilação, que são usualmente um reflexo direto de uma ação do usuário ou do ambiente físico sob o qual o dispositivo se encontra.

Uma das principais características dos sensores de movimento é sua representação por pontos que possuem ordem temporal. Dessa forma, cada sensor de movimento retorna um vetor $T = \{t_i\}_{i=1}^N$, $t_i \in \mathbb{R}^3$ que representa uma série temporal, em que t_i é um valor temporal. Essas séries são medições em três direções, expressas em valores reais. Por

exemplo, a Figura 2-2 mostra os dados do sensor acelerômetro correspondentes a execução das atividades andar e correr.

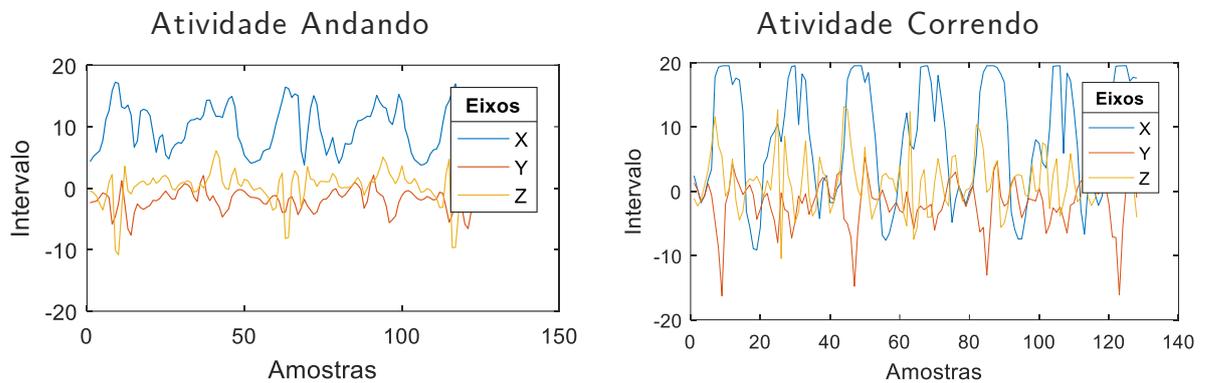


Figura 2-2: Exemplo de dados do sensor de acelerômetro para as atividades andar e correr.

Outra característica dos sensores de movimento é a taxa de amostragem. A taxa de amostragem é a quantidade de pontos capturados por segundo pelo sensor e possui um papel fundamental no processo de RAH, pois a escolha de taxas muito baixas pode não ser suficiente para representar as atividades. Em contrapartida, escolher taxas muito altas pode gerar problemas no método utilizado devido ao volume grande de informações. As taxas mais utilizadas no problema de RAH variam entre 20 a 100 Hz (Shoaib, et al., 2015).

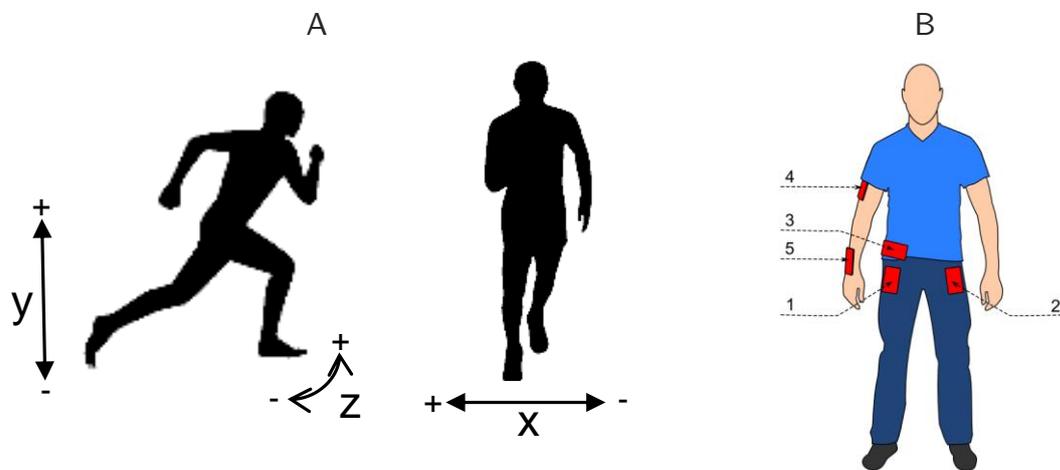


Figura 2-3: A) Representação de cada um dos eixos do acelerômetro relativos ao usuário. B) Principais posições do smartphone em um usuário comum. Adaptado de Shoaib, et al. (2014).

Os smartphones costumam estar em diferentes partes do corpo do usuário quando utilizados no dia-a-dia (Shoaib, et al., 2014). Isso significa que diferentes tipos de sinais são coletados quando uma mesma atividade é executada por causa da posição no qual o dispositivo se encontra. Por causa dessa variedade, cada um dos eixos do acelerômetro ou giroscópio desempenha um papel importante para cada tipo de atividade. Por exemplo, em um cenário em que o smartphone esteja próximo à perna do usuário, o

eixo z captura os movimentos dianteiros da perna, o eixo y captura os movimentos de cima e baixo e o eixo x os movimentos horizontais, como mostrado na Figura 2-3A.

Outra informação relevante é que os dados coletados podem variar de acordo com a posição no qual se encontra o dispositivo. Os lugares mais utilizados são mostrados na Figura 2-3B:

1. Localizado no lado direito da calça;
2. Localizado no lado esquerdo da calça;
3. Fixado no cinto da calça, na altura da cintura e próximo à perna esquerda;
4. Localizado no lado direito do braço;
5. Localizado no pulso.

Depois de selecionar os sensores e fixar a taxa de amostragem, os dados adquiridos estarão em um formato bruto e sem nenhum tipo de processamento, podendo ainda conter ruídos. Além disso, podem ser de natureza contínua ou discreta. Entretanto, independentemente da natureza dos dados, eles devem ser processados e segmentados para posterior utilização (Banos, et al., 2014). Essa etapa de pré-processamento deve converter os dados em um formato legível e compatível com os métodos que serão utilizados posteriormente.

2.2. Pré-processamento e segmentação de sinais

Em geral, os dados brutos coletados pelos sensores de movimento precisam ser processados, convertidos e adaptados em um formato homogêneo para que possam ser utilizados por etapas futuras da aplicação de RAH. Uma informação importante é que esses sensores coletam dados vistos como pontos ordenados no tempo e, por isso, são tratados como séries temporais. A manipulação dos dados dos sensores como séries temporais permite que técnicas de pré-processamento comuns a essas séries possam ser aplicadas (Bagnall, et al., 2016) (Schäfer, 2016). Dessa forma, é possível definir as nomenclaturas utilizadas no decorrer desta pesquisa:

Pela definição, uma série temporal é uma sequência de n valores reais, $n \in \mathbb{R}^1$, que são registrados ao longo do tempo, conforme mostra a equação

$$T = (t_1, \dots, t_n), t_i \in \mathbb{R}^1 \quad \text{Equação 2-1}$$

Uma base de dados de séries temporais D é uma coleção de N séries temporais:

$$D = (T_1, \dots, T_N), \quad |DS| = N \quad \text{Equação 2-2}$$

Uma subsequência de uma série temporal é uma série temporal que omite alguns valores de uma série temporal maior, mas não muda a ordem dos valores restantes. Dada uma

série temporal T , uma subsequência S é uma série temporal com w valores contínuos começando em uma posição i em T :

$$S(i, w) = (t_i, \dots, t_{i+w-1}), \quad \text{com } 1 \leq i \leq n - w + 1 \quad \text{Equação 2-3}$$

Uma subsequência no intervalo i pode ser inferida de seu predecessor $i - 1$ por uma soma e subtração:

$$S(i, w) = S(i - 1, w) + x_i - x_{i-w}, \quad \text{para } i > 1 \quad \text{Equação 2-4}$$

Após definir os dados dos sensores como séries temporais, o próximo passo corresponde à segmentação das séries. A segmentação é a tarefa de divisão das séries temporais em segmentos apropriados para o tipo de problema estudado. Por exemplo, qual o tamanho de segmento mais adequado para reconhecer as atividades humanas? Alguns trabalhos costumam tratar o pré-processamento e a segmentação como um só problema, aplicando métodos que minimizam os ruídos causados pelo ambiente, como limpeza e normalização da base de dados. Por último, os sinais são divididos em partes menores, chamadas de segmento. Esse último consiste na divisão do sinal (ou série temporal) em janelas de tempo com um determinado tamanho.

O processo de segmentação dos sinais dos sensores é também um tema de estudo no RAH. A Figura 2-4 apresenta um exemplo de técnica de segmentação. A primeira é a técnica de segmentação é mais simples, sem sobreposição nos dados. Um dos motivos para a adotá-la está em sua simplicidade em termos de implementação e processamento, o que a torna ideal para aplicações em tempo real. Uma das desvantagens é que os dados associados a uma característica particular ou atividade podem ser divididos entre as janelas, o que pode resultar em perda de informações importantes.

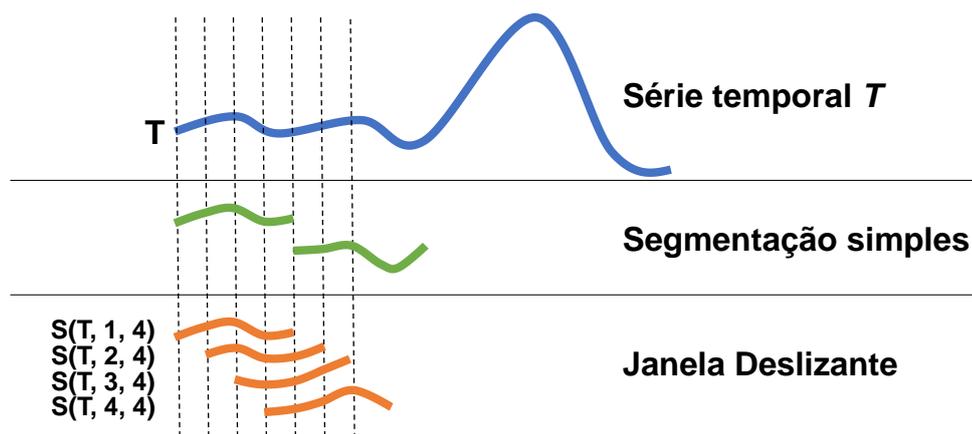


Figura 2-4: Aplicação da técnica de segmentação simples e sem sobreposição de janelas na série temporal T . Logo, abaixo, a técnica de segmentação com sobreposição de janelas de 75% aplicada na série T .

A segunda técnica é a de janela deslizante comumente adotada em RAH (Banos, et al., 2014). A abordagem baseada em janela deslizante tem se mostrado útil no

reconhecimento das atividades dinâmicas (e.g. andar e correr) e atividades estáticas (e.g. sentar e deitar).

Nesse contexto de atividades é possível associar o conceito de subsequência de séries temporais com o segmento de dados dentro da janela temporal. As subsequências de tamanho fixo w podem ser extraídas de cada posição na série temporal T (Figura 2-4). Por exemplo, uma janela deslizante sl é obtida através de um deslocamento de posição, com $1 \leq sl \leq n - w$, e a subsequência $S(T, i, w)$ é extraída da posição i . Há um total de $\frac{n-w}{sl} + 1$ subsequências em T :

$$janela(T, sl, w) = \bigcup_{i=0}^{\frac{n-w}{sl}} S(T, i * sl + 1, w) \quad \text{Equação 2-5}$$

De forma intuitiva, escolher segmentos ou subsequências de tamanhos menores permite a detecção de atividades mais rapidamente, pois a espera para se montar o segmento é menor e a necessidades de recursos necessários no processo também é reduzida (Banos, et al., 2014). Por outro lado, utilizar segmentos maiores permite que atividades mais complexas sejam reconhecidas, no entanto, será necessária uma carga adicional de tempo para se montar e processar o segmento. Por esse motivo, a qualidade da segmentação também está relacionada com a qualidade da classificação de atividades.

Todavia, um dos desafios dessa etapa é definir qual o tamanho ideal dos segmentos de tal forma que as atividades sejam reconhecidas rapidamente com uma boa precisão. Uma variedade de taxas de amostragem e de tamanhos de segmentos tem sido utilizada na literatura, conforme mostram (Shoaib, et al., 2014). Os trabalhos apresentados por (Banos, et al., 2014) (Kwapisz, et al., 2011) (Wu, et al., 2012) (Bao & Intille, 2004) mostram bons resultados, com reconhecimento acima dos 90% com o uso dessas técnicas de segmentação no RAH, com variação da taxa de amostragem entre 20Hz e 100Hz e tamanho de segmento que varia de 1 segundo até 10 segundos. Nesta pesquisa, o processo de segmentação consiste na divisão do sinal (série temporal) em janelas de tamanhos iguais e com tempos iguais, sem espaço entre as janelas ou sobreposição entre as janelas.

2.3. Fusão e agregação de dados

Um dos objetivos da fusão de dados é combinar múltiplas fontes de dados para obter uma informação de maior qualidade e mais relevante (Nakamura, et al., 2007). Em muitos casos, os dados são provenientes de múltiplas fontes, possivelmente heterogêneas, o que implica em um processo fundamental e não-trivial que impacta

diretamente no comportamento da aplicação. Uma instância da fusão de dados utilizada em muitos contextos é a agregação. A agregação de dados visa a obtenção de um conjunto menor de dados a partir de múltiplas fontes (Nakamura, et al., 2007). É a capacidade de sumarizar. Algumas funções de sumarização conhecidas, como a média, realiza a redução dos dados manipulados.

No contexto de RAH, aplicar técnicas de fusão de dados visa diminuir o problema da grande quantidade de dados que os múltiplos sensores geram a todo instante. A ideia básica é obter uma única fonte que represente os sinais de entrada, conforme mostrado na Figura 2-5

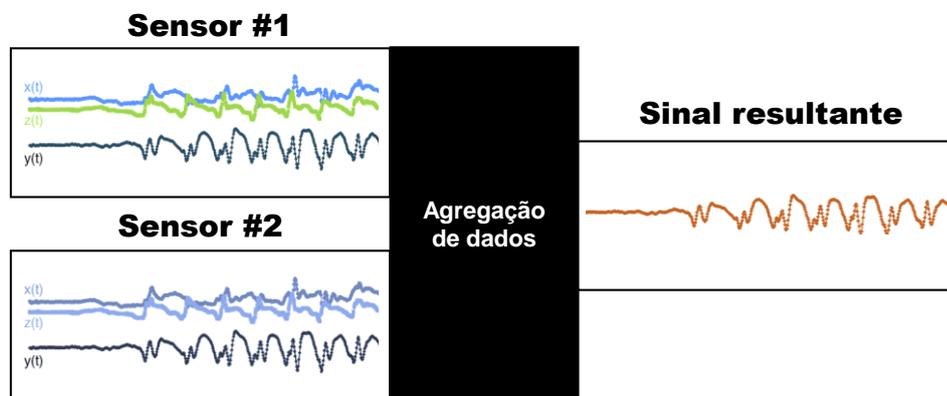


Figura 2-5: Exemplo da fusão dos sinais de dois sensores x_1 e $x_2 \in \mathbb{R}_3$, em um único sinal $r_1 \in \mathbb{R}$.

O sinal resultante deve ser representativo o suficiente para permitir que os métodos utilizados posteriormente consigam extrair os padrões relevantes do sinal. Esse processo necessita de algoritmos de baixo custo computacional porque os smartphones possuem recursos computacionais limitados, como energia, processamento e capacidade de armazenamento que devem ser levados em consideração (Gravina, et al., 2017). A escolha do método de fusão inadequado pode introduzir perda de informações, e por consequência pode ocasionar problemas nas próximas etapas do processo, prejudicando o desempenho da aplicação (Nakamura, et al., 2007).

É possível categorizar a fusão de dados conforme o nível de abstração dos dados: baixo nível, nível intermediário e alto nível (Nakamura, et al., 2007) (Gravina, et al., 2017).

No nível mais baixo, os dados brutos são combinados em um novo conjunto de dados mais preciso (e sem ruídos) que os dados originais tomados individualmente (Nakamura, et al., 2007). Nesse grupo estão inseridas funções como a magnitude, a Análise de Componentes Principais (PCA), o filtro de média móvel e o filtro de *Kalman*.

A Análise de Componentes Principais consiste na utilização de técnicas de redução de dimensionalidade de dados, e pode funcionar como uma técnica de agregação de dados,

por exemplo. O PCA busca características distintas que separam os dados de maneira não-supervisionada, mas seu foco é em utilizar menos informação para estas tarefas. As informações obtidas a partir dessa redução de dimensionalidade permitem visualizar estruturas complexas utilizando um espaço dimensional mais simples. Além disso, pode servir como entrada para métodos da aprendizagem supervisionada (Seção 2.5). Em resumo, o objetivo do PCA é a representação dos dados em uma dimensão menor (Duda, et al., 2000).

Outra técnica mais simples utilizada como agregação de dados é a magnitude, que é comumente utilizada no contexto RAH para tolerar mudanças na orientação de um smartphone. Os algoritmos de classificação que são treinados utilizando apenas uma orientação ou orientações específicas passam por problemas de desempenho quando a orientação do dispositivo muda (Sun, et al., 2010). Dado um segmento do sinal do acelerômetro, o objetivo da magnitude é avaliar o grau de intensidade de movimento com base em limiares derivados dos picos de aceleração (Khusainov, et al., 2013). Isso é possível devido à magnitude destacar as variações de sinais causadas pela fusão de valores entre diferentes coordenadas dos sensores inerciais. A Equação 2-6 mostra como calcular a magnitude de sinal em dados extraídos de sensores inerciais cujas coordenadas são x , y e z :

$$M(X) = \sqrt{x^2_i + y^2_i + z^2_i} \quad \text{Equação 2-6}$$

No nível intermediário, a fusão de dados ocorre considerando o conjunto de características. Basicamente, as características são extraídas a partir de múltiplas fontes de dados e um vetor de alta dimensão é criado para representar os dados de entrada. Nesse nível de fusão, é comum aplicar métodos de seleção de características mais relevantes para que métodos de aprendizagem de máquina possam ser aplicados na etapa de classificação.

No nível mais alto, a fusão de dados ocorre na camada de decisão. Ele toma como entrada decisões de diferentes tipos de modelos de classificação e as combina para obter uma decisão global mais precisa (Nakamura, et al., 2007). A saída desse nível é uma única decisão obtida de decisões locais. Métodos de inteligência artificial e de sistemas especializados podem ser utilizados nesta categoria, como por exemplo, os *ensembles*. Nenhum método desta categoria é aplicado nesta pesquisa.

2.4. Extração de características dos sinais

O processo de extração de características tem sido o núcleo dos trabalhos encontrados na literatura de RAH. Por esse motivo, esses trabalhos são categorizados conforme a natureza das características extraídas do sinal. Uma categorização proposta por Figo, et al. (2010) é apresentada na Figura 2-6, que se baseia na extração manual de características (do inglês, *Handcraft Features - HF*). Essas características são divididas em três domínios: tempo, frequência e discreto.

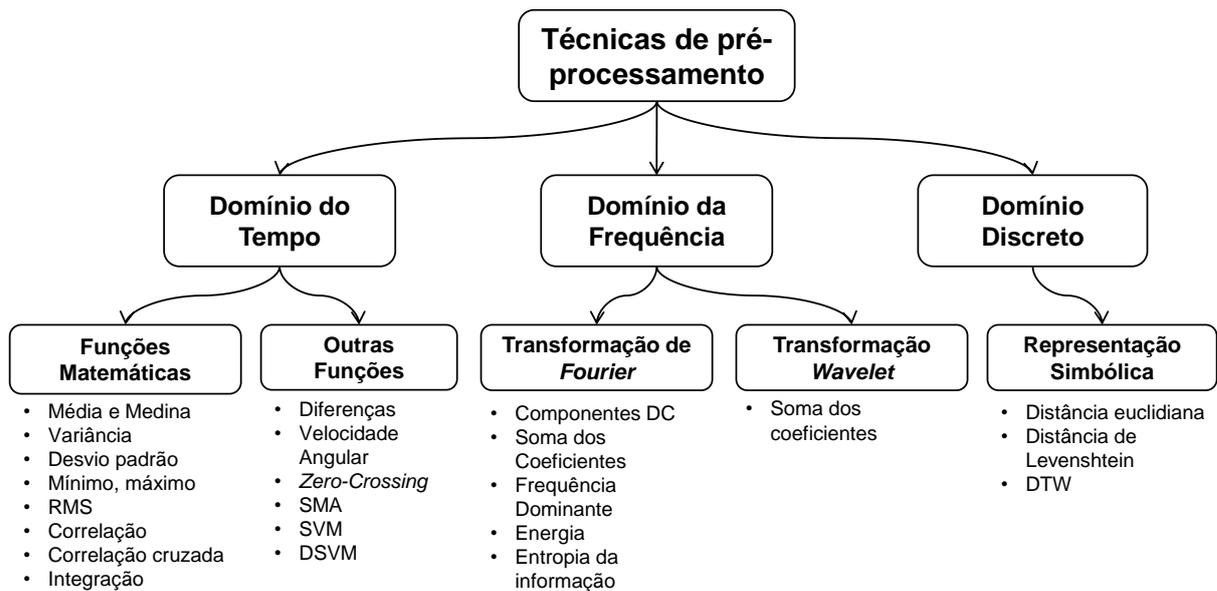


Figura 2-6: Categorização baseada na natureza das características extraídas dos sinais dos sensores inerciais proposta por (Figo, et al., 2010).

Para cada domínio existem funções ou métodos específicos que transformam os sinais dos sensores inerciais em representações de dados que são utilizadas por algoritmos de aprendizagem de máquina para gerar modelos de reconhecimento de atividades (Anguita, et al., 2013) (Figo, et al., 2010). Essas funções ou métodos extraem uma série de informações estatísticas e não-estatísticas dos sinais, como média, mediana e desvio padrão. As características estatísticas são de fácil entendimento e implementação, além de possuírem baixa complexidade computacional (Figo, et al., 2010).

A Tabela 2-1 mostra o conjunto de fórmulas das características do domínio do tempo mapeadas na literatura. Todas as características são aplicadas nos eixos x_i, y_i e z_i dos sensores inerciais de um smartphone.

Tabela 2-1: Principais características do domínio do tempo utilizadas pelas abordagens de extração manual de características. Todas as características da tabela são utilizadas nesta pesquisa.

Características no domínio do tempo			
Característica	Fórmulas	Característica	Fórmulas
Mínimo	$Min(X)$	Média	$\frac{1}{n} \sum_{k=0}^n x_k$
Máximo	$Max(X)$	Média absoluta	$\frac{1}{n} \sum_{k=0}^n x_k $
Amplitude	$ Max(X) - Min(X) $	Média quadrada	$\sqrt{\frac{1}{n} \sum_{k=0}^n x_k^2}$
Soma	$\sum_{k=0}^n x_k$	Desvio médio absoluto	$\frac{1}{n} \sum_{k=0}^n x_k - Media(X) $
Soma absoluta	$\sum_{k=0}^n x_k $	Soma quadrada de erro	$\sum_{k=0}^n (x_k - Media(X))^2$
Energia	$\sum_{k=0}^n x_k^2$	Variância	$\frac{1}{n} \sum_{k=0}^n (x_k - Media(X))^2$
Energia (Log)	$\sum_{k=0}^n \log(x_k^2)$	Skewness	$\frac{\frac{1}{n} \sum_{k=1}^n (x_k - Mean(X))^3}{(\frac{1}{n} \sum_{k=1}^n (x_k - Mean(X))^2)^{\frac{3}{2}}}$
Norm. Euclidiana	$\sqrt{\sum_{k=0}^n x_k^2}$	Kurtosis	$\frac{\frac{1}{n} \sum_{k=1}^n (x_k - Mean(X))^4}{(\frac{1}{n} \sum_{k=1}^n (x_k - Mean(X))^2)^2} - 3$
Power	$\frac{1}{n} \sum_{k=0}^n x_k^2$	Bin	$\bigcup_{i=1}^n \frac{i * (Max(X) - Min(X))}{n}$
Área	$\sum_{k=0}^n x_k + y_k + z_k$	Média absoluta da magnitude do sinal	$\frac{1}{n} \sum_{k=0}^n (x_k + y_k + z_k)$
Área absoluta	$\sum_{k=0}^n x_k + y_k + z_k $	Função da diferença de magnitude média (AMDF)	$\frac{1}{N} \sum_{k=0}^{N-1} x_k - x_{k+N} $
Desvio padrão	$\sqrt{\frac{1}{n} \sum_{k=0}^n (x_k - Media(X))^2}$	Amplitude de pico	$Max(N) - Min(N)$
Coefficiente de Pearson	$\frac{\frac{1}{n} \sum_{k=0}^n (x_k - Media(X))^2}{\frac{1}{n} \sum_{k=0}^n x_k}$	Magnitude média do sinal	$\frac{1}{n} \sum_{k=0}^n \sqrt{x_k^2 + y_k^2 + z_k^2}$
Zero crossing	$\frac{1}{2n} \sum_{k=0}^n sign(x_k) - sign(x_{k-1}) $		
Percentil	$P(p) = \left\lceil \frac{p}{100} * n \right\rceil, p = kth \text{ value}$		
Correlação	$\frac{n(\sum_{k=1}^n x_k y_k) - (\sum_{k=1}^n x_k)(\sum_{k=1}^n y_k)}{\sqrt{n(\sum_{k=1}^n x_k^2) - (\sum_{k=1}^n x_k)^2} + \sqrt{n(\sum_{k=1}^n y_k^2) - (\sum_{k=1}^n y_k)^2}}$		
Autocorrelação	$\frac{\sum_{k=1}^{n-1} (x_k - Mean(X))(x_{k+1} - Mean(X))}{\sum_{k=1}^n (x_k - Mean(X))^2}$		
Correlação cruzada	$\frac{\sum_{k=1}^n (x_k - Mean(X))(y_k - Mean(Y))}{\sqrt{\sum_{k=1}^n (x_k - Mean(X))^2 \sum_{k=1}^n (y_k - Mean(Y))^2}}$		

As características baseadas no domínio da frequência, apresentadas na Tabela 2-2 são usadas para capturar repetições naturais através da decomposição da série temporais em um conjunto de valores reais e imaginários, que representam componentes de ondas. As componentes são obtidas através da transformada de Fourier e transformada

Wavelet. As repetições dos sinais se relacionam com a periodicidade de uma determinada atividade, como por exemplo, andar ou correr. Esse conjunto é utilizado em muitos trabalhos da literatura (Figo, et al., 2010) (Anguita, et al., 2013) (Shoaib, et al., 2014). A Tabela 2-2 apresenta as principais características do domínio da frequência utilizadas pelas abordagens de extração manual de características.

Tabela 2-2: Principais características do domínio da frequência utilizadas pelas abordagens de extração manual de características. Todas as características desta tabela são utilizadas nesta pesquisa.

Domínio da Frequência	Fórmulas
Energia	$\sum_{k=1}^n x^2$
Energia normalizada	$\sum_{k=1}^n \left(\frac{x_k}{\sum_{k=1}^n x_k}\right)^2$
Power	$\frac{1}{n} \sum_{k=1}^n x^2$
Centroid	$\frac{\sum_{k=1}^n x_k * n}{\sum_{k=1}^n x_k}$
Entropia	$-\sum_{k=1}^n F_n \log F_n$, tal que $F_n = \frac{x_k}{\sum_{k=1}^n x_k}$
Componente dominante	$\frac{1}{n} \sum_{k=0}^n x_k$
Soma dos Coeficientes	$\sum_{k=0}^n x_k$

Por fim, Figo, et al. (2010) apresentam os primeiros resultados no domínio RAH utilizando as características pertencentes ao domínio discreto. Essas características representam o sinal dos sensores em uma sequência de símbolos, através de um processo de discretização, permitindo que dados sejam compactados em representações menores que as representações originais. Após o processo de discretização, duas séries temporais discretizadas podem ser comparadas através de uma métrica de distância como, por exemplo, a distância euclidiana, a *Dynamic Time Warping* (DTW) ou a distância de *Levenshtein*.

As abordagens de extração manual de características utilizam principalmente as medidas baseadas no domínio do tempo e da frequência, pois a partir dessas funções é possível obter um conjunto de características que discriminam bem as atividades executadas. As características são utilizadas por métodos de aprendizagem de máquina supervisionada, cujo objetivo é identificar padrões nos dados a partir de um conjunto de amostras e gerar um modelo de classificação para reconhecer as atividades por meio da previsão de amostras desconhecidas (Duda, et al., 2000) (Cichosz, 2014). Os métodos de aprendizagem de máquina serão apresentados na próxima seção.

2.5. Algoritmos de aprendizagem supervisionada

Um algoritmo de aprendizagem pode detectar automaticamente padrões em um conjunto de dados e pode ser usado para tomar decisões em situações de incerteza (Murphy, 2012). Podemos dividir os algoritmos de aprendizagem em: algoritmos supervisionados, não-supervisionados, aprendizagem reforçada e semi-supervisionada. O objetivo de cada uma dessas técnicas é derivar de um conjunto de dados, de forma indutiva, um modelo capaz de classificar novos dados dedutivamente. Antes explicar qualquer detalhe sobre esses algoritmos, é importante definir algumas informações sobre a entrada e saída do conjunto de dados envolvidos no processo (Cichosz, 2014):

- Domínio: o domínio, designado por X , é o conjunto de todas as entidades que são consideradas em uma dada tarefa da aprendizagem indutiva. Nesse contexto, é representado por uma base de dados contendo atividades executadas por indivíduos.
- Instâncias: uma instância é qualquer elemento único do domínio, $x \in X$, ou seja, é um segmento que representa uma atividade.
- Atributos: um atributo é uma função $a: X \rightarrow A$ que assina um valor para cada instância no domínio. A princípio, pode-se assumir que há n atributos definidos no domínio X , $a_1: X \rightarrow A_1$, $a_2: X \rightarrow A_2, \dots$, $a_n: X \rightarrow A_n$. Os atributos representam as características extraídas dos sinais dos sensores. Podem ser do tipo:
 - Nominal: possui um número finito de valores discretos sem relação de ordem total;
 - Ordinal: possui um número finito de valores discretos com relação de ordem total;
 - Contínuo: possui valores numéricos.
- Classe, rótulo ou atributo alvo: em problemas de classificação e regressão, subclasses da aprendizagem supervisionada, um atributo é escolhido para ser a classe (ou rótulo) da instância. Ele representa uma propriedade da instância que o modelo criado por um método da aprendizagem de máquina deve ser capaz de prever, baseado nos outros atributos. Problemas que possuem classes são chamados de supervisionados e os problemas que não possuem classes associadas são chamados de não-supervisionados.
- Atributos de entrada: todos os atributos que não pertencem ao conjunto de classes (ou rótulos) são considerados atributos de entrada.
- Conjunto de treinamento: as informações de treino são um subconjunto do domínio X , sendo indispensável para qualquer problema classificação. O conjunto de treinamento $DS \subseteq D$ é o conjunto de instâncias atualmente usadas para a criação

do modelo, onde D denota o conjunto de todas as instâncias disponíveis (treino e teste). Se existe uma classe para um dado problema de classificação, assume-se que seus valores são conhecidos em D , e as instâncias disponíveis estão rotuladas. Nesse contexto, cada segmento de atividade possui associado expressamente um rótulo que corresponde a atividade que está sendo executada.

- **Modelo:** O objetivo da aprendizagem supervisionada é gerar, baseado no conjunto de treinamento disponível, um modelo h representando o conhecimento que pode ser aplicado a todas as instâncias $x \in X$ de maneira automática. Ele é uma função que toma uma instância como argumento e retorna uma predição da classe à qual a instância deve estar associada.
- **Desempenho:** a qualidade das predições de um modelo é chamada de desempenho do modelo. O que se espera de um modelo é que ele apresente um desempenho verdadeiro, ou seja, que ele preveja corretamente todas as novas instâncias em todo o domínio, incluindo instâncias não vistas anteriormente.
- **Generalização:** o verdadeiro desempenho é também chamado de desempenho de generalização, que é a essência da aprendizagem indutiva. O modelo deve prever bem novos dados de entrada, e para isso deve abranger generalizações apropriadas dos padrões detectados no conjunto de treinamento.
- **Superajuste ou Sobreajuste** (do inglês, *Overfitting*): uma baixa capacidade de generalização leva ao superajuste. Um modelo h é considerado superajustado para um conjunto de treinamento DS se existe outro modelo h' para a mesma tarefa que possui um desempenho pior que h no conjunto de treinamento, mas apresenta um desempenho melhor em todo o domínio.

A aprendizagem de máquina supervisionada, utilizada neste trabalho, precisa de uma base de dados rotulada. Seu objetivo é mapear entradas x para saídas y , dado um conjunto rotulado de dados:

$$D = \{(x_i, y_i)\}_{i=1}^N \quad \text{Equação 2-7}$$

Aqui D é todo o conjunto, e N é o número de instâncias, x_i é um vetor de características (conjunto de entradas), e y_i são os rótulos ou classes. Neste tipo de aprendizagem os dados podem ser visualizados no formato de tabela contendo uma sequência de atributos, ou características, finalizada com uma classe (Murphy, 2012), conforme mostra a Tabela 2-3.

Tabela 2-3: Exemplo do vetor de características de uma base de dados rotulada.

Instância	Atributo x_1	Atributo x_2	...	Atributo x_n	$y =$ Classe
1	0.78	0.06	0.42	0.03	Classe A
2	0.80	0.20	0.40	0.10	Classe B

...
n	0.84	0.01	0.41	0.15	Classe C

Na literatura existem diversos algoritmos de aprendizagem supervisionada, como por exemplo, árvore de decisão, *Naive Bayes*, máquina de vetor de suporte (SVM), redes neurais artificiais (ANN), regressão logística, KNN (*K-Nearest Neighbors*), entre outros. Os detalhes destes algoritmos podem ser vistos em (Duda, et al., 2000) (Barber, 2012) (Murphy, 2012) (Cichosz, 2014) (Witten, et al., 2016).

Esta pesquisa adota o algoritmo KNN no processo de classificação. O KNN é um algoritmo utilizado em problemas de classificação, pertencente à família de algoritmos de aprendizagem supervisionada. É um dos algoritmos de classificação mais simples em termos de implementação e entendimento. Uma de suas características é que esse algoritmo não aprende um modelo, da forma como o algoritmo SVM ou a árvore de decisão, por exemplo. Em vez disso, guarda todas as instâncias de treinamento para utilizá-las posteriormente na rotulagem de novos dados, funcionando como uma tabela de busca.

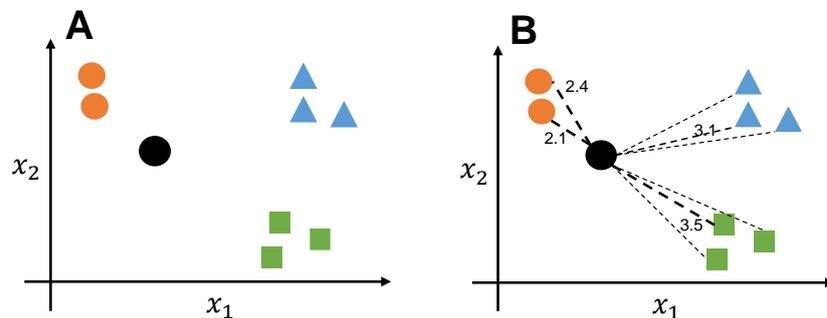


Figura 2-7: A- representação das instâncias rotuladas em um plano. A instância preta não está rotulada. B- cálculo da distância entre a instância preta e todas as outras instâncias

O objetivo do KNN é buscar o vizinho ou o conjunto de vizinhos (k) mais próximos de uma nova instância não-rotulada. Para classificar essa nova instância não-rotulada, basta associar o rótulo do vizinho mais próximo à instância não-rotulada, utilizando alguma medida de similaridade (Figura 2-7). No caso de divergências, ou seja, quando no conjunto dos k -vizinhos mais próximos existem diferentes rótulos, utiliza-se alguma regra de seleção (e.g. voto de maioria).

O cálculo de similaridade entre as instâncias do conjunto de treinamento e a nova instância de entrada é realizado de acordo com a escolha de uma métrica de distância. Tabela 2-4 mostra as métricas mais utilizadas: a métrica de distância euclidiana, manhatan, minkowski e similaridade entre cossenos.

Tabela 2-4: métricas de distância que podem ser utilizadas no algoritmo KNN.

Funções de Distância	Fórmula
Euclidiana	$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$
<i>Manhattam</i>	$\sum_{i=1}^k x_i - y_i $
<i>Minkowski</i>	$\left(\sum_{i=1}^k (x_i - y_i ^q) \right)^{1/q}$
Similaridade entre cossenos	$\frac{\sum_{i=1}^k x_i y_i}{\sqrt{\sum_{i=1}^k (x_i)^2} \sqrt{\sum_{i=1}^k (y_i)^2}}$

Dado um inteiro positivo k (adota-se um número ímpar de preferência), e um novo dado de entrada x , e uma métrica de similaridade d , O algoritmo KNN funciona da seguinte forma:

- a. Percorrer toda a base de treinamento calculando os valores de distância *dist* entre o novo dado de entrada x e cada instância contida no treinamento. Essa pesquisa deve retornar as k instâncias mais próximas.
- b. Na sequência, estima-se a probabilidade condicional para cada uma das classes existentes no problema, associando a instância x à classe com maior probabilidade.

2.6. Considerações sobre o capítulo

O capítulo 2 apresentou as cinco fases que compõem a metodologia utilizada pelos trabalhos de RAH para classificação de atividades. Essa metodologia não é aplicada apenas no contexto RAH, mas é utilizada em todos os problemas que envolvem classificação. Entender cada processo é importante porque cada etapa possui uma parcela de contribuição no desempenho do método de classificação.

Dentre as cinco etapas, a extração de características é a que mais recebe destaque. Isso acontece porque os métodos de classificação são muito sensíveis aos dados do conjunto de treinamento. Isso quer dizer que um conjunto de características que não capture boas representações de cada atividade deve resultar também na criação de um modelo

de classificação ruim. Por essa razão, o objetivo de muitos trabalhos tem sido encontrar o melhor conjunto de características, de modo que o método de classificação obtenha um bom desempenho na etapa de classificação.

As diferentes abordagens para reconhecer atividades humanas

A necessidade de técnicas mais robustas para extrair características automaticamente dos sinais dos sensores tem impulsionado muitos trabalhos no domínio RAH. O objetivo é eliminar a participação de um especialista e melhorar o desempenho dos modelos de classificação. Além disso, aspectos relacionados a custo computacional, espaço e uso de bateria também são levados em consideração. A literatura apresenta algumas direções para resolver alguns dos problemas citados, mas ainda existem muitas lacunas em aberto. Nessa perspectiva, este capítulo apresenta métodos que podem ser utilizados para resolver as lacunas ainda não exploradas na literatura de RAH. Tais métodos carecem de experimentos e resultados que mostram a sua eficácia no domínio RAH.

Este capítulo apresenta soluções para três problemas apontados pela literatura de RAH. O primeiro problema está relacionado com a extração manual de características, que é dependente da tarefa ou aplicação em questão e não pode ser transferida para atividades ou padrões similares. Isso quer dizer que se uma nova atividade for inserida no contexto, é possível que o modelo sofra uma degradação em seu desempenho por não possuir as características que representam a nova atividade.

O segundo problema é chamado de variabilidade intraclasse (Bulling, et al., 2014), é também encontrado nas abordagens de extração manual de características. A variabilidade intraclasse ocorre porque a mesma atividade pode ser executada de formas diferentes por indivíduos diferentes, mas sistemas RAH não consegue aprender esse tipo de padrão. Também pode ocorrer uma situação em que o indivíduo também execute uma mesma atividade de forma diferente, devido fatores emocionais ou físicos, como estresse ou fadiga.

Por fim, o terceiro problema caminha no sentido oposto da variabilidade intraclasse, chamado de similaridade interclasse (Bulling, et al., 2014). A similaridade interclasse acontece quando existem classes que são diferentes, mas que possuem características similares nos dados dos sensores, levando o sistema a achar que duas atividades na verdade representam a mesma atividade.

A necessidade de encontrar características mais representativas que resolvem os problemas citados acima tem incentivado a comunidade científica no emprego de métodos mais avançados para descobrir informações úteis sem a intervenção humana. Para isso, são utilizados novos algoritmos capazes de transformar os dados dos sensores em diferentes domínios de representação de forma automática, eliminando parte dos problemas relacionados às abordagens de extração de características manuais.

3.1. Extração automática de características

No novo cenário proposto nesta pesquisa, o processo de extração de características pode ser categorizado em dois grupos baseados na extração manual e aprendizagem automática de características. Esta categorização consiste em uma extensão da categorização proposta por Figo, et.al., (2010).

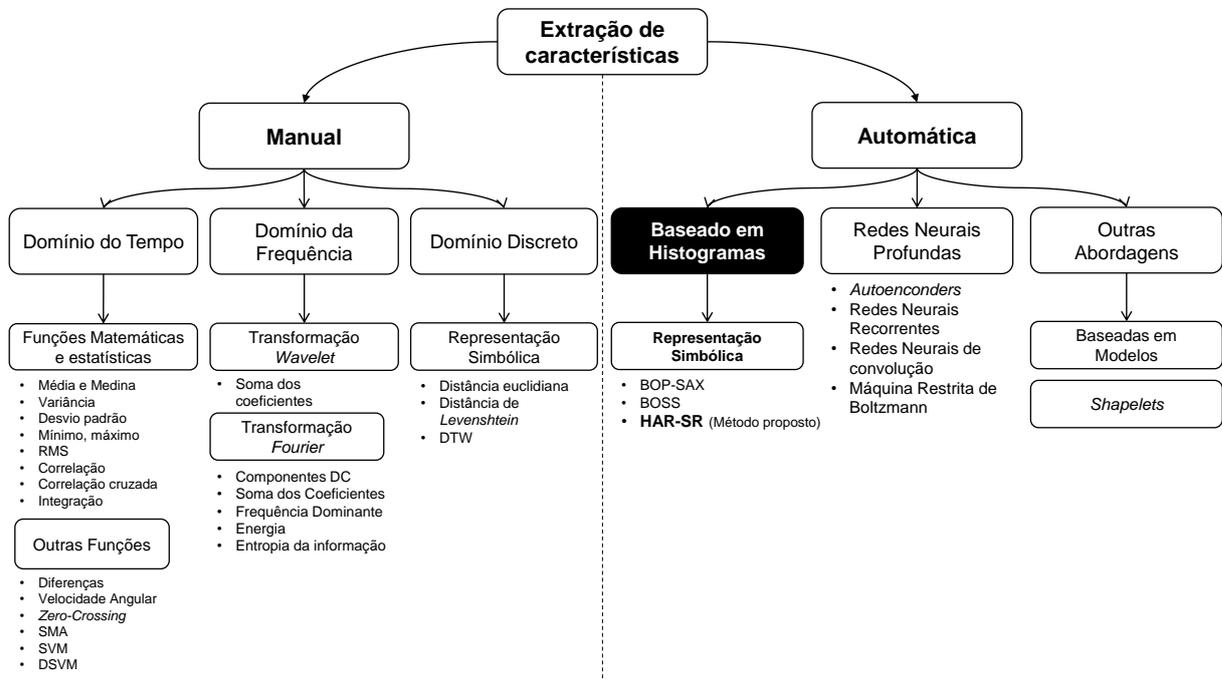


Figura 3-1. A nova categorização proposta separa os trabalhos em abordagens baseadas na extração manual e automática de características.

A Figura 3-1 mostra a nova categorização que divide os trabalhos em dois grandes grupos: a abordagem baseada na extração manual de características (HC) e a

abordagem baseada na extração automática de características (do inglês, *feature learning*). No grupo das características manuais estão contidas as características dos domínios de tempo, frequência e discreto. O grupo das características automáticas incluem os métodos que utilizam outras metodologias que de alguma forma extraem padrões automaticamente dos sinais, e que dispensam o esforço de um especialista.

A necessidade de técnicas mais robustas para extrair características discriminativas dos dados de maneira automática tem sido o foco de muitos trabalhos (Senin & Malinchik, 2013) (Ravi, et al., 2016) (Ronaldo & Cho, 2016) (Ignatov, 2018) (Lima, et al., 2018). Uma das primeiras classes de algoritmos a propor soluções automatizadas para o problema RAH foram as redes neurais profundas (do inglês, *Deep Neural Network - DNN*) (Nweke, et al., 2018). Os métodos baseados em DNN eliminam a necessidade de uma fase dedicada à extração de características, pois o todo o processo é realizado dentro das camadas ocultas da rede.

As DNN fazem uso de múltiplas camadas de abstração e, por isso, aprendem as representações intrínsecas nos dados de entrada automaticamente. Esses métodos são diferentes dos métodos de aprendizagem tradicionais como, por exemplo, as máquinas de vetores de suporte, KNN ou árvores de decisão, que precisam de um conjunto de características previamente definidas. Os métodos de DNN que têm sido utilizados no contexto de RAH são: Máquinas de Boltzmann Restritas, *Autoencoders*, *Sparse Coding*, Redes Neurais de Convolução (CNN) e Redes Neurais Recorrentes (RNN), sendo os dois últimos os métodos mais comuns (Nweke, et al., 2018)(Wang, et al., 2018).

Para cada problema existe um subtipo de algoritmos com modificações em suas estruturas internas que permitem detectar os mais variados tipos de padrões. No entanto, todo esse poder e flexibilidade possui um custo associado. O APÊNDICE A apresenta uma discussão sobre o custo computacional das DNNs e o porquê esses métodos podem não ser apropriados para dispositivos com baixo poder computacional (CPU e GPU de baixo desempenho) (He & Sun, 2014). Além disso, cada arquitetura é dependente de configurações específicas em seus parâmetros, onde pequenas modificações na arquitetura podem significar grandes mudanças na complexidade (Sze, et al., 2017). O estudo detalhado das DNN foge do escopo desta pesquisa, mas pode ser encontrada em (Raví, et al., 2017) (Nweke, et al., 2018) (Wang, et al., 2018).

A nova categorização proposta apresenta uma nova classe de métodos automáticos que podem ser aplicados no domínio RAH: os métodos baseados em histogramas, conforme destacado na Figura 3-1. Os métodos baseados em histogramas ou dicionários (*codebook*) usam as contagens de frequência de padrões recorrentes e, em seguida, constroem classificadores com base nos histogramas resultantes. Essa é uma abordagem

recente que tem mostrado bons resultados em problemas envolvendo classificação de séries (Schäfer, 2016) (Bagnall, et al., 2016) (Shirahama & Grzegorzek, 2017) (Montero Quispe, et al., 2018).

As abordagens baseadas em *shapelets* buscam encontrar padrões curtos que definem uma classe, mas que podem aparecer em qualquer parte da série. Esses padrões independentes de fase são comumente chamados *shapelets*. A classe é então distinguida pela presença ou ausência de um ou mais *shapelets* em algum ponto da série inteira. As abordagens baseadas em modelo (*model based*) ajustam um modelo generativo a cada série e medem a similaridade entre séries usando similaridade entre modelos (Bagnall, et al., 2016).

3.2. Por que utilizar métodos de discretização?

Os algoritmos de representação simbólica são uma categoria de algoritmos capazes de extrair características de forma automática (do inglês, *feature learning*). Essa classe de algoritmos permite o desenvolvimento de aplicações de baixo custo para smartphones, que devem impactar no tempo de processamento, ocupação de espaço na memória e consumo de energia das baterias. No entanto, a literatura de RAH pouco tem explorado as características do domínio discreto.

Em geral, as séries temporais do mundo real são ruidosas e podem possuir alta dimensionalidade. O problema da alta dimensionalidade de uma série temporal, por exemplo, afeta os algoritmos de classificação degradando seu desempenho em termos de velocidade, conforme sua dimensionalidade aumenta (Lin, et al., 2003). Os métodos de discretização ou representação simbólica resolvem diversos problemas no contexto de séries temporais, dentre eles: redução de dimensionalidade, redução de ruídos e redução de numerosidade (Lin, et al., 2003) (Schäfer, 2016) (Bagnall, et al., 2016).

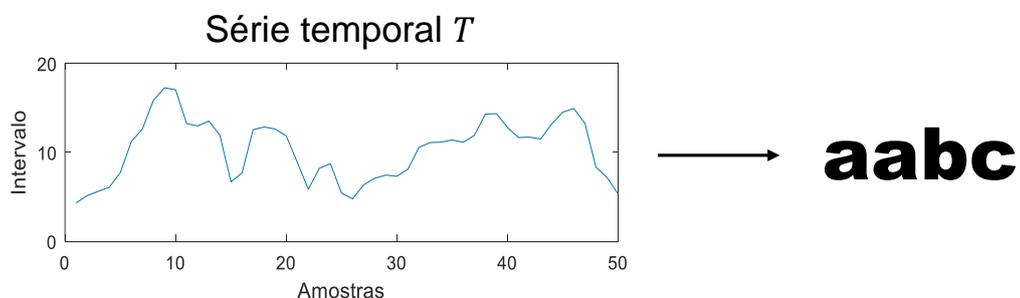


Figura 3-2. Uma série temporal com 50 amostras pode ser representada pela sequência “aabc” através da utilização de um algoritmo de discretização.

De forma resumida, o processo de discretização consiste em transformar uma série temporal, representada por valores reais, em um símbolo ou sequência de caracteres (Figura 3-2), dispostos em um domínio discreto e em um espaço muito menor que o original. A combinação dos algoritmos de discretização com técnicas de segmentação de séries temporais resulta em uma forma mais representativa e completa da série temporal. Essa nova forma de extrair padrões gera um conjunto símbolos a partir de uma única série temporal e sem a necessidade de intervenção humana, na qual o método aprende as melhores características do sinal.

Para se obter o conjunto de símbolos a partir de uma série temporal, os segmentos da série são obtidos através da técnica de janela deslizante, onde cada segmento é convertido em uma nova representação discreta, conforme mostra a Figura 3-3.

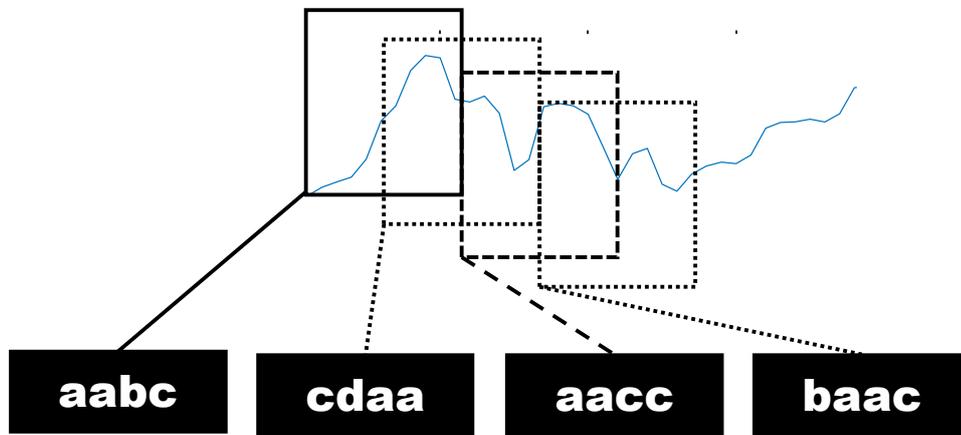


Figura 3-3: Aplicação da técnica de janela deslizante em uma série temporal. Cada segmento é transformado em uma sequência de símbolos. O exemplo mostra quatro segmentos e suas respectivas representações simbólicas.

No fim do processo o conjunto de símbolos é representado por um histograma de frequência de símbolos. Um histograma é uma representação gráfica responsável por agrupar os dados para se obter uma distribuição de frequências que, nesse caso, são as palavras. A Figura 3-4 mostra um exemplo de um histograma cujo espaço de amostras contém 16 símbolos. A distribuição de frequência dos símbolos pode ser considerada como características que representam os dados brutos. Essas características podem ser utilizadas para treinar algoritmos de classificação para descoberta de novos padrões. O processo desde a segmentação até a geração do histograma dá origem a uma nova classe de algoritmos que pode ser generalizada como modelos baseados em histogramas (ou dicionários), que resultam em representações discretas de séries temporais (Bagnall, et al., 2016).

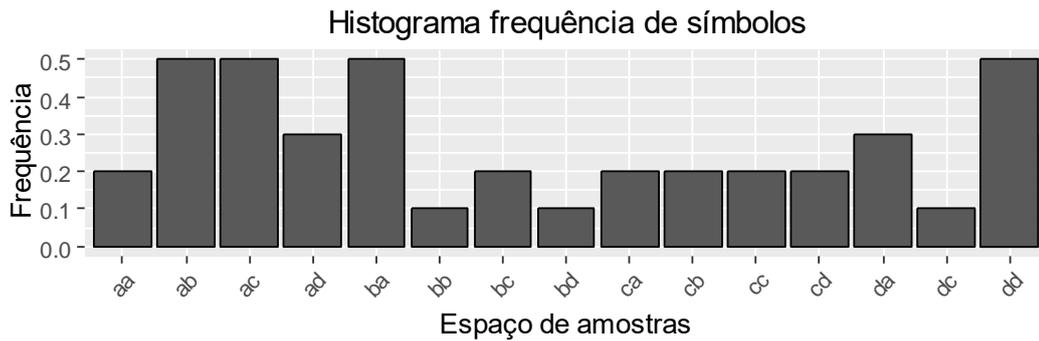


Figura 3-4: Histograma de frequência de símbolos, que contém 16 símbolos no intervalos aa – dd. Note que a frequência é normalizada no intervalo [0, 1].

A partir do processo de discretização são propostos métodos mais complexos que geram características mais representativas dos dados e melhoraram a precisão dos modelos de classificação. Por exemplo, o método SAX-VSM (Senin & Malinchik, 2013) e o método BOSS-VS (Schäfer, 2015), aplicam modelo *tf – idf* (Manning, et al., 2008) na distribuição de frequência dos histogramas.

3.3. Discretização de séries temporais

Na literatura existem algoritmos que transformam séries temporais em um conjunto de padrões simbólicos, ou representações simbólicas. Dentre os algoritmos utilizados para este propósito, estão o *Symbolic Aggregate Approximation* (SAX) (Lin, et al., 2003), o *Symbolic Fourier Approximation* (SFA) (Schäfer & Höggqvist, 2012) e *Permutation Entropy* (PE) (Bandt & Pompe, 2002). Esses algoritmos criam representações discretas de séries temporais, no qual o conjunto de símbolos obtidos no processo de discretização pode ser interpretado como uma distribuição de frequência de símbolos que representa a série ao qual está associada.

Essa nova forma de representação possui algumas vantagens que são intrínsecas ao seu processo de transformação, como redução de dimensionalidade, redução de ruídos, representação simbólica e representação no domínio da frequência ou domínio do tempo.

Os algoritmos desse grupo, como o SAX e o SFA, são divididos em duas fases que são representadas por dois parâmetros:

- Aproximação: é aplicado no mapeamento da série temporal em um espaço de baixa dimensão, representado em um vetor de valores reais. Esse parâmetro consiste no tamanho da palavra $\omega \in \mathbb{N}$, associada à série temporal, que representa o número de valores para a aproximação. Quanto menor o tamanho

da palavra maior será a redução de ruídos, mas aumenta-se a perda de informação.

- Quantização: cada valor real obtido no processo de aproximação é mapeado para um valor discreto, que é interpretado como símbolo. Esse parâmetro é definido como o tamanho do alfabeto $\alpha \in \mathbb{N}$, que é usado na quantização. Um alfabeto pequeno também resulta em uma forte redução de ruído.

O processo de aproximação e quantização pode mudar em cada método, o que resulta em diferentes tipos de representações do histograma P para a mesma série temporal. Esses processos serão descritos na próxima seção para os algoritmos SAX e SFA.

3.3.1. O Algoritmo SAX

O SAX (do inglês, *Symbolic Aggregation Approximation*) é um algoritmo de discretização que transforma uma série temporal arbitrária de tamanho N em uma cadeia de símbolos arbitrários de tamanho M (Lin, et al., 2003).

Tabela 3-1: Sumário das notações usadas no SAX.

\mathbf{T}	Uma série temporal $\mathbf{T} = \mathbf{t}_1, \dots, \mathbf{t}_N$
$\bar{\mathbf{T}}$	Um PAA da série temporal $\bar{\mathbf{T}} = \bar{t}_1, \dots, \bar{t}_M$
$\hat{\mathbf{T}}$	Uma representação simbólica de uma série temporal $\hat{\mathbf{T}} = \hat{t}_1, \dots, \hat{t}_M$
$\omega \in \mathbb{N}$	É um parâmetro que define o tamanho da palavra. Representa o número de seguimentos do PAA
$\alpha \in \mathbb{N}$	É um parâmetro que define o tamanho do alfabeto (e.g. para o alfabeto = {a, b, c}, $\alpha = 3$)

Antes de iniciar a fase de aproximação, é comum aplicar a normalização padrão no sinal para se obter invariância de deslocamento/amplitude. Essa normalização, chamada de *z-score*, é aplicada subtraindo a média μ de uma série temporal T para obter a invariância de deslocamento e dividindo-a pelo desvio padrão σ para se obter invariância para diferentes amplitudes, conforme mostra a Equação 3-2:

$$z - \text{score}(T) = \frac{t_i - \mu}{\sigma}, i \in [1..n] \quad \text{Equação 3-1}$$

Na fase de aproximação, o algoritmo PAA (do inglês, *Piecewise Aggregate Approximate*) (Lin, et al., 2003) é responsável pela redução de dimensionalidade da série temporal.

Dada uma série temporal T de tamanho N é transformada em um vetor $\bar{T} = \bar{t}_1, \dots, \bar{t}_M$ de tamanho $M \leq N$ onde cada \bar{t}_i , é calculado de acordo com:

$$\bar{t}_i = \frac{M}{N} \sum_{j=\frac{N}{M}(i-1)+1}^{(N/M)i} t_j \quad \text{Equação 3-2}$$

Essa equação significa que para reduzir a dimensionalidade de N para M , primeiro deve-se dividir a série original em M partes de tamanhos iguais e então computar a média dos valores de cada uma das partes. O vetor desses valores se torna a representação reduzida dos dados, conforme mostra a Figura 3-5. A redução de dimensionalidade do PAA concorre com técnicas mais sofisticadas como Transformadas de Fourier e Wavelets (Schäfer & Höggqvist, 2012).

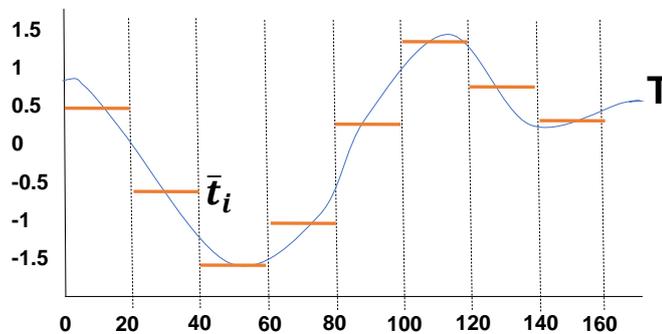


Figura 3-5: A aplicação do PAA em uma série temporal ($N=160$) resultando um novo sinal ($M=8$). O novo sinal é composto por todos os valores \bar{t}_i , que são calculados em cada intervalo.

	α							
β_i	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Figura 3-6. A tabela de pesquisa mostra os pontos de interrupções que dividem a distribuição gaussiana em um número arbitrário de regiões equiprováveis.

Na fase de quantização, uma nova transformação é realizada para obter a representação discreta de \bar{T} . Assumindo-se que a série temporal normalizada possui uma distribuição gaussiana (Larsen, et al., 1986) (Butler & Kazakov, 2015), determina-se os pontos de interrupções (do inglês, breakpoints) que irão produzir α áreas de tamanhos iguais na curva gaussiana. Os pontos de interrupção são listas ordenadas de números $\beta = \beta_1, \dots, \beta_{\alpha-1}$ tal que a área de uma curva gaussiana é $N(0,1)$, sendo de β_i até $\beta_{i-1} =$

$1/a$. Esses pontos podem ser obtidos através de uma tabela estatística pré-definida (Lin, et al., 2003), que funciona como uma tabela de pesquisa, conforme mostra a Figura 3-6.

Após a obtenção dos pontos de interrupções, uma série temporal pode ser discretizada da seguinte forma:

- i. Obtenção do PAA da série temporal e;
- ii. uma busca na tabela de pesquisa de modo que todos os coeficientes do PAA que estão abaixo do menor ponto de interrupção são mapeados para o símbolo 'a', todos os coeficientes maiores ou iguais ao menor ponto de interrupção e menor que o segundo ponto de interrupção são mapeados para o símbolo 'b', e assim por diante, conforme mostra a Figura 3-7.

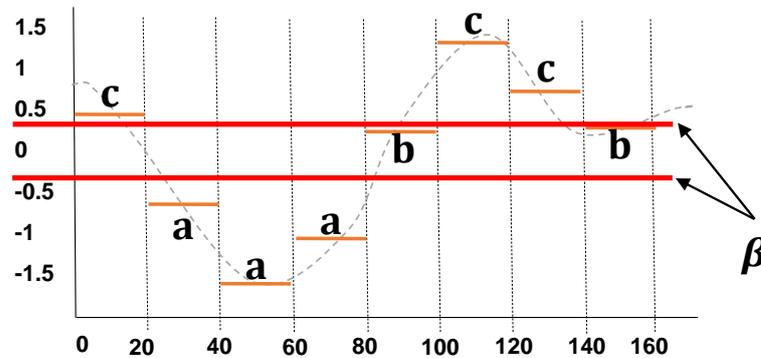


Figura 3-7. Uma série temporal é discretizada da seguinte forma: i) obter a representação em PAA; ii) utilizar os pontos de interrupção predeterminados para mapear os valores do PAA em símbolos do SAX. Na figura acima, $N = 160$, $M = 8$ e $a = 3$, a série temporal é mapeada para a palavra 'caaabccb'.

No exemplo os três símbolos 'a', 'b' e 'c' são aproximadamente equiprováveis. A concatenação de cada símbolo gerado será chamada de palavra SAX. O processo de geração de uma palavra SAX é considerado um processo simples e computacionalmente barato quando comparado com algoritmos mais robustos, como o SFA, apresentado na próxima seção.

3.3.2. O Algoritmo SFA

A Aproximação Simbólica de Fourier (do inglês, *Symbolic Fourier Approximation - SFA*) (Schäfer & Högvist, 2012) é um dos métodos mais sofisticado de discretização de séries temporais encontrados na literatura (Bagnall, et al., 2016). Em termos práticos, funciona de forma similar ao SAX, porém, na fase de aproximação utiliza a transformada de Fourier (Schäfer & Högvist, 2012) para obter \bar{T} em vez do PAA. Além

disso, sua tabela de pontos de interrupção β é dinâmica, que se adequa aos dados de um conjunto de treinamento diferente do SAX que utiliza uma tabela fixa e pré-definida. As mudanças citadas geram um custo adicional ao método e uma etapa adicional para geração da tabela β . Entretanto, esse custo adicional é compensado por resultados melhores na comparação com o SAX em diversos cenários (Bagnall, et al., 2015) (Schäfer, 2016).

O processo de transformação da série temporal T para o domínio discreto é dividido em duas fases:

- i. A fase de pré-processamento, que utiliza o conjunto de séries temporais disponíveis $D = \{T_1, \dots, T_n\}$ para realizar as operações de aproximação e discretização. O objetivo dessa fase é obter uma tabela de pontos de interrupção β dinâmica, de acordo com os padrões encontrados no conjunto D .
- ii. A fase de transformação utiliza a tabela MCB (*Multiple Coefficient Binning*) para transformar uma série temporal T de tamanho N em uma representação simbólica \hat{T} de tamanho M . A tabela MCB é conjunto de pontos de interrupção $\beta = \beta_1, \dots, \beta_{a-1}$, obtida na fase de pré-processamento.

Pré-processamento: A primeira operação da fase de pré-processamento é a aproximação. Essa operação utiliza a Transformada Discreta de Fourier (DFT) e a MFT (Transformada Momentânea de Fourier) para transformar as séries temporais $D = \{T_1, \dots, T_N\}$ em novas séries temporais no domínio da frequência $\bar{D} = \{\bar{T}_1, \dots, \bar{T}_M\}$ (Schäfer & Höggqvist, 2012). A Figura 3-8 mostra o processo de aproximação para uma série temporal utilizando dois coeficientes de Fourier. Os coeficientes de Fourier de ordem mais alta representam mudanças rápidas, como quedas ou ruídos no sinal. Esse processo funciona como um filtro passa-baixa usando os primeiros $M \ll N$ coeficientes de Fourier.

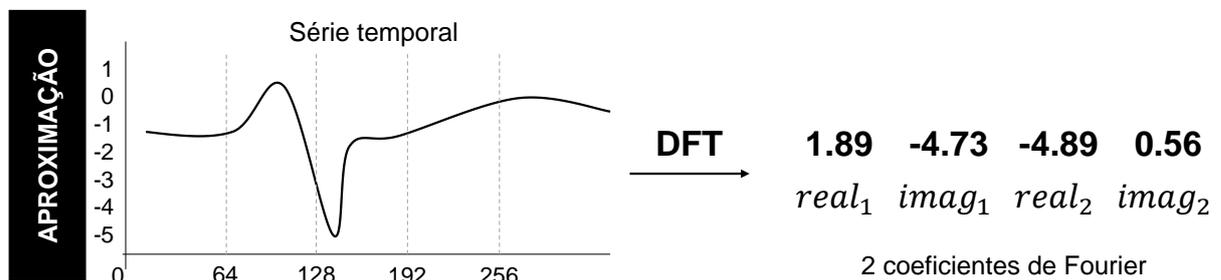


Figura 3-8: Processo de aproximação de uma série temporal T através da Transformada de Fourier.

A segunda operação da fase de pré-processamento é a quantização. Nesse processo, as discretizações são determinadas a partir das aproximações DFT obtidas na etapa de aproximação. Essa técnica, chamada de Armazenamento de Múltiplos Coeficientes (do inglês, *Multiple Coefficient Binning* - MCB), é responsável por gerar uma tabela de

consulta dinâmica β . O processo de quantização (Figura 3-9) faz o mapeamento de um coeficiente de Fourier ao seu respectivo segmento. O MCB se adapta aos dados e dessa forma minimiza a perda de informação introduzida no processo de quantização.

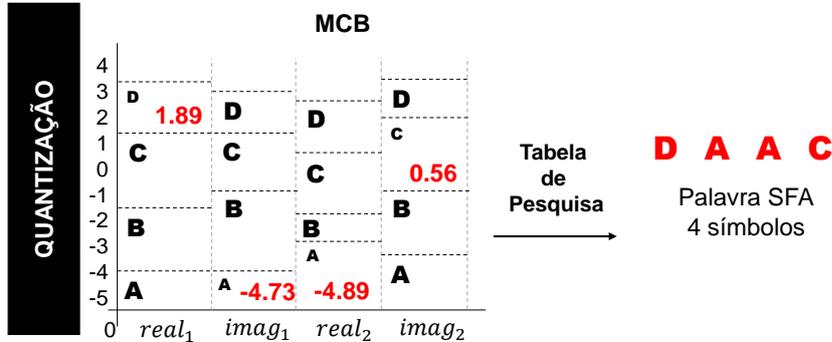


Figura 3-9: Processo de discretização utilizando a tabela de pesquisa MCB.

Transformação: A segunda fase do algoritmo SFA é a fase de aproximação e outra de discretização. O processo de aproximação é aplicado a uma série temporal T de tamanho N , no qual a DFT e a MFT são utilizadas para se obter o vetor de coeficientes de Fourier, resultando na série \bar{T} de tamanho M . A série \bar{T} é discretizada através de uma pesquisa na tabela MCB (obtida na fase de pré-processamento do SFA), resultando na representação simbólica \hat{T} .

3.3.3. Extração de subestruturas de séries temporais

Os métodos de discretização geram uma única representação discreta \hat{T}_m para uma série temporal T_n . Isso significa que ao discretizar uma série temporal, o SAX ou SFA devem gerar em uma única saída a representação simbólica, conforme mostra a Figura 3-10.

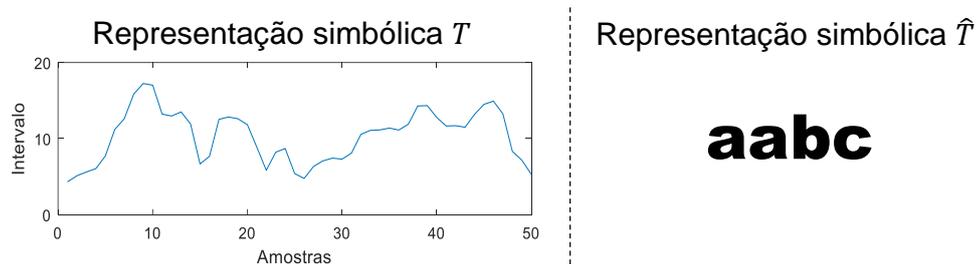


Figura 3-10: Exemplo da representação de uma série temporal em uma sequência de símbolos.

No entanto, é possível tornar o processo de discretização mais robusto. A combinação entre os algoritmos de representação simbólica e técnicas de segmentação permite que o SAX ou SFA extraiam subestruturas e outros tipos de padrões mais representativos a partir da mesma série temporal. O resultado dessa combinação passa a ser um

conjunto de símbolos (distribuição de frequências) e não mais uma única representação \hat{T} , conforme mostra a Figura 3-11.

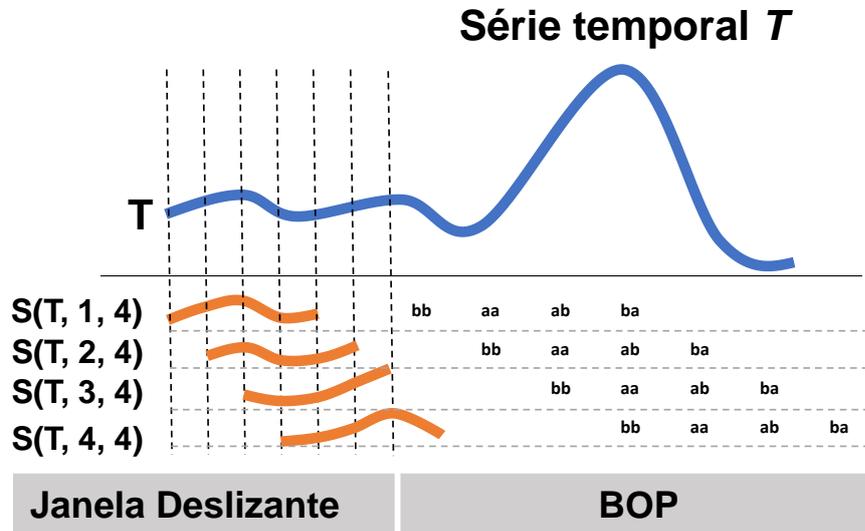


Figura 3-11. Transformação de uma série temporal T em um histograma de frequência de símbolos utilizando o modelo BOP. A série temporal é segmentada através da técnica de janela deslizante, e então discretizada por algum algoritmo de representação simbólica. O resultado é histograma de frequência de símbolos.

A nova metodologia para encontrar padrões em séries temporais, chamada de *Bag-of-Patterns* (BOP) (Lin, et al., 2012), combina os algoritmos de representação simbólica com a técnica de janela deslizante, vista na Seção 2.2. O BOP é definido como:

$$\text{BOP}(T_n) = \hat{T}_1, \dots, \hat{T}_W, \text{ onde } W \text{ é o tamanho da janela deslizante} \quad \text{Equação 3-3}$$

O algoritmo extrai W segmentos da série temporal e sem seguida realiza a discretização de cada segmento através de um algoritmo de discretização, como o SAX ou SFA. Com essa etapa adicional no processo de discretização, os modelos baseados no BOP passam a possuir os seguintes parâmetros:

- O tamanho da janela de tempo $W \in \mathbb{N}$, que representa o tamanho dos segmentos. De forma intuitiva, o tamanho de W deve ser escolhido de modo que consiga capturar as estruturas presentes na série temporal.
- O tamanho de palavra $\omega \in \mathbb{N}$ e o tamanho do alfabeto $\alpha \in \mathbb{N}$.

Conforme mostra a Figura 3-11, o algoritmo começa extraindo as janelas deslizantes de tamanho W da série temporal. Em seguida, cada janela deslizante é normalizada para possuir média igual a zero e desvio padrão igual a 1 (invariância de amplitude). Por fim, a transformação simbólica é aplicada em cada um dos valores obtidos no processo de janela deslizante. A partir do conjunto de símbolos obtidos pelo BOP, um

histograma P é construído. Esse histograma de símbolos, não-ordenado, é conhecido em outros domínios como bag-of-words (BOW).

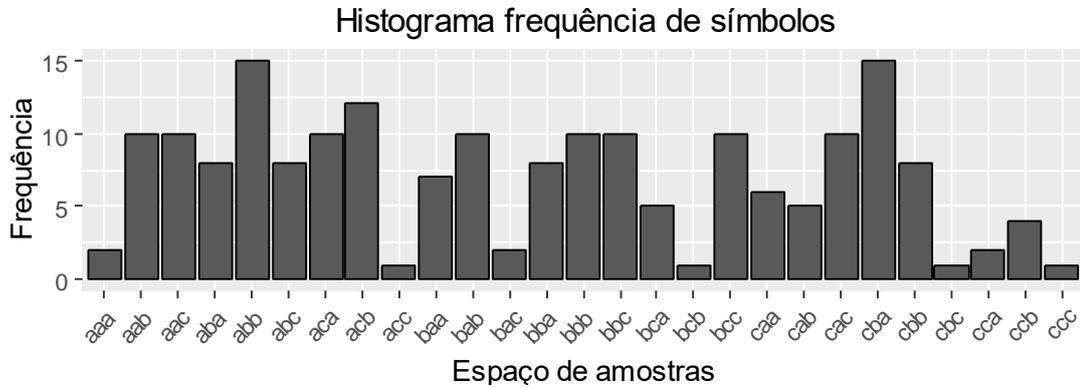


Figura 3-12: Histograma de frequência de símbolos obtido através do BOP. $\omega = 3$, $\alpha = 3$.

O histograma $P : \alpha^\omega \rightarrow \mathbb{N}$ é uma função do espaço de símbolos α^ω para números naturais. Os números representam as ocorrências de um símbolo, conforme mostra o exemplo da Figura 3-12.

O primeiro trabalho a utilizar esse conceito, chamado de BOP-SAX (Lin, et al., 2012) (Senin & Malinchik, 2013), utilizou o algoritmo SAX para discretizar cada segmento da série temporal obtido através da técnica de janela deslizante. O modelo BOSS (*Bag-of-SFA-symbols*) (Schäfer, 2015) é um modelo similar ao BOP-SAX, mas utiliza o algoritmo SFA para geração de palavras.

O histograma de frequências P pode ser tratado como um conjunto de características pelos métodos de classificação. A partir do mesmo histograma também é possível aplicar outras técnicas para se obter novas características de maior qualidade. Por exemplo, o método do modelo do espaço vetorial chamado de $tf - idf$ (Manning, et al., 2008) é utilizado para obter o vetor de frequências ponderadas (ou pesos) a partir do histograma P . Nesse espaço, as palavras que estão próximas são sintaticamente similares e as palavras mais afastadas são sintaticamente distantes. No modelo $tf - idf$ cada um dos símbolos recebe um peso de acordo com sua importância no contexto. Esse modelo consegue representar bem a relação sintática entre os símbolos e sua importância, pois o peso aplicado em cada símbolo gera um vetor de características melhor que o apresentado somente pela frequência de símbolos (Mikolov, et al., 2013). Adaptações desse modelo aplicados no domínio discreto são apresentadas em Shaeffer (2015), Shaeffer (2016), Lin, et al. (2012) e Senin e Malinchik (2013).

Uma nova forma de se extrair outro tipo de características do histograma P é apresentada nesta pesquisa. A nova abordagem aplica os conceitos da teoria da

informação no histograma P. As novas características são medidas que representam as séries temporais ao qual estão associadas, conforme apresentada no capítulo 4.

3.4. Trabalhos Relacionados

Esta seção descreve os principais trabalhos de reconhecimento de atividades humanas utilizando os sensores inerciais embutidos em smartphones. Os trabalhos estão divididos conforme a nova taxonomia proposta nesta pesquisa, baseando-se na natureza da extração de características. Dessa forma, os trabalhos serão divididos em:

- HCFM - Trabalhos baseados na extração de características manuais (do inglês, *handcraft features model*);
- AFEM - Trabalhos baseados na extração de características automáticas (do inglês, *automatic feature extraction model*);

3.4.1. Abordagens baseadas na extração manual de características

Para que um método de classificação de atividades consiga um bom desempenho, é essencial que os dados sejam convertidos em uma representação que capture as características mais representativas dos sinais dos sensores, pois os métodos de aprendizagem clássicos, como KNN e árvore de decisão não funcionam bem se aplicados diretamente nos dados brutos dos sensores (Dernbach, et al., 2012) (Kwapisz, et al., 2011). Por esse motivo, diversos trabalhos buscam formas de extrair uma representação que capture as características mais representativas desses sinais dos sensores. Os trabalhos apresentados nesta seção aplicam a metodologia de extração manual de características, chamada de HCFM (do inglês, *handcraft features model*).

Em uma importante revisão da literatura, Figo et al. (2010) apresentam uma avaliação das principais características de extração manual no domínio RAH. No total, foram analisadas quinze características no domínio do tempo, seis no domínio da frequência e três no domínio discreto. Tais características foram descritas e avaliadas de acordo com a taxa de acurácia e complexidade computacional. Em seus testes são utilizados dados de três atividades de locomoção (pular, correr e caminhar). Os resultados mostram que as características de cálculos simples como a média, mínimo, máximo e desvio padrão, pertencentes ao domínio do tempo, possuem baixo custo computacional e são eficazes, atingindo 97,3% de acurácia. As características no domínio da frequência apresentam

custo computacional mais elevado com menor desempenho de classificação, atingindo 70,1% de acurácia. Os resultados no domínio discreto atingiram 52% de acurácia, no entanto, o autor conclui que as características no domínio discreto são viáveis para aplicações em dispositivos móveis e que a literatura ainda não expressa soluções eficientes nesse domínio.

Kwapisz et al. (2011) disponibilizam a base de dados WISDM, que é utilizada em muitos trabalhos RAH. A WISDM contém seis classes de atividades (correr, caminhar, subir escadas, descer escadas, sentar e levantar) contendo informação do acelerômetro de 36 usuários. Os experimentos foram executados usando as características no domínio do tempo e os dados do sensor acelerômetro. A métrica de avaliação utilizada é a validação cruzada (10-folds). Os resultados mostram uma acurácia de 85,1% para árvore de decisão (J48), 78,1% para o Regressão Logística e 91,7% para Redes Neurais.

Siirtola et al. (2011) apresentaram um método cuja principal contribuição está na combinação das características nos três domínios usando dez tipos de características no domínio do tempo, frequência e uma no domínio discreto. As novas características, denominadas SAXS (*Symbolic Aggregate approXimation Similarity*) se baseiam no cálculo de similaridade entre palavras obtidas através do algoritmo SAX (*Symbolic Aggregate approXimation*) (Lin, et al., 2007). Os autores também utilizam um modelo de referência para cada atividade, para que a similaridade entre o modelo de referência e uma nova instância de atividade seja calculada. O método proposto foi avaliado em cinco base de dados em diferentes contextos, como classificação de atividades cujo indivíduo está utilizando determinada ferramenta, reconhecimento de gestos, estilos de natação e atividades esportivas. Os resultados demonstram uma acurácia de 83,8% na atividade de natação, 84,3% para gestos, 85,7% e 82,96% na utilização de ferramentas e 83,26% na prática de esportes. Um dos pontos negativos desse trabalho é que os autores não utilizam uma base de dados comum aos trabalhos do domínio RAH para efeitos de comparação.

Anguita et al. (2013) disponibilizam a base de dados UCI-HAR, que contém dados dos sensores acelerômetro e giroscópio extraídos de 30 usuários que realizaram 06 atividades físicas (deitar, caminhar, subir escadas, descer escadas, sentar e levantar). No total, 561 características foram extraídas no domínio do tempo e frequência. O modelo de classificação gerado pelo SVM obteve uma acurácia de 96%. A métrica de validação utilizada é a de *hold-out*.

Shoaib et al. (2013) utilizaram sete classificadores para reconhecer seis atividades físicas coletadas de quatro posições do corpo diferentes: cintura, bolso esquerdo, bolso direito, braço e pulso. No total, foram extraídas 24 características dos sinais de acelerômetro,

giroscópio e magnetômetro. O melhor resultado é representado pela acurácia de 98% quando o smartphone está na posição da cintura. Os resultados mostraram que o acelerômetro e o giroscópio são complementares no processo de reconhecimento de atividades. Além disso, o giroscópio apresenta bom desempenho, cerca de 90%, quando utilizado sozinho.

Micucci et al. (2017) disponibilizam uma nova base de dados pública para o domínio RAH, que contém dados de atividades físicas e dados que representam quedas. Para reconhecer os 09 tipos de atividades físicas, foram extraídas 453 características do sinal do acelerômetro e utilizados quatro classificadores: o KNN, SVM, ANN e *Random Forest*. No melhor cenário, seu método obteve 88% de acurácia utilizando o método *Random Forest* (5-fold CV) e 73% utilizando o método RF com a métrica de validação cruzada por indivíduo. Os resultados obtidos pelos autores mostram que os classificadores tendem a perder desempenho quando utilizam a métrica de validação cruzada por indivíduo. Esse resultado também mostra que os métodos com maiores acurácias que utilizam a validação cruzada tradicional nem sempre possuem bons resultados em um domínio real, pois são sobreajustados para o conjunto na qual foram treinados.

Em geral, os trabalhos apresentados na literatura baseados na extração manual de características utilizam sempre o mesmo conjunto de características apresentados na Seção 2.4. Terminado o processo de seleção e extração de características, um *framework*, chamado WEKA (Witten, et al., 2016), que agrega técnicas de pré-processamento e métodos de aprendizagem de máquina é comumente utilizado para testar os classificadores.

Tabela 3-2: Resumo dos trabalhos de reconhecimento de atividades em smartphones utilizando sensores de movimento que utilizam abordagens baseadas na extração manual de características.

Autores	Abordagem	Base de dados	Nº de características	Método	Avaliação	Acurácia
(Kwapisz, et al., 2011)	HCFM	WISDM	43	ANN	10-fold CV	91.7 %
				RL		78.1 %
				DT		85.1 %
(Siirtola, et al., 2011)	HCFM	-	40-144	SAXS + KNN	10-fold CV	86,7 % (média)
(Anguita, et al., 2013)	HCFM	UCI	561	SVM	Hold-out	96.37 %
(Shoaib, et al., 2013)	HCFM	SHOAIB	24	ANN	10-fold CV	90% (média)
(Micucci, et al., 2017)	HCFM	UniMib SHAR	453	KNN	5-fold CV	87.77 %
				SVM		81.62 %

		(A-9)		ANN		72.13 %
				RF		88.41 %
(Micucci, et al., 2017)	HCFM	UniMib	453	KNN	LOSO	63.79 %
		SHAR		SVM		63.32 %
		(A-9)		ANN		63.63 %
				RF		73.17 %

3.4.2. Abordagens Baseadas na extração automática de características

A qualidade dos algoritmos de classificação é dependente da qualidade das características usadas. A literatura mostra uma área emergente que busca soluções para encontrar padrões automaticamente nos sinais de sensores. Esse novo horizonte para o RAH deu origem a um novo grupo de trabalhos baseados na extração de características automáticas, denominados AFEM (do inglês, *automatic feature extraction models*);

Kolosnjaji e Eckert (2015) utilizam um tipo de redes neurais profundas, o *autoencoder*, para extração automática de características do sensor de acelerômetro. Esse método é muito utilizado para redução de dimensionalidade e compressão de dados. Primeiro, os dados são pré-processados utilizando a transformada rápida de Fourier (FFT) e, em seguida, são alimentados na rede com quatro camadas. O método de classificação utilizado é o *Random Forest* (RF), e a métrica de validação utilizada nos experimentos é a validação cruzada e a validação cruzada por indivíduo. Seus melhores resultados são de 85.35% para a base WISDM e 76.26% para a base UCI.

Ronao e Cho (2016) utilizam redes de convolução profunda (CNN) nos dados brutos dos sensores para extrair características automaticamente. A acurácia do seu método (*convnet*) é de 94.79%. Os autores também apresentam resultados melhores quando a transformada rápida de Fourier (FFT) é aplicada nos sinais dos sensores, cuja acurácia obtida é de 95.75%.

No trabalho de Ravi et al. (2016), os sinais dos sensores são convertidos para o domínio do espectrograma e introduzidos em uma rede neural profunda. A extração de características é realizada por um método de convolução temporal. Essa abordagem possibilita que as características aprendidas sejam invariantes a mudanças em diferentes propriedades. Os resultados obtidos para a base WISDM utilizando a métrica de validação 10-fold CV apresentam acurácia de 98.2%.

Ignatov (2018) utiliza uma rede neural de convolução (CNN) em combinação com outras características estatísticas clássicas para extrair características dos sinais dos sensores de um smartphone. Essa metodologia permite que a CNN encontre características locais dos sinais enquanto as medidas estatísticas (HC) preservem informações sobre a forma global do sinal. Seus resultados apresentam acurácia de 93.32% para a base WISDM e 97.63% para a base UCI.

Um dos primeiros trabalhos que utiliza representação simbólica para reconhecer atividades baseadas em smartphones é descrito por (Figo, et al., 2010). Este trabalho utilizou o algoritmo SAX no processo de discretização dos sinais do sensor acelerômetro. No entanto, seu melhor resultado em termos de acurácia ficou em 50.5% para três atividades (correr, andar e pular). O segundo cenário foi constituído de duas atividades (andar e correr) e apresentou uma acurácia 84,17% utilizando a métrica DTW. Os autores deixam como trabalhos futuros a investigação de técnicas de pré-processamento para melhorar os resultados em um cenário com mais de duas atividades.

Lima et al. (2018) apresentam uma análise comparativa dos algoritmos de representação simbólica, como o SAX-VSM, BOSS MODEL, BOSS-VS e WEASEL e a abordagem tradicional de extração manual (HCFM) com características no domínio do tempo e frequência. Seus melhores resultados mostram que os algoritmos de representação simbólica conseguem extrair padrões automaticamente dos dados com alta taxa de acurácia, obtendo 100% de acurácia quando a métrica de validação utilizada é a 10-fold CV. Isso mostra que os algoritmos de representação simbólica apresentam uma capacidade de especialização excelente na base de treinamento, superando tanto os métodos clássicos de extração manual quando os trabalhos estado da arte apresentados na literatura.

Tabela 3-3: Resumo dos trabalhos de reconhecimento de atividades em smartphones utilizando sensores de movimento que utilizam abordagens baseadas na extração automática de características.

Autores	Abordagem	Base de dados	Método	Avaliação	Acurácia
(Kolosnjaji & Eckert, 2015)	AFEM	WISDM	Autoencoder + ANN	LOSO	85.36 %
			Autoencoder + RF		83.46 %
(Kolosnjaji & Eckert, 2015)	AFEM	UCI	Autoencoder + ANN	LOSO	76.26 %
			Autoencoder + RF		77.81 %
(Ronao & Cho, 2016)	AFEM	UCI	CNN	Hold-out	94.79 %
(Ronao & Cho, 2016)	AFEM	UCI	FFT + CNN	Hold-out	95.75 %
(Ignatov, 2018)	AFEM	WISDM	CNN + HF	Hold-out	93.32 %
(Ignatov, 2018)	AFEM	UCI	CNN + HF	Hold-out	97.63 %

(Lima, et al., 2018)	HCFM	UCI	SAX-VSM	10-fold CV	95.7 %
			BOSS		100 %
			BOSS-VS WEASEL		98.5 %
					52.9 %
(Lima, et al., 2018)	HCFM	SHOAIB	SAX-VSM	10-fold CV	91.3 %
			BOSS		100 %
			BOSS-VS WEASEL		99.9 %
					48.4 %
(Lima, et al., 2018)	HCFM	WISDM	SAX-VSM	10-fold CV	75.1 %
			BOSS		100 %
			BOSS-VS WEASEL		94.4 %
					59.9 %

3.4.3. Discussão sobre os trabalhos relacionados

Os trabalhos que utilizam as características do domínio do tempo e frequência (HCFM) em geral constroem um modelo de classificação utilizando os métodos de aprendizagem de máquina clássicos (e.g. árvores de decisão, SVM e KNN). As abordagens AFEM utilizam métodos mais robustos para extrair características automaticamente dos dados.

O método apresentado nesta pesquisa utiliza algoritmos de representação simbólica para aprender os padrões característicos de cada indivíduo em um processo de *feature learning*, como nas redes neurais. Além disso, o espaço de características utilizada pelo algoritmo de aprendizagem de máquina é pequeno, porém representativo o suficiente para gerar modelos simples e com boa acurácia na classificação de atividades.

Quanto aos métodos de avaliação, a literatura possui três grupos. No primeiro grupo, os trabalhos dirigem seus esforços para criar métodos que são especialistas em reconhecer padrões de um mesmo indivíduo, chamados de métodos personalizados. No segundo grupo, os trabalhos buscam um método capaz de generalizar o reconhecimento de padrões para qualquer indivíduo. O terceiro grupo é misto, pois utiliza a validação cruzada nos dados de mais de um indivíduo, mas não se preocupa em separar dados de um mesmo indivíduo no conjunto de treino e teste.

A maioria dos estudos em RAH usa o tipo misto, que é a validação cruzada tradicional (10-fold CV) para avaliar o desempenho dos algoritmos. Na validação personalizada, o algoritmo é treinado com apenas um indivíduo e avaliado com o mesmo indivíduo. Os dados desse indivíduo são divididos entre treino e teste, em que ambos possuem

amostras tipicamente similares. Nesse caso, não há problema caso o conjunto de treino possua amostras similares às encontradas no conjunto de teste. Isso explica porque os algoritmos de classificação possuem acurácias maiores nessa métrica de validação, já que o modelo de classificação foi treinado com instâncias que são muito similares às encontradas do conjunto de teste.

Para construir um método genérico em cenários que envolvem grupos de indivíduos, a validação cruzada tradicional (*k-fold*) pode não ser a melhor solução. Um dos problemas encontrados na sua utilização é que os segmentos pertencentes a um determinado indivíduo estejam presentes tanto no conjunto de treino como no de teste. Isso faz com que o modelo de treinamento conheça parte do conjunto de teste. No contexto de RAH, isso significa que o modelo treinado utilizando a validação cruzada poderá conhecer os padrões de atividade de um indivíduo específico que está presente no conjunto de teste, obtendo maiores acurácias na classificação. Nesse cenário, essa situação pode levar ao Sobreajuste (do inglês, *Overfitting*) do método, que é caracterizado por uma baixa capacidade de generalização.

Para minimizar o problema da baixa capacidade de generalização, um segundo método de avaliação pode ser utilizado: a validação cruzada por indivíduo (do inglês, *Leave-one-subject-out Cross Validation* - LOSO). A LOSO leva a informação do indivíduo no cálculo, impedindo que o conjunto de treinamento e de testes sejam viciados por padrões de um mesmo indivíduo. Dessa forma, o número N de partições utilizadas na validação cruzada deve corresponder ao número de indivíduos no conjunto. Esse método de validação é um indicador do potencial de generalização do modelo, sem a necessidade de obtenção de dados do indivíduo antes da utilização do sistema (Morales & Akopian, 2017).

É possível notar na Tabela 3-2 que os resultados obtidos através da avaliação LOSO são menores se comparado com os resultados obtidos com a validação cruzada tradicional (*10-folds*) na mesma base de dados. Isso acontece porque a avaliação LOSO não generaliza da mesma forma que o 10-fold CV. Em alguns casos observados na Tabela 3-3, o conjunto de treinamento é separado em duas partes: 1/3 são separados como dados de testes e 2/3 para treinamento. Os algoritmos apresentam um bom desempenho de classificação, mas esse tipo de avaliação também pode ser pouco representativo para avaliar o método se comparado com o LOSO.

No capítulo 5 são apresentados os resultados obtidos através do LOSO e da validação cruzada tradicional na comparação do método proposto com trabalhos da literatura. Os resultados apresentados devem mostrar a importância da escolha correta da métrica

de validação para obtenção de modelos de classificação de atividades generalistas e personalizados.

A comparação e avaliação dos trabalhos relacionados serviram como uma fonte para descobrir quais são os principais desafios encontrados no RAH. As soluções mais recentes são um ponto de partida para novas propostas e melhorias aos métodos já existentes. Um dos principais desafios no momento está na busca por um limiar entre métodos que aprendem os padrões dos dados de forma automática com uma boa precisão no reconhecimento de atividades e ainda apresentem uma baixa complexidade computacional nesse processo, obedecendo as limitações de hardware e software dos smartphones.

CAPÍTULO

4

Reconhecimento de atividades baseado em representação simbólica

O método HAR-SR (do inglês, *Human Activity Recognition based on Symbolic Representation*), é um método de classificação de séries temporais proposto nesta pesquisa que é aplicado no contexto RAH. O HAR-SR utiliza outras interpretações da medida de complexidade estatística com base nos histogramas de frequências de cada classe do problema. As novas interpretações são usadas como características pelo algoritmo de classificação. No decorrer deste capítulo será detalhado todo o funcionamento do método HAR-SR e como pode ser aplicado no reconhecimento de atividades humanas. O HAR-SR utiliza métodos de aproximação e discretização de séries temporais, combinando técnicas da teoria da informação, com o objetivo de obter características relevantes para classificação das atividades.

4.1. Visão Geral

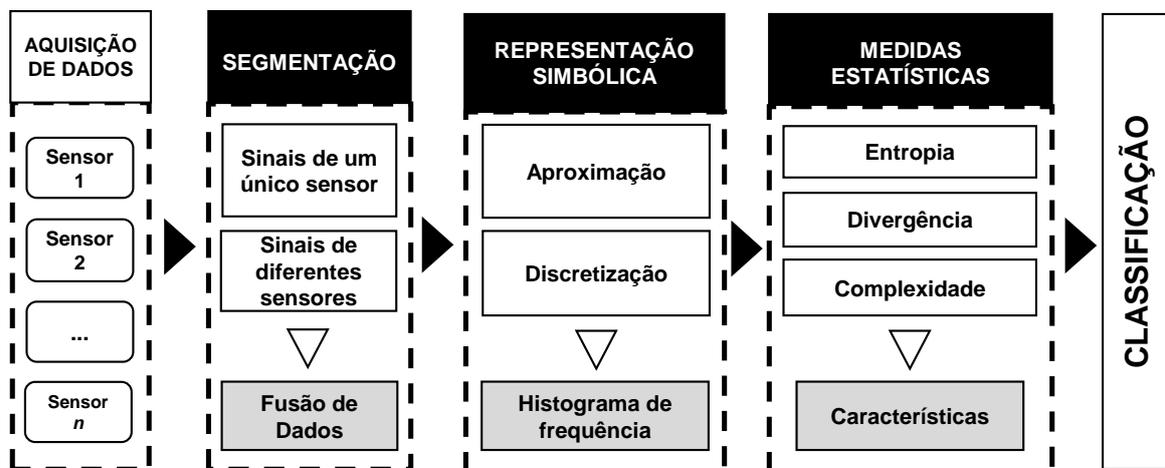


Figura 4-1: Visão geral do método HAR-SR.

O processo de reconhecimento de atividades humanas proposto é composto por cinco etapas principais, conforme mostra a Figura 4-1:

- i. Aquisição de dados e seleção dos sensores;
- ii. A segmentação do sinal, que inclui seleção de sensores e o emprego de métodos de fusão de dados;
- iii. A representação simbólica, que faz a extração automática de padrões da série, além de reduzir a dimensionalidade dos dados e eliminar ruídos e;
- iv. A extração de novas características ainda não exploradas na literatura de RAH.
- v. A última etapa consiste na construção do modelo de classificação de atividades utilizando um método de aprendizagem de máquina.

Devido à diversidade de características que podem ser extraídas dos dados brutos coletados de diferentes sensores, a etapa de pré-processamento e segmentação tem o propósito de formatar e adequar os dados brutos de modo que eles possam ser utilizados pelos algoritmos das próximas etapas da metodologia de RAH. Em geral, essa etapa inclui um processo de pré-processamento e segmentação, que divide o fluxo de dados em blocos menores, conforme demonstrados na Seção 2.2. Além disso, as técnicas utilizadas no pré-processamento devem variar de acordo com o tipo do sensor, nas quais diferentes métricas podem ser utilizadas para redução de ruídos e eliminação de *outliers*, bem como a seleção da frequência de amostragem do sensor.

A fase de segmentação pode dar mais vida útil à bateria e apresentar melhoras quanto ao espaço de armazenamento do smartphone. Ela visa resolver o problema das altas taxas de amostragem que podem ser adotadas pelas aplicações, agravado pelo fato de que os dados gerados pelos sensores são representados no espaço $T = \{t_i\}_{i=1}^N, t_i \in \mathbb{R}^3$, gerando três séries temporais para cada sensor de movimento adotado. Para lidar com esse problema, o método HAR-SR realiza uma fusão de dados ainda no processo de segmentação, cujo objetivo é obter uma série temporal $\{t_i\}_{i=1}^N, t_i \in \mathbb{R}^1$ unidimensional que contém informações discriminantes dos sinais de entrada.

Para extrair padrões da série resultante e reduzir a dimensionalidade dos dados, o HAR-SR transforma cada segmento da série temporal em um conjunto de símbolos através de um processo de aproximação e discretização. Dessa forma, a etapa de representação simbólica consiste desde o processo de discretização da série temporal, até a geração de um histograma de frequências de símbolos.

As medidas estatísticas são utilizadas para extrair características do histograma. Essas medidas são definidas em valores obtidos pelo cálculo de entropia de Shannon, divergência de Jensen-Shannon e medida de complexidade estatística. O resultado desse

processo fornece um conjunto de características que representam as particularidades de cada atividade, só que em um espaço significativamente menor se comparado com outros métodos de extração de características, como os que utilizam apenas a frequências de termos ou a frequência de termos ponderadas (tf – idf).

Por fim, a última etapa consiste no processo de classificação de atividades comum em trabalhos no domínio de aprendizagem supervisionada. Dentre os algoritmos de classificação mais comuns da literatura que utilizam representação simbólica está o KNN (Senin & Malinchik, 2013) (Schäfer, 2015). Entretanto, outros algoritmos de aprendizagem de máquina podem ser utilizados, como árvores de decisão, redes bayesianas e redes neurais.

4.2. Aquisição, segmentação e fusão dos dados

O processo de segmentação utilizado neste trabalho segue as recomendações encontradas em (Banos, et al., 2014) para problemas que envolvem o reconhecimento de atividades humanas. Portanto, no processo de segmentação, a janela de tempo utilizada por padrão será de 2.5 segundos (50 Hz) sem sobreposição. A utilização de janelas temporais maiores não necessariamente melhora o desempenho do classificador e pode acrescentar mais demora na etapa de classificação. A segmentação das séries multidimensionais deverá permitir que os métodos de fusão de dados e os algoritmos de representação simbólica sejam aplicados.

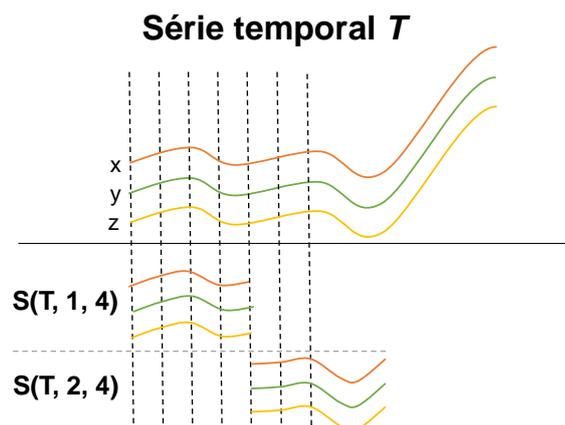


Figura 4-2: Processo de segmentação aplicada aos sinais multidimensionais dos sensores de movimento.

Uma função importante dessa fase é a preparação dos dados para os métodos de discretização. Conforme apresentado na Seção 3.3, os métodos de discretização (e.g. SAX) não aceitam séries multidimensionais e por isso não podem receber diretamente os dados dos sensores de movimento. Assim, a fase de fusão de dados é empregada para gerar um único sinal, a partir dos segmentos, que servirá como entrada para o algoritmo

de representação simbólica. O resultado do processo da fusão de dados é uma série unidimensional, conforme mostra a Figura 4-3.

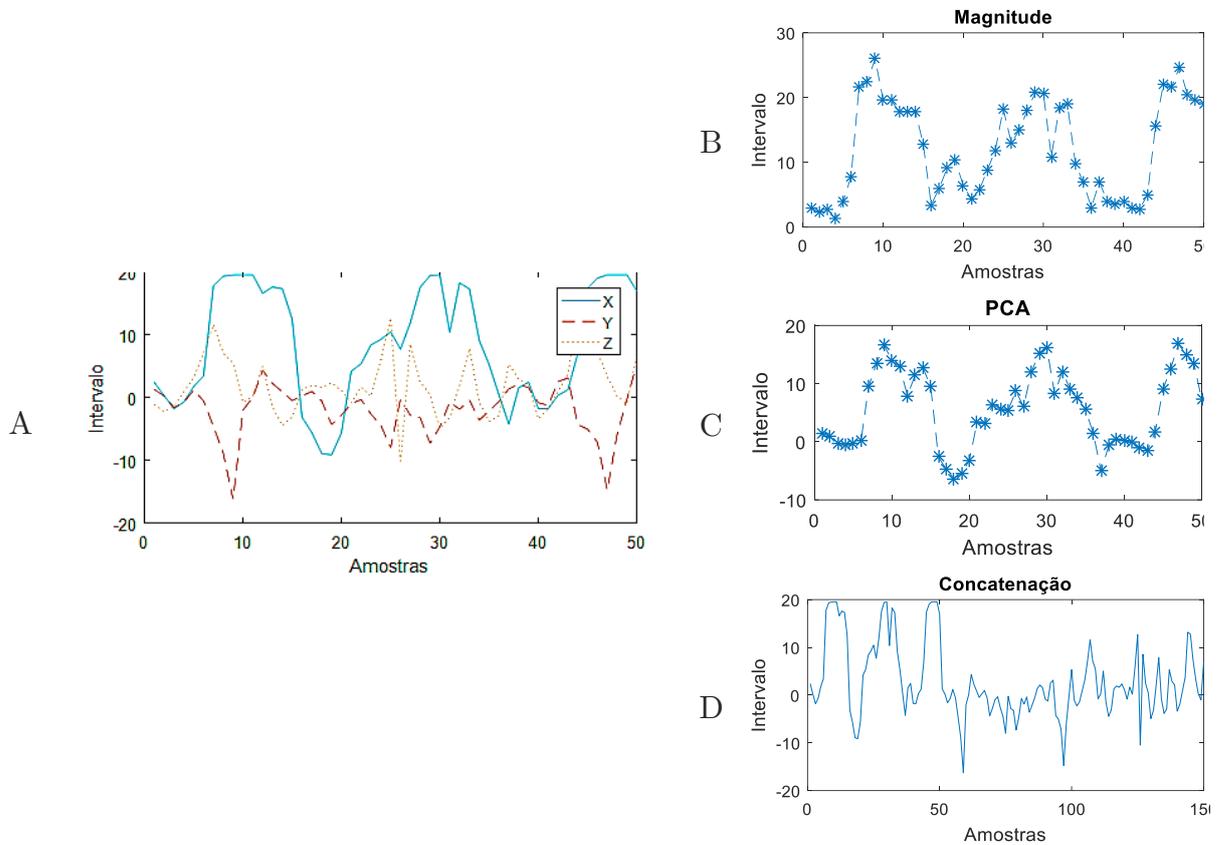


Figura 4-3: A: série temporal multidimensional. B: método da magnitude aplicado à série temporal da Figura A. C: método do PCA aplicado à série temporal da Figura A. D: método da concatenação aplicado à série temporal da Figura A.

No processo de fusão de dados, diferentes métodos podem ser empregados como a Magnitude (método de agregação) e o PCA (método de agregação). Além destes métodos, este trabalho propõe um método de concatenação de sinais para se obter um único sinal $T = \{t_i\}_{i=1}^J$, $t_i \in \mathbb{R}$, porém com a mesma dimensão dos sinais originais. O valor J é obtido em função da quantidade de dimensões da série e também do número de sensores utilizados, conforme mostra a Equação 4-1.

$$J = SG * V * N \quad \text{Equação 4-1}$$

onde SG é o número de sensores utilizados, V é a dimensão do espaço da séries T , como por exemplo, para a série $T = \{t_i\}_{i=1}^N$, $t_i \in \mathbb{R}^3$, $V = \mathbb{R}^3 = 3$, e N é o tamanho do segmento (ou janela de tempo) utilizada em T .

Por exemplo, considere os três sensores de movimento do smartphone, no qual o sinal do sensor de acelerômetro é representado por $T_1 = \{t_i\}_{i=1}^N$, $t_i \in \mathbb{R}^3$, o giroscópio por $T_2 = \{t_i\}_{i=1}^N$, $t_i \in \mathbb{R}^3$ e o magnetômetro por $T_3 = \{t_i\}_{i=1}^N$, $t_i \in \mathbb{R}^3$, onde N representa o

tamanho do segmento e $V = 3$. Seja $SG = (T_1 + T_2 + T_3) = 3$ e seja $N = 128$, o valor de J pode ser obtido de acordo com:

$$J = 3 * 3 * 128 = 1152 \quad \text{Equação 4-2}$$

Assim, a série resultante no processo de concatenação simples para o exemplo é $T = \{t_i\}_{i=1}^J$, $t_i \in \mathbb{R}$, com $J = 1152$.

Essa função é importante na fase de fusão de dados porque ela prepara os sinais para os métodos de discretização. Entretanto, uma desvantagem é que nenhum processo de redução de dimensionalidade é aplicado no método de concatenação simples.

4.3. Representação simbólica dos sinais dos sensores

Após a realização da segmentação dos sinais dos sensores e da fusão de dados, um novo sinal unidimensional $T = \{t_i\}_{i=1}^J$, $t_i \in \mathbb{R}$ deverá ser a entrada dos métodos de discretização. Em resumo, essa etapa tem os seguintes passos:

- i. Utilização do método de janela deslizante para extração de subestruturas de uma série T .
- ii. Cada segmento $S = \{s_i, \dots, s_{s1}\} \in T$ passa por um processo de aproximação simbólica, cujo objetivo é reduzir a dimensionalidade do segmento e remover ruídos, resultando no conjunto $\{\bar{s}_1, \dots, \bar{s}_{s1}\}$.
- iii. Em seguida, uma tabela de pesquisa é utilizada no processo de discretização para se obter a representação simbólica do segmento.
- iv. Por fim, cada segmento discretizado é utilizado para se obter um histograma de frequências dos símbolos.

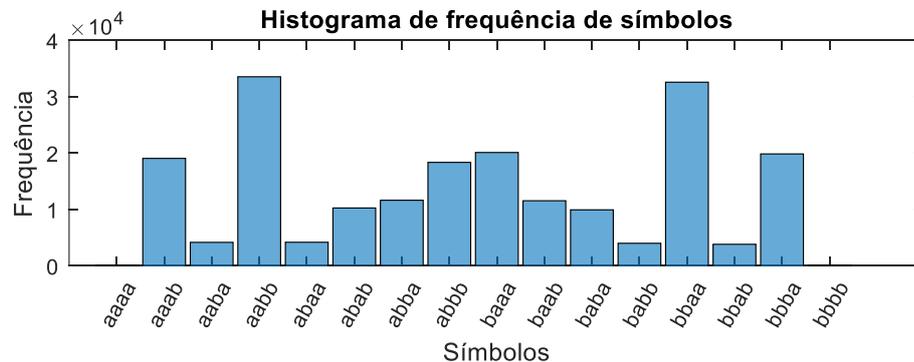


Figura 4-4: Histograma de frequência de símbolos obtido através do método BOSS para uma série temporal.

Os algoritmos utilizados para a discretização nesta pesquisa são o SAX e o SFA, conforme mostrado na Seção 3.3. Para extração de subestruturas e geração do

histograma, o BOP-SAX e o BOSS são utilizados. O resultado do processo de discretização é um histograma de frequências para cada série temporal do conjunto D , conforme mostra a Figura 4-4

A partir do histograma é possível obter uma matriz de frequência dos símbolos (ou palavras). O espaço de característica do histograma é formado por todas as combinações de letras que geram uma palavra, conforme mostra a Figura 4-5. Nesse caso, o tamanho da palavra é 2 e o tamanho do alfabeto é 3, ou seja, o espaço de palavras do histograma é $3^2 = 9$. As linhas correspondem às palavras e as colunas correspondem à quantidade de histogramas (ou segmentos). Se uma palavra está contida em um histograma, então o valor da frequência é diferente de zero, caso contrário, o valor é igual a 0. Cada histograma possui um rótulo da atividade ao qual pertence, que será útil na tarefa de classificação de atividades.

		Série Temporal			
		1	2	...	n
Dicionário de símbolos	aa	0	0	0	0
	ab	10	15	2	20
	ac	5	20	10	5
	ba	10	10	15	10
	bb	5	5	10	5
	bc	0	0	0	0
	ca	5	2	1	8
	cb	2	1	2	1
	cc	0	0	0	0

Figura 4-5: Exemplo do *Bag-Of-Patterns* (BOP) para uma série temporal. Cada linha representa uma palavra do método de representação simbólica e cada coluna representa um segmento da série (ou o histograma gerado a partir desse segmento).

O conjunto de todos os histogramas pode ser utilizado como característica por um método de classificação. Uma abordagem utilizada na literatura para gerar um vetor de características mais relevantes é através da utilização do modelo $td - idf$ aplicado em cada elemento da matriz para transformar os valores das frequências em coeficientes ponderados. Depois de computar todos os valores, a matriz de frequência dos termos se torna a matriz de pesos, cujas colunas são usadas como vetores de peso para cada classe.

No entanto, esta pesquisa utiliza outra metodologia no processo de extração de características. A partir do histograma, são obtidos quantificadores estatísticos como a entropia, a divergência e medida de complexidade. Essa nova metodologia será apresentada na próxima seção.

4.4. Extração de características utilizando medidas estatísticas

O principal objetivo da maioria dos trabalhos de RAH tem sido extrair características que consigam representar bem as atividades humanas. O modelo HAR-SR apresenta uma nova forma de extrair características utilizando os conceitos da teoria da informação aplicados no contexto RAH. As novas características são representadas por um conjunto de valores obtidos através da Medida de Complexidade Estatística (do inglês, *Statistical Complexity Measure* - SCM). A utilização das medidas estatísticas ainda não foi adotada nos sistemas RAH existentes.

A complexidade de uma série temporal é expressa em termos da quantidade de informações que essa série possui. A informação é utilizada para quantificar o grau de ordem e desordem dessa série e pode revelar informações que até então eram desconhecidas. Na literatura diferentes medidas de complexidade foram desenvolvidas para comparar e distinguir comportamentos em séries temporais, como comportamentos periódicos, caóticos e aleatórios conforme mostram (Bandt & Pompe, 2002) (Riedl & Wessel, 2013).

A escolha das medidas corretas a serem utilizadas na SCM depende das características particulares dos dados, como estacionariedade, tamanho da série temporal, variação de parâmetros, nível de ruídos, entre outros. Em todos os casos os aspectos globais da série temporal são capturados de alguma forma, mas os diferentes métodos não são equivalentes em discernir todos os detalhes físicos relevantes (Rosso, et al., 2010). Sendo assim, a medida de complexidade estatística adotada pelo HAR-SR utiliza as medidas sugeridas em (Rosso, et al., 2010), que combina o conceito de desordem, obtido através da entropia de Shannon H_S , como o conceito de desequilíbrio Q_j , obtido pela divergência de Jensen-Shannon (Rosso, et al., 2007).

No contexto RAH, a medida de complexidade é expressa na quantidade de informação guardada pelas séries temporais obtidas por meio dos sensores de movimento de um smartphone. Através dessa medida é possível definir quantificadores (ou valores) associados às atividades executadas. Se as séries temporais forem transformadas em um histograma de símbolos, então a complexidade do sistema pode ser acessada pela informação contida em qualquer símbolo.

Um dos desafios para a obtenção da medida de complexidade é a escolha correta do método para a obtenção da Função de Densidade de Probabilidade (do inglês, *Probability Density Function* - PDF) (Rosso, et al., 2010). Diversos algoritmos podem ser utilizados para se obter uma distribuição de frequência P utilizada para se obter o

PDF. Dentre os algoritmos estão o BOP-SAX e o modelo BOSS, apresentados na Seção 3.3.3. No entanto, cada método utiliza diferentes técnicas que resultam em diferentes representações que devem ser analisadas e escolhidas conforme a natureza do problema.

Uma Função de Distribuição de Probabilidade (PDF) é comumente utilizada para quantificar uma distribuição dos símbolos que representam a série temporal. Esses quantificadores da teoria da informação podem ser definidos como medidas que caracterizam propriedades de um PDF, permitindo a extração de informações das séries temporais, onde determinismo e estocástico são dois extremos do processo (Antonelli, et al., 2018).

A SCM é uma nova representação do histograma, cujos resultados são valores que estão associados ao comportamento da série temporal a qual o histograma está associado. Esses valores correspondem às novas características que podem ser utilizadas por um algoritmo de aprendizagem de máquina para obter um modelo para classificação de atividades mais robusto. A vantagem dessa abordagem é que total de características é significativamente menor que às geradas pelo *tf* ou *tf – idf*.

A entropia de Shannon é a primeira medida obtida no cálculo do SCM. Ela mede a informação contida em uma distribuição de probabilidades relacionada a uma série temporal. A entropia de Shannon (Shannon & Weaver, 1963) é a mais adotada ao lidar com conteúdo da informação.

Para calcular a entropia, é necessário que o histograma *P* seja normalizado, conforme mostra a Equação 4-3, onde p_i é o valor de frequência do *i*-ésimo símbolo do espaço de amostras de tamanho η . Essa normalização é a representação do PDF necessária para o cálculo da medida de entropia.

$$P = \frac{p_i}{\sum_{i=1}^{\eta} p_i} \quad \text{Equação 4-3}$$

Definindo-se o PDF, cujo $P = \{p_i; i = 1, \dots, \eta\}$ com $\sum_{i=1}^{\eta} p_i = 1$, onde Ω é o espaço de amostras de tamanho η . Então, a Entropia de Shannon é definida conforme mostra a Equação 4-4:

$$H[P] = - \sum_{i=1}^{\eta} p_i \ln[p_i] \quad \text{Equação 4-4}$$

O p_i representa as frequências relativas de todos os possíveis símbolos do espaço Ω . Quando $H[P] = H_{\min} = 0$, a previsão de que o resultado *i* irá ocorrer é de completa certeza. A Figura 4-6A mostra um exemplo onde o valor $H[P] = 0$, pois a ocorrência do símbolo 'ab' possui valor de probabilidade máximo. No contrário, temos o mínimo

de conhecimento no caso de uma distribuição uniforme $P_e = \{p_i = \frac{1}{\eta}; i = 1, \dots, \eta\}$, desde que cada resultado tenha a mesma possibilidade de ocorrer, e a incerteza é máxima, $H[P_e] = H_{\max} = \ln \eta$ (Figura 4-6B).

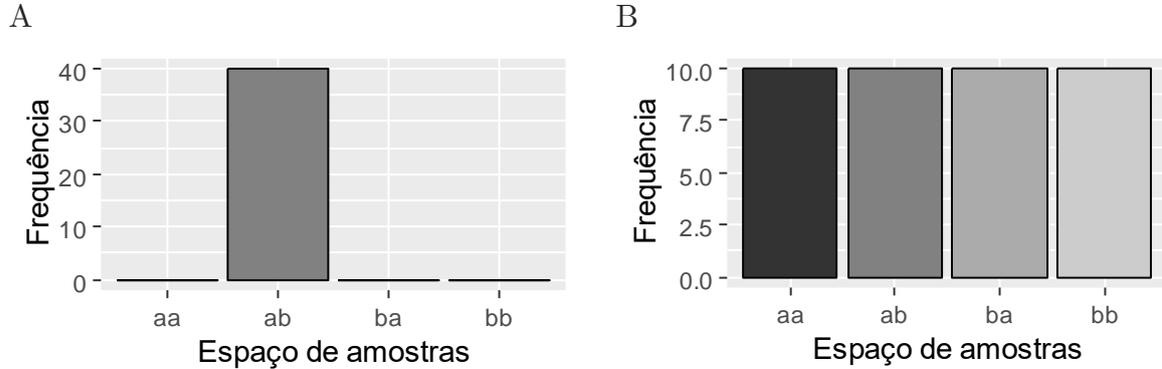


Figura 4-6: A entropia mínima pode ser representado pela distribuição à esquerda, onde $H[P] = H_{\min} = 0$. A distribuição uniforme P_e , à direita na figura, chamada também de histograma de uniforme (H_{REF}), representa uma série aleatória na qual todas as ocorrências possuem o mesmo valor de probabilidades (e.g. um ruído aleatório). No exemplo existem quatro possibilidades de símbolos, cada um com probabilidade $p = \frac{1}{4}$.

Dadas as duas situações extremas, é interessante utilizar a entropia normalizada de Shannon, $0 \leq H \leq 1$, definida na Equação 4-5, onde S_{\max} é o espaço de símbolos do histograma P.

$$H[P] = H[P]/S_{\max} \quad \text{Equação 4-5}$$

Em alguns casos, duas distribuições diferentes P e Q podem apresentar o mesmo valor de entropia, como mostra a Figura 4-7. Para resolver esse problema, utiliza-se o cálculo de divergência, que será a segunda medida estatística adotada nesta pesquisa para calcular a SCM.

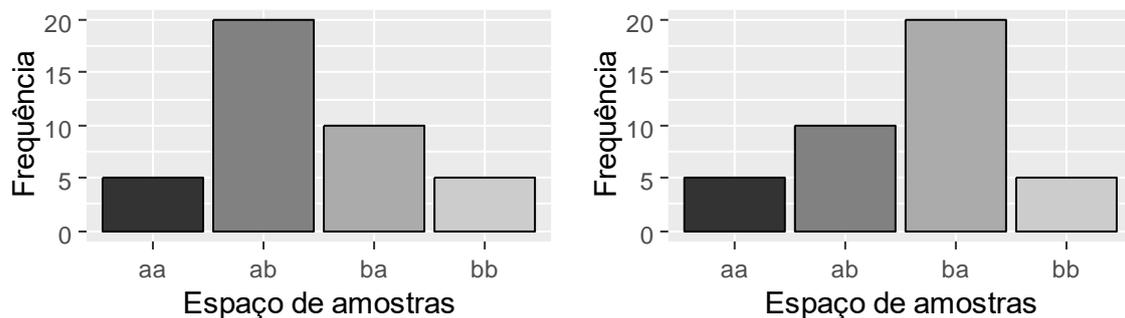


Figura 4-7: As distribuições P e Q são diferentes, mas apresentam a mesma entropia $H = 0.875$ (valor da entropia normalizada)

Na teoria de probabilidade e estatística, a divergência de Jensen-Shannon é um método de medir a similaridade entre duas distribuições de probabilidade P e Q. Essa será a

medida de divergência adotada neste trabalho conforme adotada em (Rosso, et al., 2010).

Adota-se o termo *desequilíbrio* para se referir a divergência entre dois histogramas que, nesse caso, é entre o histograma P da série temporal e o histograma de referência $P_e = \{p_i = \frac{1}{\eta}; i = 1, \dots, \eta\}$ que representa o estado estacionário do sistema. Após esta definição, a divergência é definida conforme a Equação 4-6:

$$Q_j[P, P_e] = Q_0 H[P, P_e] = Q_0 H\left\{\left[\frac{P + P_e}{2}\right] - \frac{H[P]}{2} - \frac{H[P_e]}{2}\right\} \quad \text{Equação 4-6}$$

Q_0 é uma constante de normalização tal que $0 \leq Q_j \leq 1$, que pode ser obtida através da Equação 4-7. Esse valor é obtido em uma situação totalmente determinista, onde apenas um componente de P possui um valor nulo igual a um.

$$Q_0 = -2 \left\{ \frac{\eta + 1}{\eta} \ln(\eta + 1) - \ln(2\eta) + \ln \eta \right\}^{-1} \quad \text{Equação 4-7}$$

Por fim, a última medida estatística a ser utilizada é a medida de complexidade SCM . A medida de complexidade estatística utilizada nesta pesquisa combina o conceito de desordem, obtido através da entropia de Shannon H , com o conceito de *desequilíbrio* Q_j , obtido pela divergência de Jensen-Shannon (Rosso, et al., 2007).

A complexidade C atribui valores baixos tanto a dados aleatórios não-correlacionados (entropia Shannon máxima) quanto a dados perfeitamente ordenados (entropia nula de Shannon). Então, se tivermos uma sequência ordenada, como uma simples oscilação ou tendência, a complexidade estatística seria baixa, e o mesmo aconteceria com uma sequência não-ordenada, como o ruído branco (sinal aleatório com igual intensidade em diferentes frequências) não-correlacionado. A caracterização dos dados é mais difícil para séries situadas entre os dois extremos apresentados e, portanto, a complexidade seria maior (Antonelli, et al., 2018). Dessa forma, a complexidade estatística é definida conforme a Equação 4-8:

$$C[P] = Q_j[P, P_e]. H[P] \quad \text{Equação 4-8}$$

Essas três medidas podem ser utilizadas em um plano Entropia-Complexidade, onde cada segmento de atividade corresponde a um par ordenado (H, C) . Nesse plano, é possível visualizar a estrutura física das séries que representam cada classe. Um exemplo do plano Entropia-Complexidade para duas classes pode ser visto na Figura 4-8.

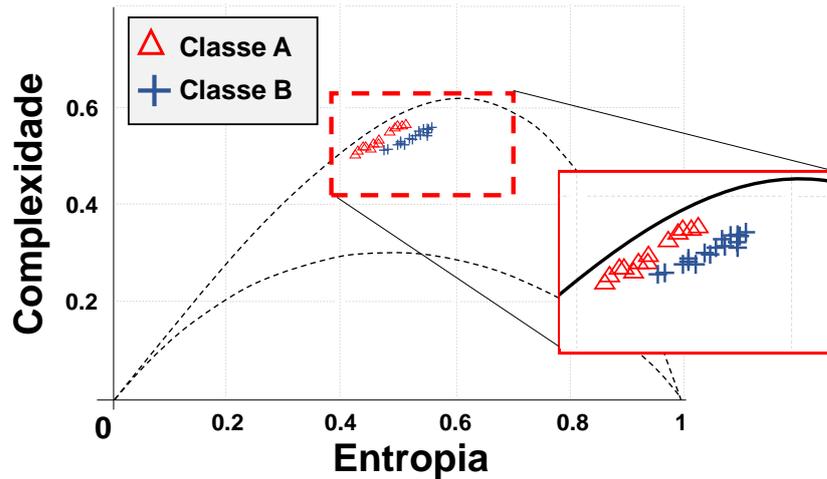


Figura 4-8: Exemplo do plano Entropia-Complexidade para duas classes A e B.

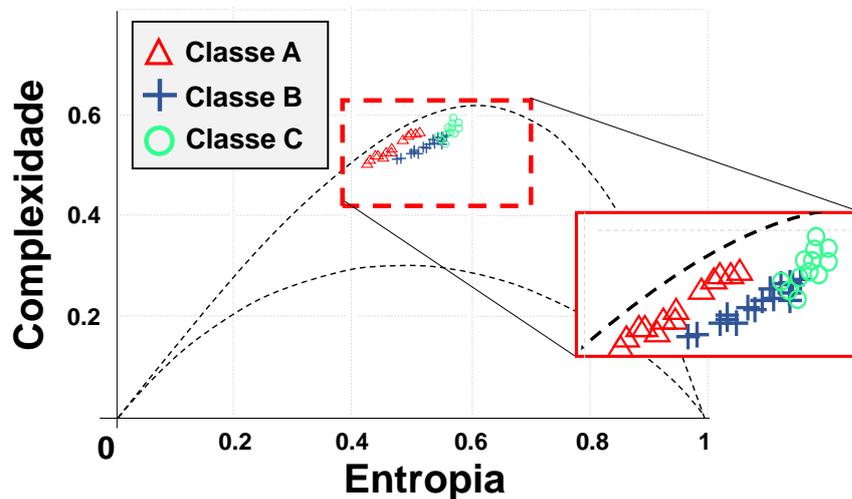


Figura 4-9: Exemplo do plano Entropia-Complexidade para três classes. Note que as classes B e C se sobrepõem em alguns pontos.

Para problemas de classificação multiclass, ou seja, com um número superior a duas classes, utilizar somente as três características estatísticas pode não ser uma boa solução para se obter um bom desempenho de classificação. Um dos motivos pode ser visualizado na Figura 4-9, no qual é possível perceber que existem classes que estão sobrepostas. O algoritmo de aprendizagem também enfrentará dificuldades em classificar essas classes sobrepostas, e por esse motivo seu desempenho de classificação é afetado.

Para contornar este problema, esta pesquisa cria uma derivação das medidas estatísticas de divergência Q_j e complexidade C . A medida de entropia permanece inalterada, pois depende apenas da distribuição de frequências de uma dada série. No entanto, o cálculo de divergência passa por algumas alterações e, conseqüentemente, o

cálculo da medida de complexidade. Isso acontece porque são introduzidos novos tipos de histogramas de referência no cálculo dessas medidas.

A primeira mudança é influenciada pelo número de classes do problema. Além do histograma uniforme P_e utilizado no cálculo da divergência, são utilizados também os histogramas de referências para cada uma das classes existentes, conforme mostra a Figura 4-10. O processo de geração dos histogramas de referência para cada classe segue as seguintes etapas:

- i. Todas as séries temporais pertencentes a uma classe são agrupadas em um conjunto. Por exemplo, as séries temporais da *classe A* são agrupadas $R_A = \{T_1^A, T_2^A, T_3^A, \dots, T_n^A\}$;
- ii. Cada série temporal do conjunto R_A passa pela fase de segmentação, fusão de dados e discretização, conforme a Figura 4-10, para se obter o histograma de frequências $P = \{P_1^A, P_2^A, P_3^A, \dots, P_n^A\}$.
- iii. Os valores de frequência de cada símbolo de P são somados, resultando em único histograma: o histograma de referência para a *classe A*. Por exemplo, todos as frequências para o símbolo 'aa' do conjunto P são somadas, depois para o símbolo 'ab' e assim por diante.
- iv. O processo se repete em todas as classes do problema. Por exemplo, um problema com duas classes resulta no conjunto $REF = \{P_e, P_A, P_B\}$.

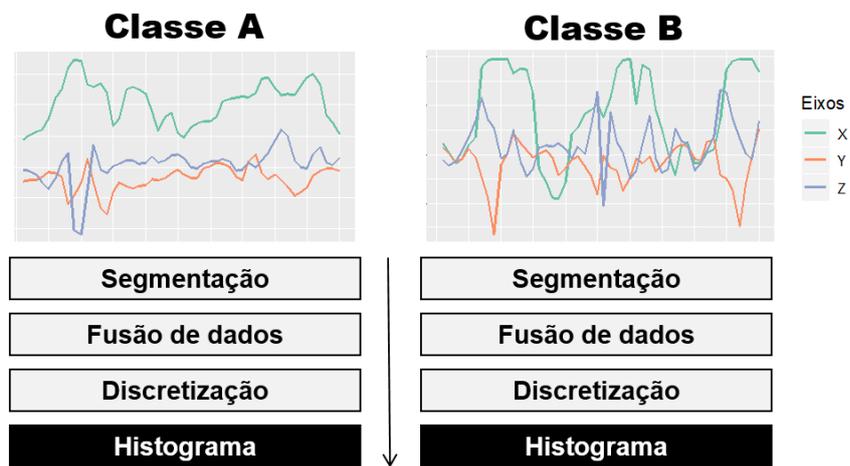


Figura 4-10: Processo de geração dos histogramas de referência P_A e P_B para cada classe A e B.

Assim, além das características de entropia, divergência e complexidade geradas a partir do P_e , para cada nova classe existirá uma característica adicional de divergência e complexidade. A Tabela 4-1 apresenta os histogramas utilizados em um problema de classificação com duas classes.

Símbolo	Classe
P_e	Histograma de referência geral, cujo valor é obtido através de uma distribuição uniforme que representa a entropia de uma série completamente aleatória
P_A	Histograma de referência para a classe A
P_B	Histograma de referência para a classe B

Tabela 4-1: A Tabela mostra a relação dos histogramas de referência utilizados em um problema de classificação com duas classes.

A Figura 4-11 mostra um exemplo do histograma de referência P_A e P_B para a classe Classe A e Classe B respectivamente, em um espaço de 9 símbolos. Note que os histogramas de cada classe podem possuir diferenças que podem ser observadas pela distribuição do histograma. Essa diferença será observada também nos valores obtidos pela medida de divergência e de complexidade.

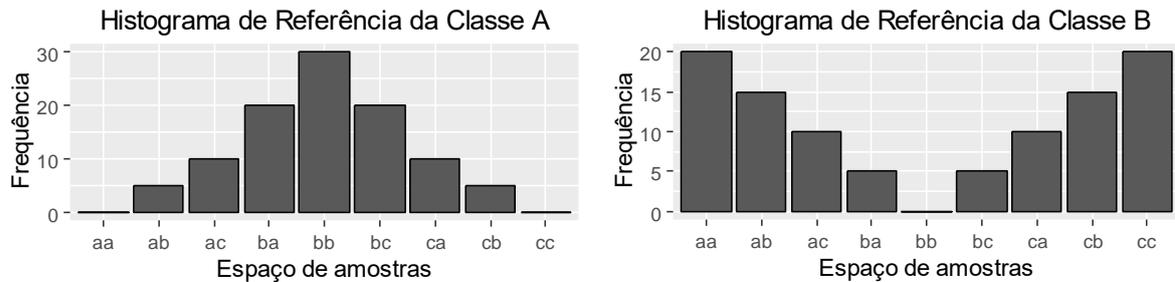


Figura 4-11: histograma de referência P_A e P_B para a classe A e classe B em um espaço de 9 símbolos.

Após a obtenção do histograma de referência P_A e P_B , os valores da entropia, da divergência e da medida de complexidade são calculados. Os primeiros valores são obtidos com base no P_e . Em seguida, a divergência é calculada a partir de cada histograma de referência (P_A , P_B). Por exemplo, no problema de duas classes, classe A e classe B, com seus respectivos histogramas de referências P_A e P_B , os valores da medida de divergência podem ser obtidos através das fórmulas da Tabela 4-2:

Divergência	Fórmula
a	
Classe A	$Q_A[P, P_A] = Q_0 J[P, P_A] = Q_0 S \left\{ \left[\frac{P + P_A}{2} \right] - \frac{H[P]}{2} - \frac{H[P_A]}{2} \right\}$
Classe B	$Q_B[P, P_B] = Q_0 J[P, P_B] = Q_0 S \left\{ \left[\frac{P + P_B}{2} \right] - \frac{H[P]}{2} - \frac{H[P_B]}{2} \right\}$

Tabela 4-2: cálculo da divergência com base no histograma de referência P_A e P_B

Esses novos valores da medida de divergência também resultam em novos valores da medida de complexidade, conforme mostra a Tabela 4-3

Complexidade	Fórmula
Classe A	$C[P] = Q_J[P, P_A].H[P]$
Classe B	$C[P] = Q_J[P, P_B].H[P]$

Tabela 4-3: cálculo da medida de complexidade com base no histograma de referência P_A e P_B

O processo acaba quando todos os histogramas de referências (P_e , P_A , P_B) são utilizados para se obter as medidas de divergência e complexidade. O conjunto de características obtidas para um problema de duas classes é semelhante ao mostrado na Tabela 4-4.

H_s	$Q_{uniforme}$	Q_{classe_A}	Q_{classe_B}	$C_{uniforme}$	C_{classe_A}	C_{classe_B}	Classe
0.543884	0.785297	0.108659	0.066576	0.427110	0.059098	0.036210	Classe A
0.497008	0.805370	0.065534	0.206573	0.400275	0.032571	0.102668	Classe B
...
0.495400	0.840364	0.121913	0.010469	0.416316	0.005186	0.153602	Classe A

Tabela 4-4: tabela de valores obtidos através do cálculo da entropia, divergência e complexidade para um problema com duas classes.

A partir desse novo conjunto de medidas estatísticas, diversos planos Entropia-Complexidade podem ser criados com base nas novas características. Por exemplo, cada medida de complexidade pode ser utilizada em um plano Entropia – Complexidade_i, onde i é a medida de complexidade para um histograma de referência de uma classe específica. Esse exemplo pode ser visto na Figura 4-12, no qual são utilizados os valores de entropia e os valores de complexidade calculados com base no histograma de referência de duas classes.

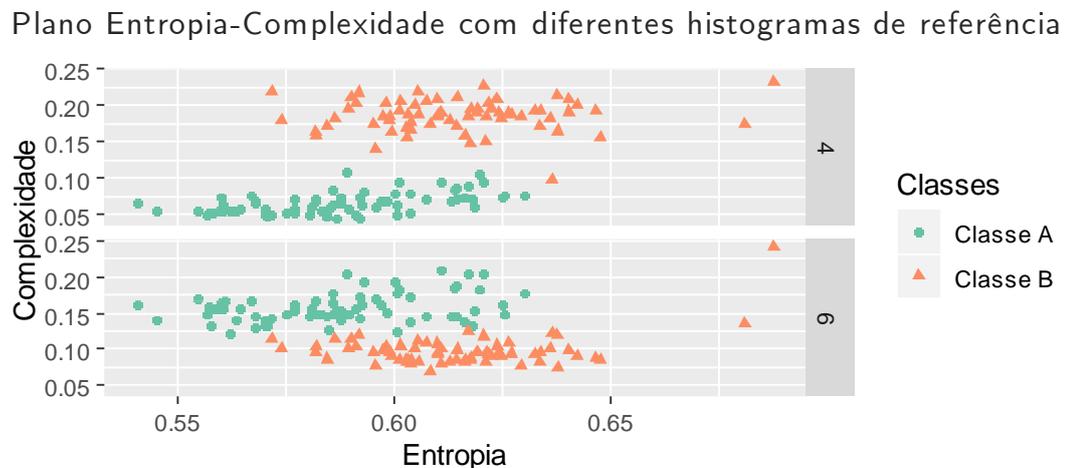


Figura 4-12: Plano $H \times C_i$, onde C_i são as medidas de complexidade que utilizam os histogramas de referência de duas classes. Os valores são $4 = P_A$, $6 = P_B$.

A Figura 4-12 mostra que os valores da medida de complexidade diminuem em relação a uma classe A , considerando que o histograma de referência utilizado é em relação a classe A . Isso acontece porque o cálculo de divergência entre os histogramas que pertencem a uma mesma classe resulta em valores de complexidade menores. Essas diferenças são causadas por cada histograma de referência, ainda que pequenas, são fundamentais para que os métodos de aprendizagem consigam encontrar padrões e gerar modelos de classificação com bom desempenho.

Essa nova metodologia de extração de características faz com que o HAR-SR seja um método que gera poucas características, obtidas em função do número de classes do problema, conforme mostra a Equação 4-9:

$$\text{características} = H + Q_i + C_i \quad \text{Equação 4-9}$$

Onde, i representa o número de classes do problema. Em um cenário onde são avaliadas duas classes, ou seja, $i = 3$, pois são dois histogramas de referência (P_A, P_B), um para cada classe e o histograma uniforme P_e . Isso quer dizer que utilizando os mesmos parâmetros $\omega = 6$ e $\alpha = 6$, o número total de características para este exemplo será igual 7, muito melhor do que as 46.656 características produzidas pela utilização do $tf - idf$, por exemplo, ou as 145 características manuais utilizadas na literatura pelas abordagens HCFM. Isso pode ser considerado uma vantagem, principalmente porque o modelo de classificação construído será mais simples.

A próxima seção descreve a última etapa do processo de classificação de atividades. A utilização de poucas características reduz a carga de processamento dos métodos de aprendizagem em geral, o que é ideal para ser aplicado em dispositivos móveis como smartphones.

4.5. Método de classificação do HAR-SR

O problema abordado nesta pesquisa é um problema de classificação. Diversos métodos de classificação podem ser utilizados para gerar modelos a partir das características geradas pelo HAR-SR, como árvore de decisão, redes neurais, SVM, *naive bayes*, entre outros. No entanto, este trabalho escolheu somente o método KNN para facilitar a comparação e a avaliação do HAR-SR com os demais trabalhos similares na literatura, conforme apresentado no capítulo 5. Como medida de distância do KNN foi escolhida a similaridade dos cossenos comumente utilizada por métodos de classificação baseados em representação simbólica como BOSS-VS.

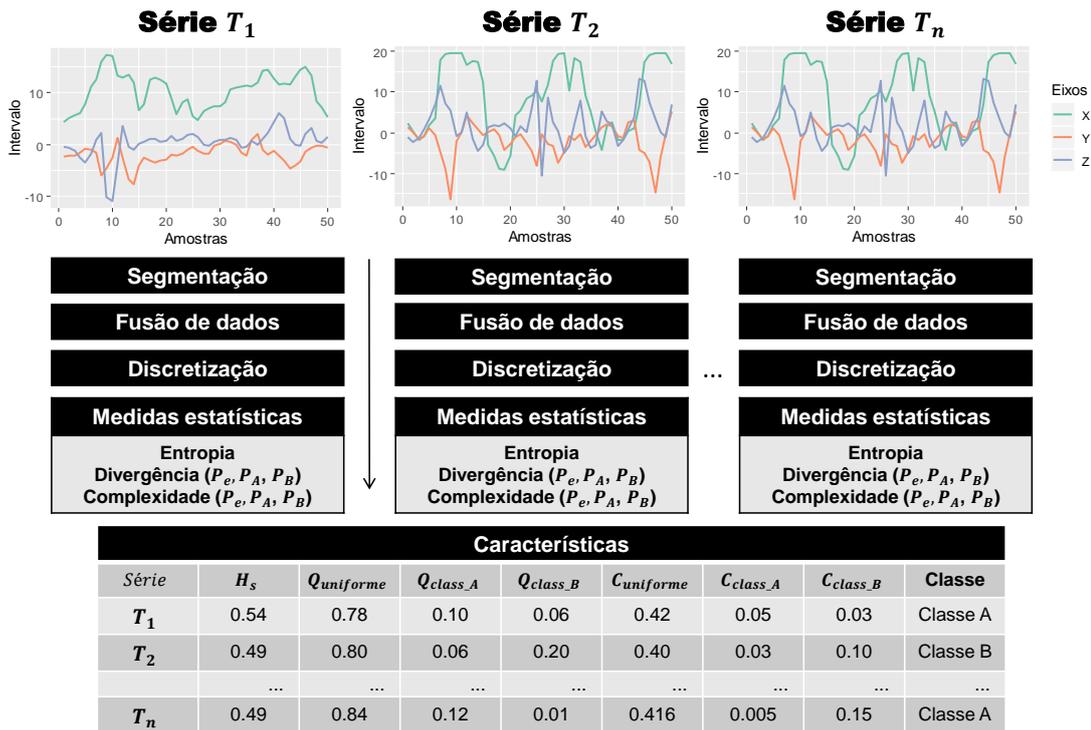


Figura 4-13: O processo de extração de características adotado pelo método HAR-SR.

O processo de classificação consiste em associar um rótulo de uma classe existente a uma nova série temporal T_u ainda não-rotulada. Para esta tarefa, o modelo passa por duas etapas:

- i. O treinamento
 - a. Criação de uma base de dados $D = \{(T_1, y_1), (T_2, y_2), (T_3, y_3), \dots, (T_N, y_N)\}$ com informações de cada uma das classes do problema, onde N é o número de instâncias, T_i é a série temporal, e y_i é o rótulo associado à série T_i .
 - b. Criação de um histograma de referência para cada uma das classes. Por exemplo, $REF = \{P_e, P_A, P_B\}$ para duas classes;
 - c. Para cada série temporal T_i , aplicar a segmentação $S = \{s_i, \dots, s_{s_i}\} \in T$, a fusão de dados, realizar a discretização para obtenção do histograma P_i ;
 - d. Cálculo das medidas estatísticas utilizando cada histograma de referência.
 - e. Geração do conjunto de treinamento DS a partir das medidas estatísticas.
- ii. A classificação.
 - a. Dada uma nova série não rotulada T_u ;
 - b. aplicar a segmentação $S = \{s_i, \dots, s_{s_i}\} \in T_u$, a fusão de dados, realizar a discretização para obtenção do histograma P_u ;
 - c. Calcular as medidas estatísticas que utilizam os histogramas de referência, obtidos na fase de treinamento, resultando no vetor de características da série T_u .

- d. O rótulo associado à série $T_i \in DS$ mais próxima da série T_u será atribuído à série não-rotulada T_u , utilizando a métrica de similaridade entre cossenos.

A Figura 4-13 apresenta um exemplo do processo de extração de características na fase de treinamento para um problema com duas classes.

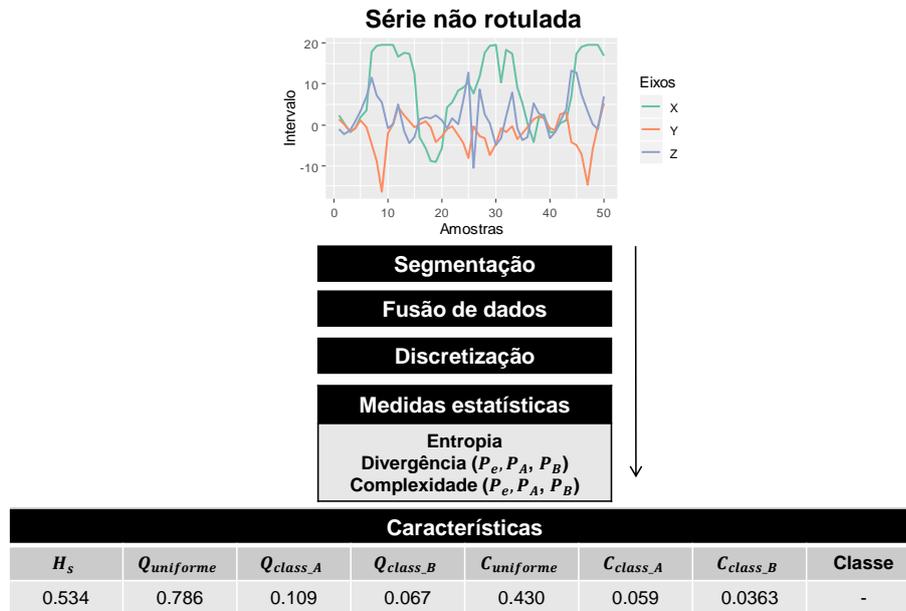


Figura 4-14: Processo de extração de características de um novo sinal de atividade.

Para classificar uma nova série não rotulada T_u , o procedimento é semelhante ao realizado na fase de treinamento. A diferença é que não há necessidade de obtenção de novos histogramas de referência. O processo de extração de características é mostrado na Figura 4-14.

Por fim, o rótulo associado à série mais próxima será atribuído à série não-rotulada utilizando uma métrica de similaridade dos cossenos. Com a nova atividade classificada, o processo de classificação é finalizado.

$$\operatorname{argmax} \left\{ \text{similaridade}(T_i, T_u) = \frac{T_i \cdot T_u}{\|T_i\| \cdot \|T_u\|} \right\} \quad \text{Equação 4-10}$$

onde T_i pertence ao conjunto de treinamento e T_u a série não-rotulada

Como mencionado anteriormente, a escolha do método de classificação adotada no método proposto foi o KNN, mas não há impedimentos de que outros métodos sejam utilizados.

4.6. Considerações Finais do Capítulo

Neste capítulo foi apresentado o HAR-SR, um método de classificação de séries temporais aplicado no contexto de reconhecimento de atividades humanas. O HAR-SR aplica fusão de dados para reduzir a dimensionalidade dos dados multidimensionais gerados pelos sensores de movimento de um smartphone. Além disso, transforma os dados em uma representação discreta através de algoritmos de representação simbólica. Esse procedimento permite que diversos métodos aplicados no contexto de mineração de texto sejam aplicados às séries temporais.

Todo o processo de extração de padrões é automático, cuja qualidade das características extraídas dos sinais depende dos parâmetros dos métodos de discretização, como tamanho da janela, palavra e alfabeto. A partir dos histogramas de frequências obtidos pelos métodos de discretização, novas características podem ser obtidas através das medidas estatísticas de entropia, divergência e medida de complexidade.

Essas medidas são provenientes da teoria da informação e são muito úteis para quantificar o grau de ordem e desordem de uma série temporal e podem revelar informações que até então eram desconhecidas. Isso significa que é possível descobrir diferenças entre os padrões de atividades humanas, por exemplo, através das diferenças encontradas entre as medidas.

O processo de classificação de uma atividade segue o mesmo modelo visto na aprendizagem supervisionada, sendo necessária uma fase de treinamento e outra fase de teste. As novas características são utilizadas por métodos de classificação para reconhecer atividades. A utilização dessas medidas no processo de classificação é interessante porque o modelo gerado é mais simples se comparado com o modelo de extração manual de características, já que poucas características são utilizadas e ainda assim, pode apresentar um bom desempenho no domínio de utilização.

Experimentos e resultados

Este capítulo apresenta os experimentos utilizados para avaliar o método HAR-SR na classificação de atividades humanas. Para alcançar este objetivo, seis cenários serão avaliados e são realizadas comparações com diferentes métodos encontrados na literatura. A métrica de desempenho utilizada é a medida F1 e a acurácia. Uma etapa importante desse processo é a utilização da validação cruzada por indivíduos como métrica de validação aplicada nas bases de dados, de modo que seja reduzido o viés na classificação. Também serão realizados diversos testes para a seleção dos melhores parâmetros para o método HAR-SR. Essa etapa é importante pois influencia no desempenho geral do HAR-SR. Por fim, o método é comparado com vários trabalhos relacionados e são realizadas análises e discussões sobre os resultados.

5.1. Protocolo Experimental

Esta seção mostra o planejamento dos experimentos realizados para reconhecer atividades através dos smartphones.

5.1.1. Conjunto de dados

A primeira base de dados, chamada de SHOAIB AS (SH), é disponibilizada por (Shoaib, et al., 2013). A seleção dessa base de dados foi motivada pelos seguintes aspectos:

- i. A base SHOAIB contém dados coletados em cinco posições diferentes do corpo;

- ii. Contém seis atividades com dados coletados dos sensores de acelerômetro, giroscópio e magnetômetro
- iii. A base de dados está balanceada e deve representar uma avaliação mais justa dos métodos. Isso significa uma redução no viés causado tanto por indivíduos com mais dados de atividades e quanto a rótulos de atividades mais numerosos que outros.

Na base SHOAIB são coletados dados de seis atividades de participantes do sexo masculino, com idades entre 25-30 anos, usando um smartphone Samsung galaxy SII. A quantidade de tempo de cada atividade é de 3-5 minutos por participante. Cada um dos participantes carregava um smartphone em cinco partes do corpo: lado direito e esquerdo da calça, cintura, braço e pulso. Os dados foram coletados em cada uma das posições a uma taxa de 50Hz de forma simultânea.

A segunda base de dados, chamada de WISDM, é disponibilizada por (Kwapisz, et al., 2011). O experimento possui 36 participantes portando o smartphone no bolso. Os dados do acelerômetro foram coletados a 20Hz. Nesta pesquisa, somente os indivíduos que realizavam todas as atividades foram selecionados, totalizando 19 indivíduos cujo identificadores são {3, 5, 6, 7, 8, 12, 13, 18, 19, 20, 21, 24, 27, 29, 31, 32,33, 34, 36}.

A última base utilizada, chamada UCI HAPT, disponibilizada por (Anguita, et al., 2013). Contém dados de 30 voluntários com idades entre 19 e 48 anos. Cada um dos participantes usava um smartphone Samsung Galaxy SII na cintura. Cada participante seguiu o protocolo duas vezes: no primeiro teste o smartphone foi fixado ao lado esquerdo do cinto e no segundo teste o participante poderia escolher um lugar de sua preferência. Foram capturados dados dos sensores de 3D-acelerômetro e 3D-giroscópio, a uma taxa de 50Hz. Os dados selecionados são referentes a seis atividades.

A Tabela 5-1 apresenta um sumário das bases de dados selecionadas para esta pesquisa.

Base de dados	SHOAIB SA	WISDM	UCI HAPT
Ano	2014	2012	2015
Nº de indivíduos	10	19 (36)	30
Idade	25-30	-	19-48
Frequência (Hz)	50	20	50
Atividades	Levantando	Levantando	Levantando
	Sentando	Sentando	Sentando
	Subindo Escadas	Subindo escadas	Subindo escadas
	Descendo Escadas	Descendo escadas	Descendo escadas
	Andando	Andando	Deitando

	correndo	correndo	Sentando
Instâncias	2700000	893702	748406
Sensores	Acelerômetro, giroscópio, magnetômetro	Acelerômetro	Acelerômetro
Localização dos sensores	Cintura Bolso esquerdo Bolso direito Braço Pulso	Cintura	Cintura

Tabela 5-1: Sumário das bases de dados selecionadas para esta pesquisa.

5.1.2. Métricas de validação

Conforme mostram os trabalhos apresentados na seção 3.4, não existe um padrão fixo de comparação de desempenho das soluções de RAH na literatura. Na fase de teste, muitos parâmetros diferem de pesquisa para pesquisa. Primeiro, o conjunto de atividades a serem classificadas pelos algoritmos refletem diretamente na acurácia dos métodos. Algumas atividades podem ser confundidas com outras e, de modo geral, a acurácia pode ser manipulada através da escolha de uma base de dados conveniente (Morales & Akopian, 2017). Muitos trabalhos não dispõem de uma comparação dos diferentes métodos para uma mesma base de dados (Shoaib, et al., 2015). Para minimizar esses problemas, esta pesquisa utiliza três bases de dados públicas e amplamente adotadas em outros trabalhos. O método proposto foi desenvolvido na ferramenta MATLAB, desde a fase de pré-processamento até a criação do modelo de classificação. Os detalhes de complexidade computacional dos métodos utilizados na criação do HAR-SR são apresentados no Apêndice A.

Nos trabalhos apresentados pela literatura foram identificadas as três formas mais comuns de testar a acurácia dos algoritmos de RAH:

- i. A validação personalizada, que utiliza dados de apenas um único indivíduo (Bao & Intille, 2004);
- ii. A validação cruzada (CV) que utiliza dados de mais de um indivíduo (Kwapisz, et al., 2011);
- iii. Validação cruzada por indivíduo (LOSO, do inglês, *leave-one-subject-out cross validation*).

Para avaliar o desempenho dos métodos de classificação, costuma-se dividir a base de dados em dois conjuntos: o conjunto de treinamento e o conjunto de teste. O conjunto de treinamento é utilizado pelo algoritmo de classificação para gerar um modelo. O desempenho desse modelo será avaliado através da utilização do conjunto de testes.

Na validação personalizada, os dados do conjunto de treino e conjunto de teste são do mesmo indivíduo, conforme mostra a Figura 5-1A. O algoritmo é treinado com os dados do mesmo indivíduo no qual é testado. A acurácia nesse tipo validação é tipicamente maior porque em outros tipos de validações os algoritmos RAH sofrem degradação em seu desempenho por causa da introdução de indivíduos desconhecidos na fase de teste.

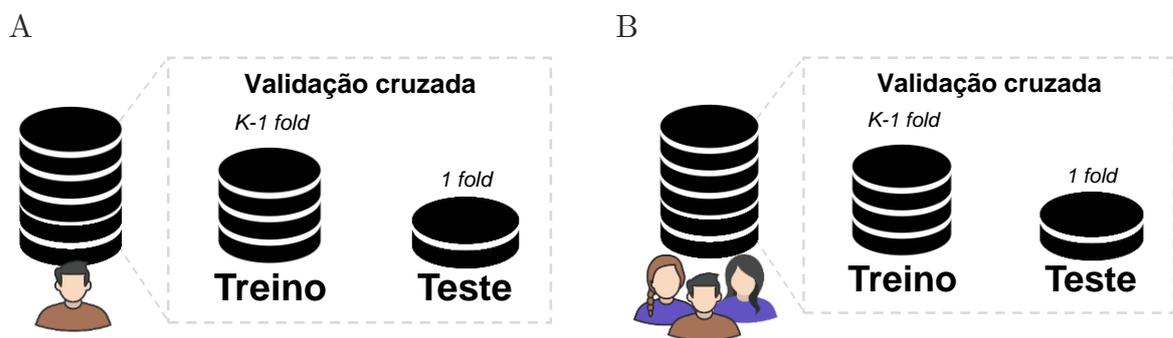


Figura 5-1: A: validação cruzada utilizando dados do mesmo indivíduo. B: Validação cruzada mista com dados de diferentes indivíduos.

A maioria dos estudos em RAH usam a validação cruzada tradicional (ou mista) para avaliar e comparar o desempenho dos algoritmos. Pela definição, a validação cruzada (CV), toma toda a base de dados e define um número fixo de partições, que será o número de partes em que os dados serão divididos igualmente. Tomando N partições, separa-se uma partição de N para o teste e o restante $N - 1$ para treinar o modelo. Feito isso, calcula-se em seguida a taxa de erro utilizando o conjunto que foi separado inicialmente como dados de teste. Esse procedimento irá acontecer N vezes com diferentes conjuntos de treinamento, mas utilizando a mesma base de dados. Ao fim, a média das N taxas de erros é utilizada para representar a taxa de erro geral. Após uma série de testes utilizando os mais diversos cenários, uma boa estimativa de erro é alcançada com a utilização de 10 partições e esse tem sido o método padrão de avaliação de diversos problemas (Witten, et al., 2016).

Para os métodos cujo objetivo é gerar um modelo de classificação personalizado a validação cruzada é uma ótima forma de validação. No entanto, pode não ser uma boa métrica de validação para modelos generalistas. Isso acontece porque ao omitir a informação dos indivíduos, fato comum quando se utiliza a CV, alguns segmentos pertencentes ao mesmo indivíduo do conjunto de teste são introduzidos no treinamento.

Essa mistura entre os dois conjuntos resulta em um modelo de classificação que já conhece parte do seu conjunto de teste. No contexto de RAH, isso quer dizer que o modelo treinado utilizando o CV poderá conhecer os padrões de atividade de um indivíduo específico que deveria estar presente somente no conjunto de teste. O resultado desse processo é a criação de modelos com acurácias maiores na classificação, no entanto, não são um reflexo da realidade.

Para encontrar um modelo generalista, que apresente um bom desempenho em todo o domínio, é preciso primeiramente possuir uma base de dados que represente bem o domínio em estudo, o que no contexto RAH ainda é um desafio a ser superado. Em seguida, é necessário separar com cuidado o conjunto de treinamento do conjunto de teste, para que não haja nenhuma contaminação.

O segundo problema pode ser resolvido com a validação cruzada por indivíduo (do inglês, *Leave-one-subject-out cross validation* - LOSO). A LOSO considera a informação do indivíduo na separação do conjunto, impedindo que o conjunto de treinamento e de testes sejam contaminados por padrões de um mesmo indivíduo. Dessa forma, o número N de partições do CV corresponde ao número de indivíduos no conjunto (Figura 5-2). Esse método de validação é importante porque é indicador do potencial de implantação para massas, sem a necessidade de obtenção de dados do indivíduo antes da utilização do sistema (Morales & Akopian, 2017). Dessa forma, essa pesquisa apresenta seus resultados baseados no LOSO.

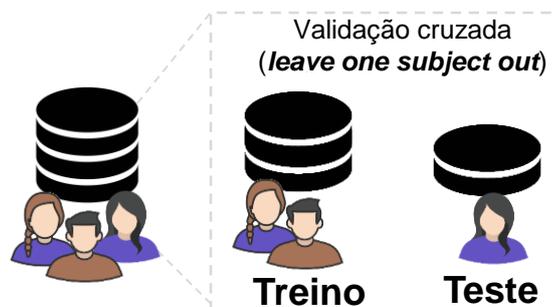


Figura 5-2: Validação cruzada *leave-one—subject-out* (LOSO).

Essa metodologia de validação também é utilizada em outros processos do HAR-SR. Para gerar o conjunto de histogramas de referências, também são necessários alguns cuidados para que nenhum deles contenham informações acerca do conjunto de testes. Por exemplo, para uma instância que pertence ao *usuário 1*, é necessário os histogramas utilizados não contenham nenhum tipo de informação acerca do *usuário 1*, e assim por diante, conforme mostra a Figura 5-3. Esse cuidado evita que informações dos indivíduos sejam associadas ao histograma de referência, que devem resultar em um conjunto de características sem nenhuma mistura.

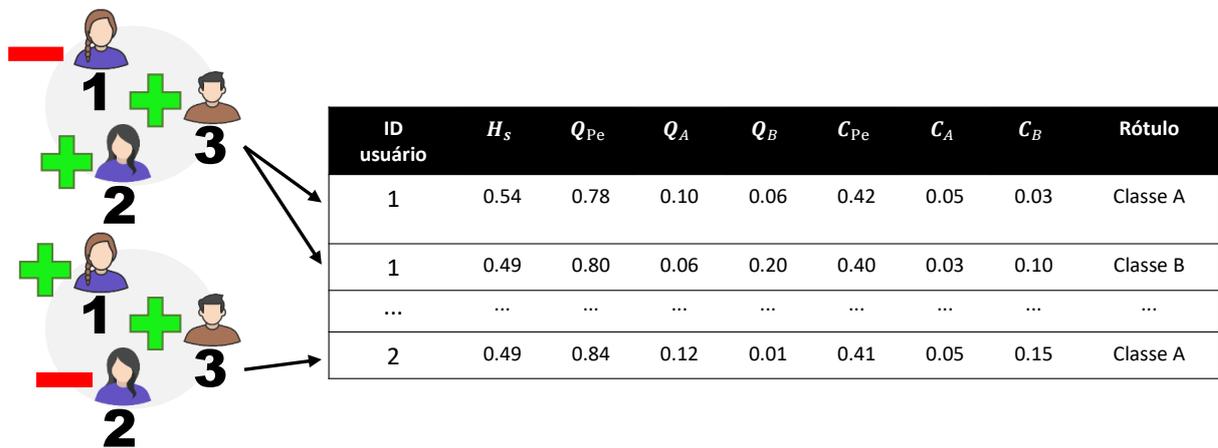


Figura 5-3: A métrica de validação utilizada na geração dos histogramas de referência das atividades é o *leave one out*.

Em alguns casos isso é permitido, como por exemplo, no modelo personalizado, o histograma de referência pode ser formado com dados do mesmo indivíduo (Figura 5-4) ou com dados de indivíduos diferentes (Figura 5-3). O histograma que contém somente os dados do próprio indivíduo tende a beneficiar o algoritmo de classificação, que em geral, apresentará um maior desempenho na tarefa de reconhecer padrões.

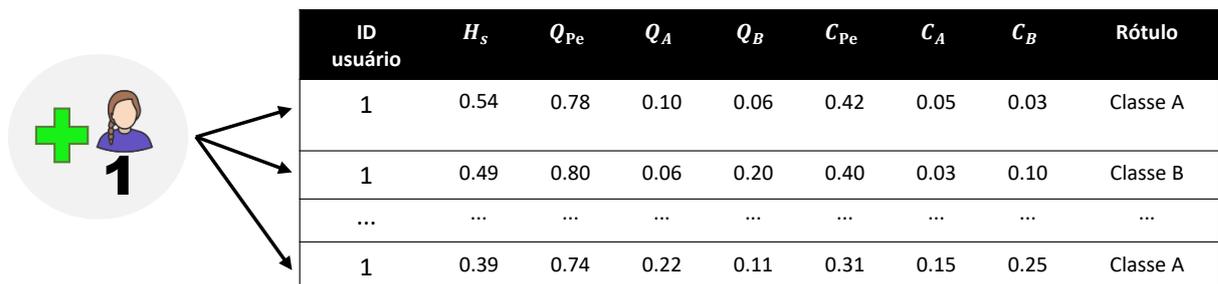


Figura 5-4. Em modelos personalizados, é utilizada a validação cruzada tradicional ou mista para geração dos histogramas de referência das atividades.

5.1.3. Métricas de Avaliação

Os resultados obtidos a partir do LOSO podem ser analisados através da matriz de confusão para identificar as atividades que mais afetam o desempenho do método proposto. A matriz de confusão $M_{n \times n}$ pode ser utilizada em problemas envolvendo n classes (*multiclass*) onde é possível mostrar o resultado gerando uma matriz bidimensional. Essa matriz é formada por uma linha e coluna para cada classe existente.

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}$$

Figura 5-5. A matriz mostra a diagonal principal representada pelos valores $\{d_1, d_2, \dots, d_n\}$.

O processo de avaliação pode ser realizado através da observação da diagonal principal $\{d_1, d_2, \dots, d_n\}$ da matriz, vista na Figura 5-5. As linhas representam as atividades atuais e as colunas as atividades que devem ser previstas.

Tabela 5-2: Terminologia utilizada na matriz de confusão

Verdadeiro Positivo (<i>TP</i>)	Amostra atual da classe <i>A</i> que é classificada corretamente como classe <i>A</i> .
Verdadeiro Negativo (<i>TN</i>)	Amostra atual da classe <i>B</i> que é classificada corretamente como classe <i>B</i> .
Falso Positivo (<i>FP</i>)	Amostra atual da classe <i>B</i> que é classificada incorretamente como classe <i>A</i> .
Falso Negativo (<i>FN</i>)	Amostra atual da classe <i>A</i> que é classificada incorretamente como classe <i>B</i> .

Um bom resultado agrupa todas as atividades na diagonal principal, enquanto o restante dos elementos não pertencentes à diagonal deve possuir o valor zero (Witten, et al., 2016). A matriz de confusão permite uma análise mais detalhada se comparada apenas com a acurácia. A Tabela 5-2 mostra uma matriz de confusão simplificada de duas classes *A* e *B*.

Tabela 5-3. Os quatro números fundamentais para estimar as medidas estatísticas de performance de um classificador.

Classe atual	Classe alvo	
	A	B
A	Verdadeiro Positivos	Falso Negativos
B	Falso Positivos	Verdadeiro Negativos

Tabela 5-4: Matriz de confusão para duas classes.

Utilizar somente a acurácia pode não ser a forma mais confiável para se medir o real desempenho de um método de classificação. Seus resultados podem ser influenciados pelo desbalanceamento da base de dados, por exemplo. Esta pesquisa utiliza a medida *F* como medida de avaliação principal, além da acurácia. A medida *F* considera outras

medidas em seu resultado, como a precisão e a sensibilidade. As medidas mais utilizadas nos trabalhos (Lara & Labrador, 2013) são sumarizadas na Tabela 5-5.

Tabela 5-5: Sumário das métricas de avaliação utilizadas nesta pesquisa.

Acurácia	$\frac{TP + TN}{TP + FP + FN + TN}$	Pode ser obtida através da proporção dos resultados verdadeiros (TP e TN) com o total do número de instâncias da população
Erro	$1 - \text{acurácia}$	A medida de erro (<i>error</i>) pode ser utilizada para expressar o oposto. Ela denota o desvio da medida das amostras verdadeiras e pode ser obtida em termos de acurácia
Sensibilidade	$\frac{TP}{TP + FN}$	Conhecido como verdadeiros positivos (<i>recall</i>), mede o quão bom é o classificador para classificar exemplos positivos corretamente
Especificidade	$\frac{TN}{TN + FP}$	Na direção oposta, a medida de especificidade (<i>specificity</i>), mostra a habilidade de classificar corretamente os exemplos negativos
Precisão	$\frac{TP}{TP + FP}$	é a taxa TP em relação a todos os positivos classificados
Medida F1	$2 \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}}$	é uma medida geral da acurácia do classificador que combina a precisão e sensibilidade

5.1.4. Cenários de avaliação

Para entender como diversos aspectos adotados neste trabalho influenciam no desempenho do método proposto, foram criados seis cenários de avaliação: A1, B1, C1, D1, E1 e F1.

- O cenário A1 avalia os parâmetros utilizados pelos métodos de discretização;
- O cenário B1 avalia os métodos de fusão de dados;
- O cenário C1 avalia a seleção de sensores por localização do smartphone;
- O cenário D1 avalia os métodos de representação simbólica;

- O cenário E1 avalia o desempenho do método proposto com outros trabalhos da literatura.
- O cenário F1 avalia a capacidade de generalização e especialização dos trabalhos.

Os algoritmos de representação simbólica (SAX e SFA) precisam de parâmetros como os tamanhos da palavra, do alfabeto, do segmento e da janela. Esses parâmetros possuem influência direta no desempenho do método de classificação. Por esse motivo, o Cenário A1 consiste na combinação de parâmetros utilizados pelo método HAR-SR, visando descobrir a melhor combinação para que o método seja executado com bom desempenho nos outros cenários. A escolha correta dos parâmetros deve influenciar tanto em tempo de execução quanto no desempenho de classificação.

No cenário B1, serão avaliados os métodos de fusão de dados em cada método de discretização. Os resultados deste cenário devem mostrar quais métodos extraem as melhores representações dos sinais dos sensores.

No cenário C1, quatro conjuntos de diferentes sensores (SG1, SG2, SG3, SG4) são utilizados para treinar e testar o método de classificação. Os resultados obtidos nesse cenário devem mostrar qual o impacto que a adição de um novo sensor possui no desempenho do método de classificação. Outro resultado obtido nesse cenário mostra como cada conjunto de sensor se comporta de acordo com as posições em que está localizado o smartphone.

No cenário D1, os métodos de representação simbólica SAX e SFA são avaliados. Esse cenário será útil para mostrar as diferenças entre as representações e seu impacto no desempenho do método de classificação.

O cenário E1 avalia o desempenho do método proposto, o HAR-SR, com outros três trabalhos. O primeiro trabalho utiliza a abordagem de extração manual de características e possui 145 características pertencentes ao domínio do tempo e frequência. Outros dois trabalhos são abordagens de extração automática de características pertencentes ao domínio discreto e similares ao método proposto.

Por fim, o cenário F1 é utilizado para avaliar a capacidade de generalização e especialização dos métodos para o modelo generalizado e o modelo personalizado. O modelo generalizado não mistura dados de um mesmo indivíduo na base de treino e teste. O modelo personalizado realiza a validação cruzada nos dados que pertencem a um mesmo indivíduo.

5.2. Cenário A1: avaliação de parâmetros

Os algoritmos de representação simbólica (SAX e SFA) possuem parâmetros como os tamanhos da palavra, do alfabeto, do segmento e da janela. A complexidade do método pode evoluir conforme o crescimento do valor dessas variáveis. Por exemplo, com um espaço de amostras de tamanho 256 é possível gerar um modelo menos complexo, no entanto, o espaço de amostras de 46.656 pode ser necessário e mais representativo quando problemas mais complexos precisam ser resolvidos. A Tabela 5-6 mostra o crescimento do espaço de amostras conforme os valores de tamanho de palavra e alfabeto. Por questões de desempenho este trabalho limita o tamanho máximo de palavra e alfabeto ao valor 6.

Tabela 5-6: Variação dos parâmetros do tamanho das palavra e tamanho do alfabeto

Palavra (ω)	Alfabeto (α)	Espaço de amostras	Intervalo
4 (<i>a, b, c, d</i>)	4 (<i>a, b, c, d</i>)	256	aaaa-dddd
4 (<i>a, b, c, d</i>)	8 (<i>a, b, c, d, e, f, g, h</i>)	4096	aaaa-hhhh
6 (<i>a, b, c, d, e, f</i>)	4 (<i>a, b, c, d</i>)	4096	aaaaaa-dddddd
6 (<i>a, b, c, d, e, f</i>)	6 (<i>a, b, c, d, e, f</i>)	46.656	aaaaaa-ffffff
8 (<i>a, b, c, d, e, f, g, h</i>)	4 (<i>a, b, c, d</i>)	65.536	aaaaaaaa-dddddddd
8 (<i>a, b, c, d, e, f, g, h</i>)	8 (<i>a, b, c, d, e, f, g, h</i>)	16777216	aaaaaaaa-hhhhhhhh

Tabela 5-7: A tabela apresenta os parâmetros utilizados no cenário A1.

Tabela de parâmetros	
Métodos de fusão de dados (φ)	Magnitude, PCA, Concatenação
Sensores	Acelerômetros, Giroscópio e Magnetômetro
Método de representação simbólica ($\hat{\mathbf{C}}$)	SAX, SFA
Tamanho de palavra (ω)	4, 6
Tamanho de alfabeto (α)	4, 6, 8
Tamanho de segmento (\mathbf{T})	2,5s (aproximadamente 128 pontos)
Tamanho de janela (\mathbf{w})	$\frac{50}{100} \cdot \mathbf{T}$ (50% janela deslizante)
Base de dados	SHOAIB

O procedimento para avaliar os melhores parâmetros conta com os resultados obtidos através do cruzamento das variáveis da

Tabela 5-7. A média dos resultados é calculada e apresentada na Figura 5-6.

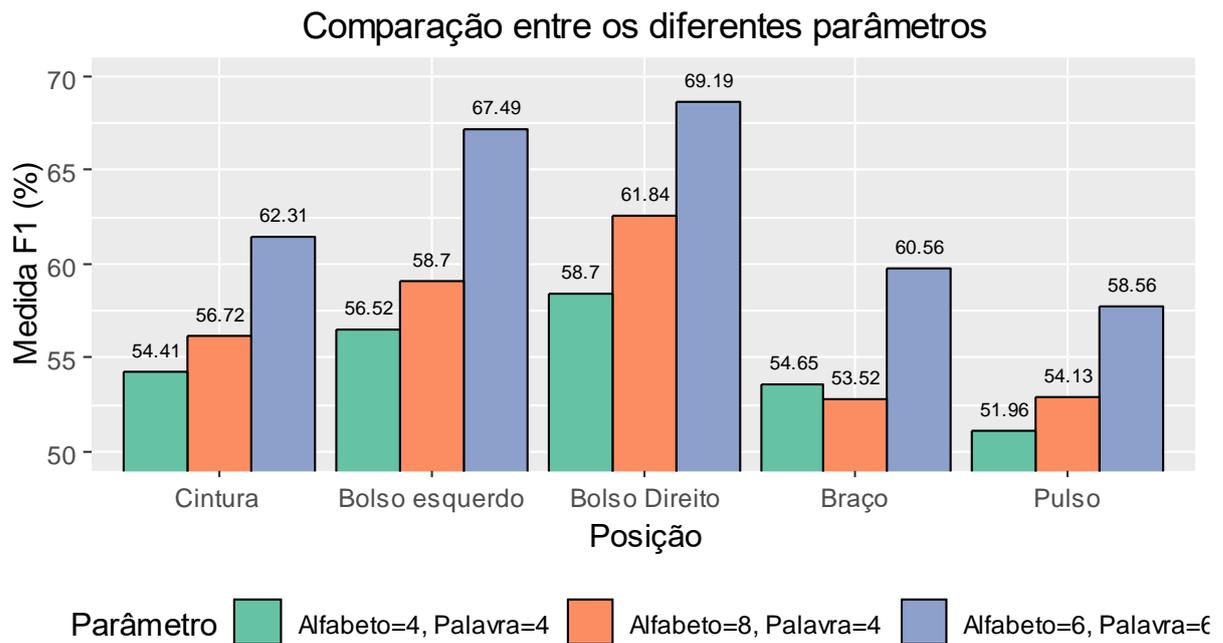


Figura 5-6. Média de desempenho dos algoritmos de representação simbólica com variação nos parâmetros ω e α para cada posição da base de dados SHOAIB.

Os experimentos realizados para selecionar o melhor conjunto de parâmetros mostram que para o intervalo de valores definidos na

Tabela 5-7, os melhores resultados são aqueles cujo valor atribuído a $\omega = 6$ e $\alpha = 6$. Este estudo foi realizado na base de dados SHOAIB por possuir mais diversidade entre os parâmetros de posição e sensores. A média dos resultados para cada posição pode ser visto na Figura 5-6, que apresenta os melhores parâmetros cujo valor atribuído a $\omega = 6$ e $\alpha = 6$ com boa margem de distância. Por esse motivo, a comparação dos resultados apresentados neste capítulo adota por padrão $\omega = 6$ e $\alpha = 6$.

5.3. Cenário B1: fusão de dados

Essa seção tem por objetivo a avaliação de três métodos de fusão de dados: A Magnitude, O PCA e a Concatenação de sinais. Os métodos de fusão de dados possuem um papel importante: prover o sinal com mais qualidade se comparado com cada sinal individualmente. Outro objetivo desses métodos é a redução da dimensionalidade dos dados de entrada. Esse estudo é importante porque o cálculo da complexidade dos algoritmos que compõem o HAR-SR, o SAX e SFA, são calculadas com base nos dados de entrada, ou seja, os dados dos sensores. Nesse caso, é necessário que o método escolhido transforme um sinal multidimensional em um sinal unidimensional, preservando as características dos sinais.

Os parâmetros utilizados no cenário B1 são apresentados na Tabela 5-8. Os resultados desse cenário mostram quais métodos extraem as melhores representações dos sinais dos sensores.

Tabela 5-8: Os métodos utilizados na fusão de dados são a Magnitude, o PCA e a concatenação. Os parâmetros utilizados são: $\hat{\mathbf{C}} = \text{SAX}$, $T=128$, $w = \frac{50}{100} \cdot T$ (50% janela deslizante), $\omega = 6$, $\alpha = 6$

Métodos de fusão de dados	
Métodos de fusão de dados	Magnitude, PCA, Concatenação
Algoritmos de representação simbólica ($\hat{\mathbf{C}}$)	SAX
Base de dados	SHOAIB
Tamanho do segmento (\mathbf{T})	2,5 segundos (128 pontos)
Tamanho da janela deslizante (\mathbf{w})	$\frac{50}{100} \cdot T$ (50% janela deslizante)
Palavra (ω)	6
Alfabeto (α)	6

A base de dados SHOAIB foi selecionada porque possui dados de três sensores, o acelerômetro, o giroscópio e o magnetômetro. Além disso, possui uma variedade de posições, permitindo assim que muitos cenários sejam analisados. Os sensores disponíveis para teste foram agrupados e avaliados em cada posição disponível na base. O grupo de sensores é apresentado na Tabela 5-9.

Tabela 5-9: Agrupamento do conjunto de sensores para realização dos testes do cenário C1.

Grupo	Sensores pertencentes ao grupo
SG1	Acelerômetro
SG2	Acelerômetro e Giroscópio
SG3	Acelerômetro, Giroscópio e Magnetômetro
SG4	Giroscópio

Os resultados apresentados a seguir são baseados na média dos resultados obtidos no LOSO nos dados obtidos de cinco posições: cintura, bolso esquerdo, bolso direito, braço e pulso.

Comparação entre os métodos de fusão de dados na posição da cintura

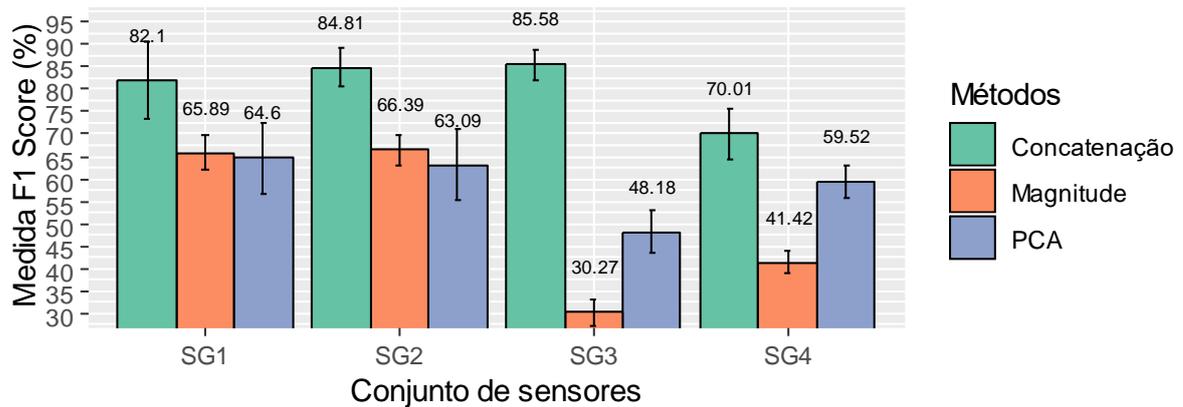


Figura 5-7: Comparação entre os métodos de fusão de dados aplicados a cada conjunto de sensores. A base de dados é a SHOAI B na posição da cintura.

Comparação entre os métodos de fusão de dados para a posição no bolso esquerdo

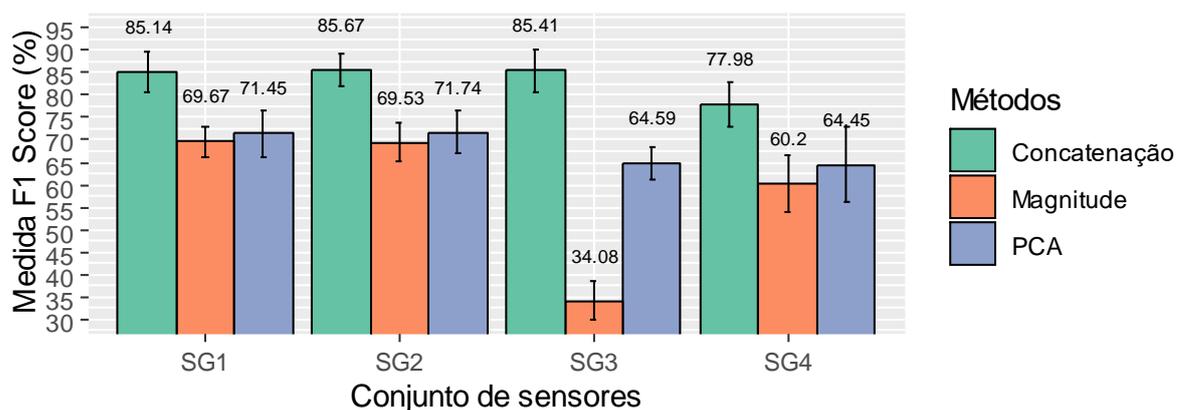


Figura 5-8: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAI B com o smartphone localizado no bolso esquerdo.

Comparação entre os métodos de fusão de dados para a posição no bolso direito

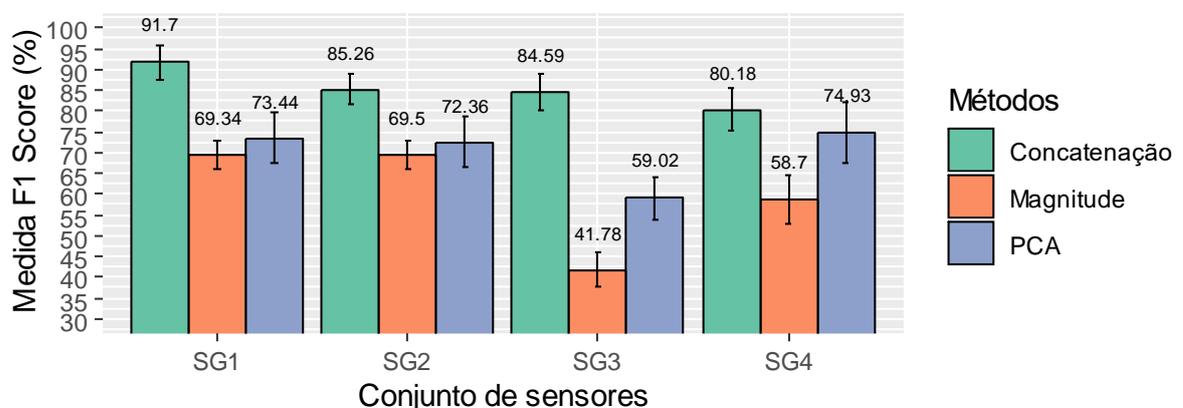


Figura 5-9: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAI B com o smartphone localizado no bolso direito.

Comparação entre os métodos de fusão de dados na posição do braço

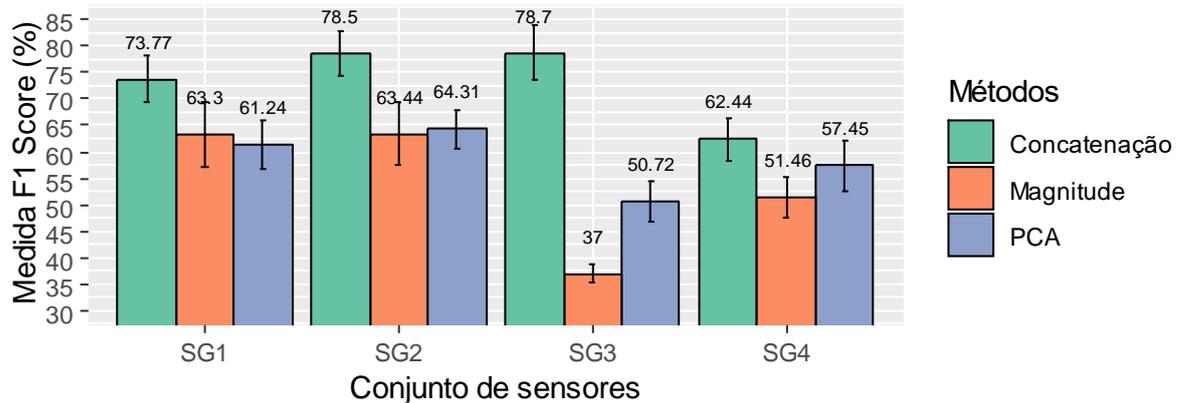


Figura 5-10: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAIIB com o smartphone localizado no braço.

Comparação entre os métodos de fusão de dados na posição do pulso

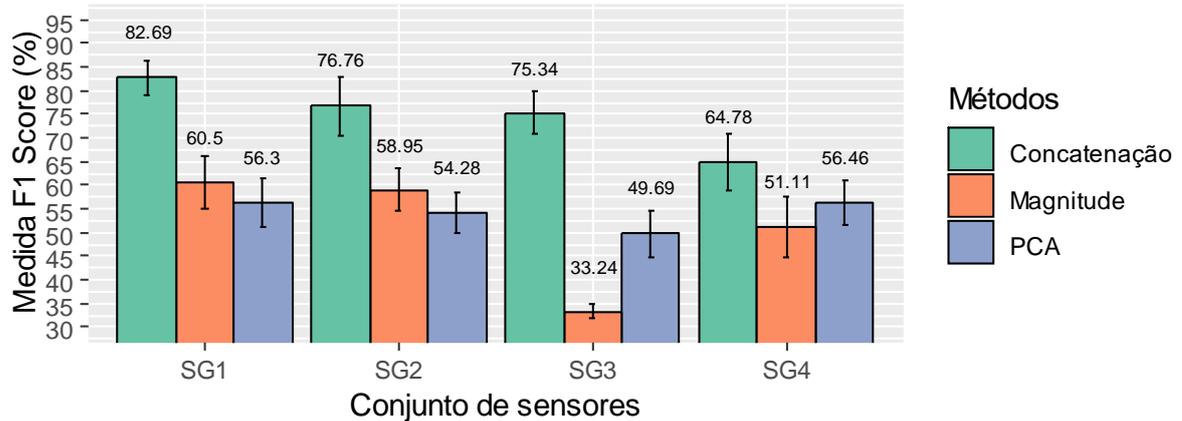


Figura 5-11: Comparação entre os métodos de fusão de dados aplicados em cada conjunto de sensores. A base de dados é a SHOAIIB com o smartphone localizado no pulso dos indivíduos.

Os resultados apresentados mostram que o método de concatenação apresenta os melhores resultados na medida F1 em todos os grupos de sensores e, em muitos casos, com alta margem de distância. Isso acontece porque o método de concatenação não realiza nenhum tipo de fusão, apenas agrega os sinais dos sensores através de uma mudança de espaço. Dessa forma, um sinal que possuir uma série com 128 pontos em cada eixo passará a ter 384 pontos e um único eixo. Assim, sem nenhuma “perda” de informações, como é o caso do método de agregação, os algoritmos de representação simbólica conseguem capturar melhor as características do sinal.

Tanto a Magnitude quanto o PCA aplicam redução de dimensionalidade nos sinais, o que implica também em perda de informações sobre as séries, que levam a uma queda no desempenho do método. Outra justificativa pode ser relacionada com o fato desses métodos não serem eficazes na construção de um sinal representativo para que o modelo consiga extrair padrões.

Os resultados da comparação entre a magnitude e o PCA são similares para o conjunto de sensores SG1 e SG2. No entanto, a Magnitude perde com alta margem no grupo SG3, no qual o sensor de magnetômetro está inserido. Utilizando apenas o giroscópio (SG4) o PCA apresenta melhores resultado que a Magnitude.

Os resultados mostram que o sensor acelerômetro apresenta um bom desempenho, sendo suficiente para reconhecer atividades em todas as posições. Um resultado positivo desse experimento foi descobrir que o método HAR-SR consegue reconhecer as atividades com boa precisão utilizando apenas o sensor acelerômetro. O acréscimo ou combinação de novos sensores, como o giroscópio e magnetômetro possuem pouco ou nenhum impacto no desempenho, se considerada a margem de erro, salvo em alguns casos. Além disso, a inclusão de novos sensores começa a degradar o desempenho do método, conforme apresentado na Figura 5-11.

Avaliando apenas do sensor do giroscópio a no melhor cenário, muito abaixo da média se comparado com o sensor acelerômetro, por exemplo. A adição de novos sensores significa mais custo de processamento e mais uso de memória, que só seria relevante se o impacto na classificação fosse realmente perceptível. Esse estudo mostra que a escolha correta de um determinado sensor, desde que conhecida a localização do smartphone, pode resultar em um melhor desempenho do método.

5.4. Cenário C1: avalia a seleção de sensores por posição

Os resultados do cenário C1 mostram como cada conjunto de sensor se comporta de acordo com as posições em que ele está localizado o smartphone. A Tabela 5-10 mostra o conjunto de posições avaliadas nesse cenário.

Tabela 5-10: Posições em que os dados do smartphone foram coletados.

Tabela de posições	
Posição	Parâmetros
Cintura	$\varphi =$ Concatenação simples $T = 2.5$ segundos (128 pontos)
Bolso esquerdo	
Bolso direito	
Braço	$\omega = 6$
Pulso	$\alpha = 6$
Todas as posições	

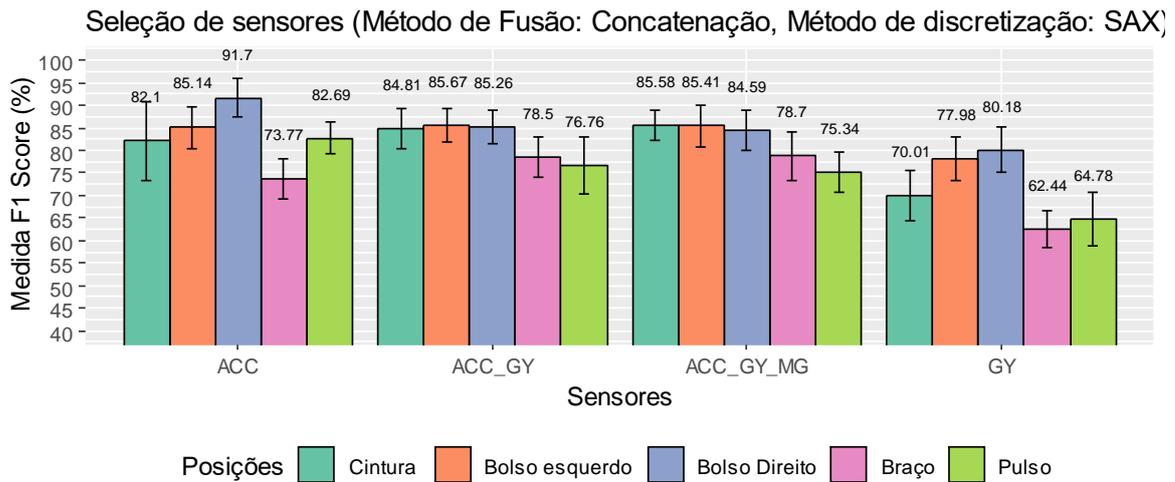


Figura 5-12: Desempenho de classificação para cada grupo de sensores em cada posição da base SHOAIB. o método de discretização é o SAX.

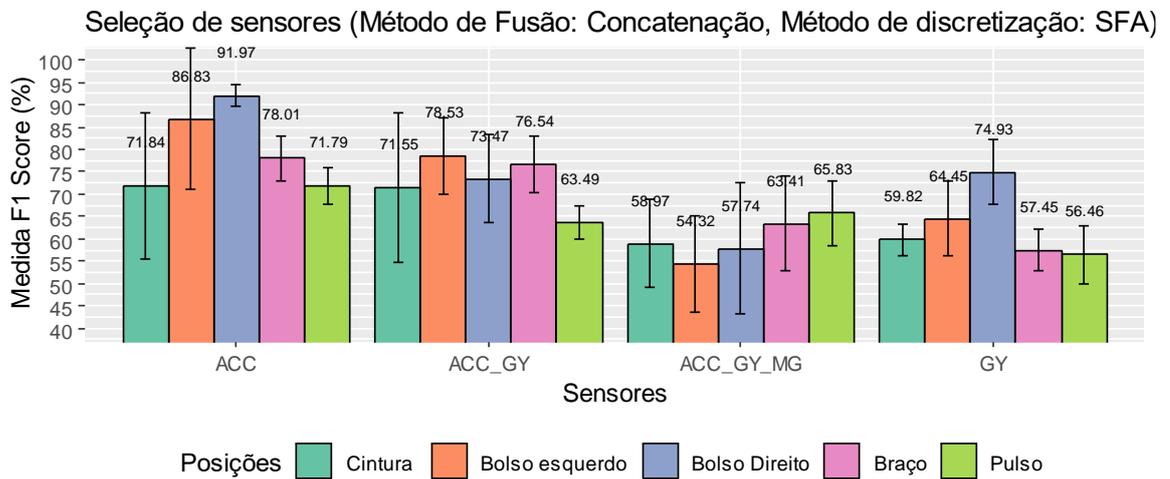


Figura 5-13: Desempenho de classificação para cada grupo de sensores em cada posição da base SHOAIB. o método de discretização é o SFA.

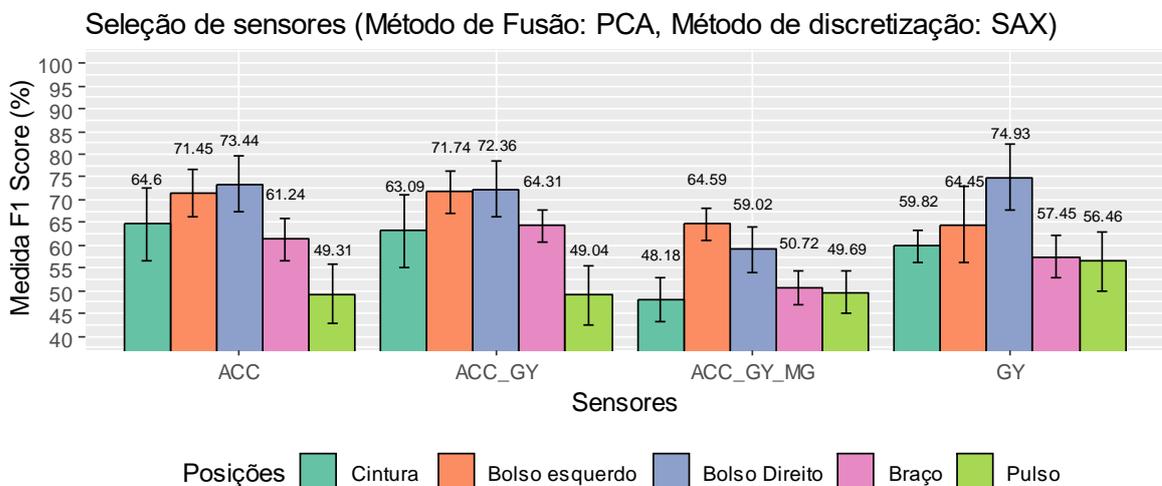


Figura 5-14: Desempenho de classificação para cada grupo de sensores em cada posição da base SHOAIB. o método de discretização é o SAX.

Os resultados desta seção mostram os algoritmos de discretização apresentam diferenças no seu desempenho de classificação, e apresentam comportamentos diferentes de acordo com a escolha dos sensores. Por exemplo, os resultados apresentados na Figura 5-16 mostram que existe algum ganho no desempenho do método ao se adicionar novos sensores para reconhecer atividades na posição da cintura e braço. Nas posições do bolso esquerdo não existe ganho e para as posições do bolso direito e pulso ocorre uma perda no desempenho do método. Além disso, na posição do bolso esquerdo e direito o sensor de giroscópio apresenta um bom desempenho.

A Figura 5-13 apresenta os resultados para o método SFA, que em todos os testes têm seu desempenho degradado quando novos sensores são introduzidos. Na comparação com o algoritmo SAX (Figura 5-16), quando utilizado apenas o sensor de acelerômetro, possui desempenho similar para as posições do bolso esquerdo e direito, um melhor desempenho para a posição do braço e nas posições da cintura e pulso apresenta desempenho pior. Utilizando apenas o sensor de acelerômetro, os melhores resultados são obtidos com a utilização do SAX.

O resultado apresentado na Figura 5-14 serve como base para avaliar se o comportamento do PCA segue o mesmo padrão de desempenho da concatenação simples, para o algoritmo SAX. De semelhante ao mostrado na Figura 5-13, não existe vantagem na adição de mais sensores. Os resultados são similares para a Magnitude, que não foi apresentado.

Esses resultados mostram que adicionar mais sensores resulta em um pequeno ou em nenhum impacto na medida F1, que em muitos casos podem ocasionar na degradação de desempenho do método de classificação. A posição da cintura é a maior beneficiada quando mais sensores são adicionados. Esses resultados são importantes porque podem servir como base para um método que selecione de forma adaptativa os sensores de acordo com a posição no qual o smartphone está presente.

5.5. Cenário D1: Avaliação dos algoritmos de representação simbólica

O cenário D1 apresenta a avaliação do desempenho dos algoritmos de discretização. Para esse teste, somente o sensor acelerômetro e o método de concatenação foram utilizados. Esse cenário será útil para mostrar as diferenças entre as representações e seu impacto no desempenho do método de classificação. Os parâmetros utilizados na avaliação são apresentados na Tabela 5-11.

Tabela 5-11: Tabela de parâmetros utilizada na avaliação dos algoritmos de representação simbólica.

Tabela de parâmetros	
Métodos de fusão de dados (φ)	Concatenação
Sensor	Acelerômetro
Algoritmos de representação simbólica ($\hat{\mathbf{C}}$)	SAX, SFA
Tamanho de palavra (ω)	6
Tamanho de alfabeto (α)	6
Tamanho de segmento (\mathbf{T})	2.5 segundos (128 pontos)
Tamanho de janela (\mathbf{w})	$\frac{50}{100} \cdot \mathbf{T}$ (50% janela deslizante)

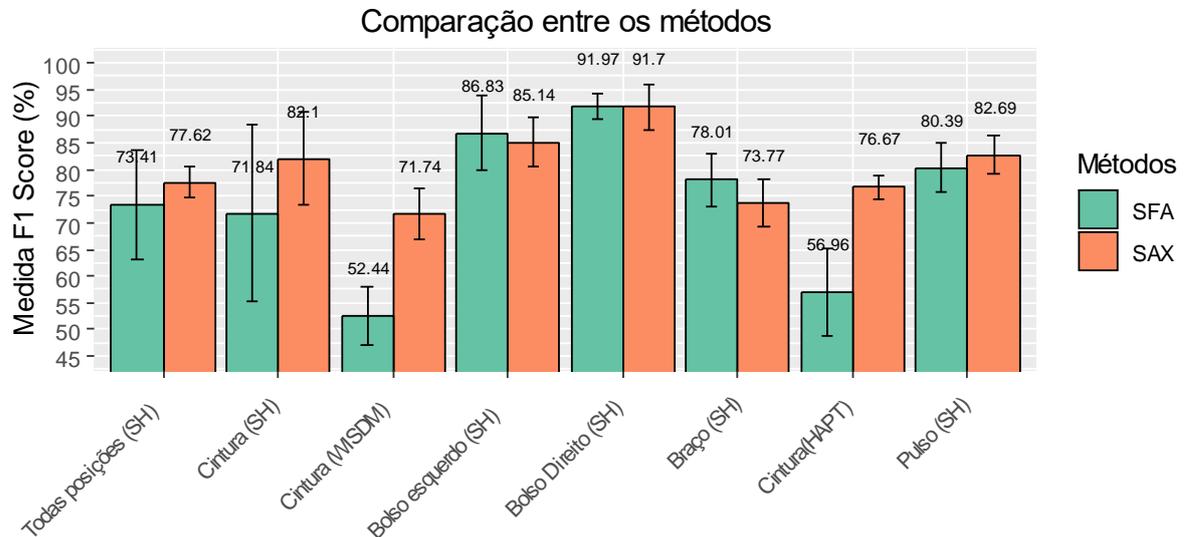


Figura 5-15: Desempenho dos métodos de discretização SAX e SFA em cada base de dados no cenário D1.

A primeira conclusão dos resultados apresentados na Figura 5-15 é que dentre os dois algoritmos, o **SAX** apresenta os melhores resultados na maioria das posições avaliadas. Observando somente os resultados da base **SHOAIB**, o **SAX** mostra resultado similar ao **SFA**, dentro da margem de erro. Na base de dados **WISDM** e **HAPT**, o **SAX** apresenta melhores resultados com ampla margem. Os resultados para os métodos de fusão de dados utilizando a magnitude e **PCA** são mostrados no apêndice.

O bom desempenho do **SAX** em relação ao **SFA** no contexto **RAH** é um resultado interessante porque mostra que nesse contexto de reconhecimento de atividades o algoritmo **SAX** consegue extrair padrões que são mais gerais para todo o conjunto de indivíduos. Em outros domínios mostrados na literatura o **SFA** apresenta melhores resultados (Schäfer, 2016) (Bagnall, et al., 2015). Um dos motivos possíveis é fundamentado na forma de criação da tabela **MCB**, que é responsável por determinar

os pontos de interrupção de acordo com as variáveis w e a . A construção da tabela MCB é supervisionada e depende da qualidade dos dados da base de dados. Nesse contexto, se a tabela é criada utilizando apenas dados de indivíduos similares, por exemplo, deverá produzir uma tabela que representa bem aquele grupo. No entanto, não deverá ter bom desempenho quando indivíduos com particularidades em seus padrões de atividades são inseridos no teste. Esse processo de treinamento da tabela MCB é dependente da base de dados de treino. Isso influencia no processo de discretização menos representativo para um grupo mais genérico de indivíduos.

O SAX utiliza uma tabela de pontos de interrupção fixa, com valores baseados em uma distribuição gaussiana. Por esse motivo, a capacidade de generalização do modelo deve ser maior que em um modelo que tome as informações somente dos conhecimentos de uma base limitada de dados.

5.6. Cenário E1: comparação do HAR-SR com outros trabalhos

Este cenário compara o desempenho dos trabalhos encontrados na literatura com o método HAR-SR proposto nesta pesquisa. O primeiro trabalho é baseado nas características manuais extraídas do domínio do tempo e frequência (TF). A lista das funções matemáticas utilizadas no processo de extração de características do sinais dos sensores é encontrada na Tabela 5-12. Os detalhes da complexidade computacional do método TF pode ser encontrado no Apêndice A.

Tabela 5-12: Sumário das características utilizadas pelo método TF.

Domínio	Características
Tempo	min, max, amplitude, amplitude peak, sum, absolute sum, Euclidian norm, mean, absolute mean, mean square, mean absolute deviation, sum square error, variance, standard deviation, Pearson coefficient, zero crossing rate, correlation, cross-correlation, auto-correlation, skewness, kurtosis, area, absolute area, signal magnitude mean, absolute signal magnitude mean, magnitude difference function.
Frequência	Energy, energy normalized, power, centroid, entropy, DC component, peak, coefficient sum.

O segundo trabalho utiliza um método de classificação do domínio discreto, o SAX-VSM. O SAX-VSM utiliza uma técnica chamada de $tf - idf$ no histograma de frequência símbolos de cada classe de atividades para obter uma matriz de frequência ponderada. O resultado é uma matriz que contém uma instância para cada atividade e será o

conjunto de características utilizadas pelo método de classificação na geração do modelo de reconhecimento de atividades.

Por fim, o terceiro trabalho, também do domínio discreto, é o BOSS-VS. Assim como o SAX-VSM, o BOSS-VS utiliza uma técnica chamada de *tf – idf* no histograma de frequência símbolos para obter uma matriz de frequência ponderada para cada atividade. Essa matriz de frequência ponderada será utilizada pelo método de classificação na geração do modelo de reconhecimento de atividades.

Neste cenário, apenas o sensor de acelerômetro foi utilizado e os parâmetros escolhidos para os métodos discretos foram aqueles que apresentaram os melhores resultados no cenário A1, apresentado na Seção 5.2. Os parâmetros utilizados na comparação dos métodos são apresentados na Tabela 5-13.

Tabela 5-13: Parâmetros utilizados em cada um dos trabalhos avaliados no cenário E1. *TF é um método de classificação que utiliza características do domínio do tempo e frequência.

Método	Nº de características	Posição	Algoritmo de classificação	Medida de distância	Parâmetros
HAR-SR	15	Cintura Bolso esquerdo Bolso direito Braço Pulso independente	K-NN (k=3)	Similaridade dos cossenos	φ = concatenação; \hat{C} = SAX ω = 6; α = 6; T = 128; w = 50%
*TF	145	Cintura Bolso esquerdo Bolso direito Braço Pulso independente	K-NN (K=3)	Distância euclidiana	-
SAX-VSM	46.656	Cintura Bolso esquerdo Bolso direito Braço Pulso independente	K-NN (K=1)	Similaridade dos cossenos	φ = concatenação; ω = 6; α = 6; T = 128; w = 50%
BOSS-VS	46.656	Cintura Bolso esquerdo Bolso direito Braço Pulso independente	K-NN (K=1)	Similaridade dos cossenos	φ = concatenação; ω = Variável; α = Variável; T = 128; w = 50%

O primeiro resultado apresentado na Figura 5-16 mostra que o método TF e o método HAR-SR possuem um desempenho similar na classificação de atividade em todas as posições. Isso mostra, em um primeiro momento, que o método HAR-SR possui as mesmas capacidades do método TF, mas utiliza um número menor de características. Os detalhes de cada uma das posições serão discutidos na Seção 5.6.1 através da matriz de confusão.

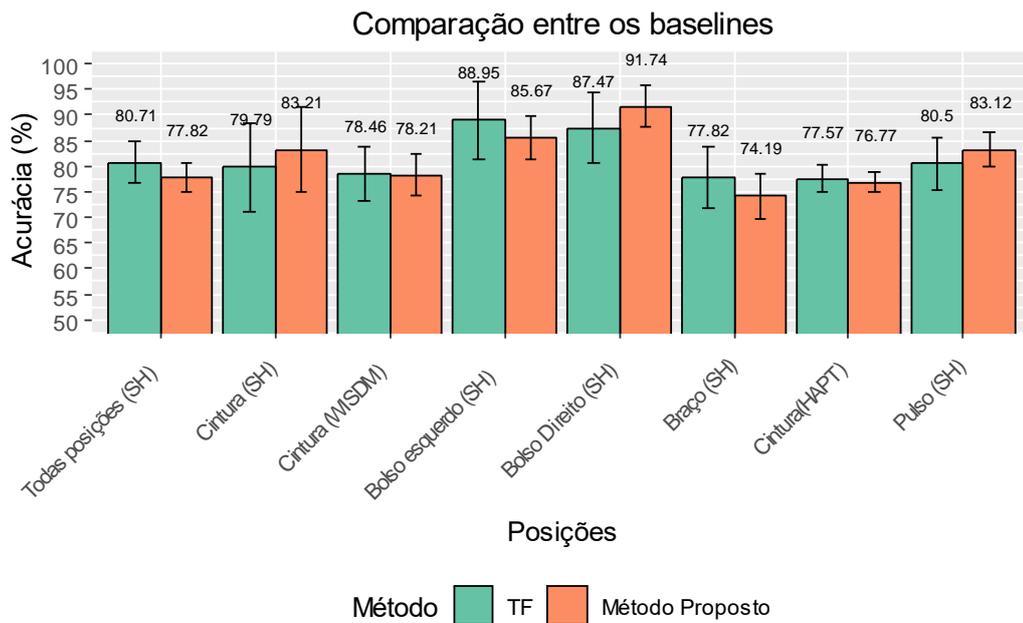


Figura 5-16: Comparação entre o método TF e o método proposto HAR-SR. Apenas o sensor do acelerômetro foi utilizado.

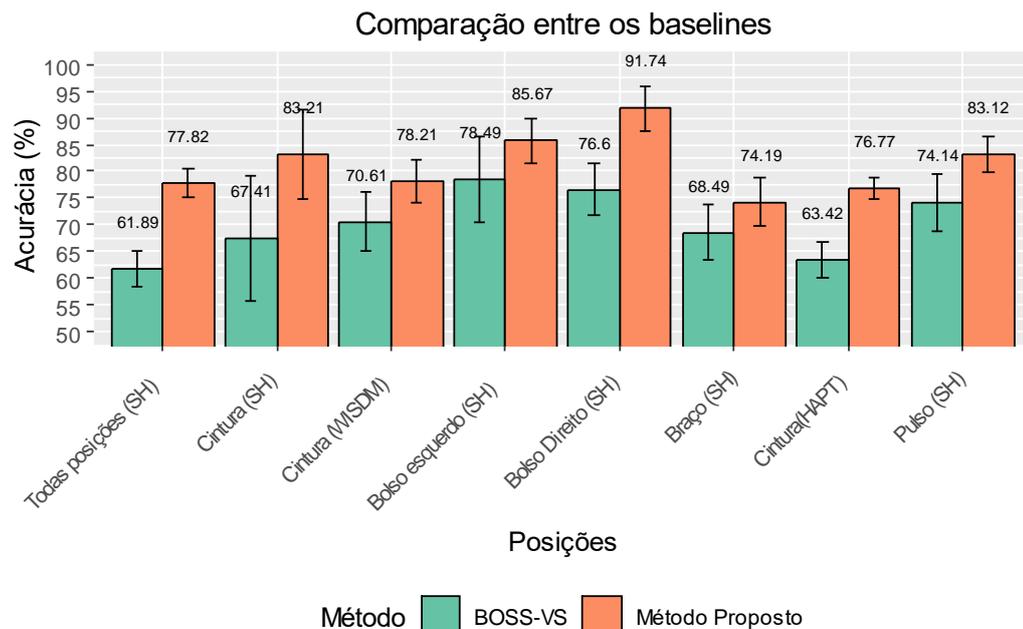


Figura 5-17: Comparação entre o método BOSS – VS e o método proposto HAR-SR. Apenas o sensor do acelerômetro foi utilizado.

O segundo resultado deste cenário é mostrado na Figura 5-17. Comparado com o modelo BOSS-VS, o método HAR-SR apresenta os melhores resultados em todas as posições analisadas com boa margem de diferença. Esse resultado já é um indício de que o modelo $tf - idf$ não agrega características discriminativas o suficientes para que o método apresente um desempenho similar ao HAR-SR.

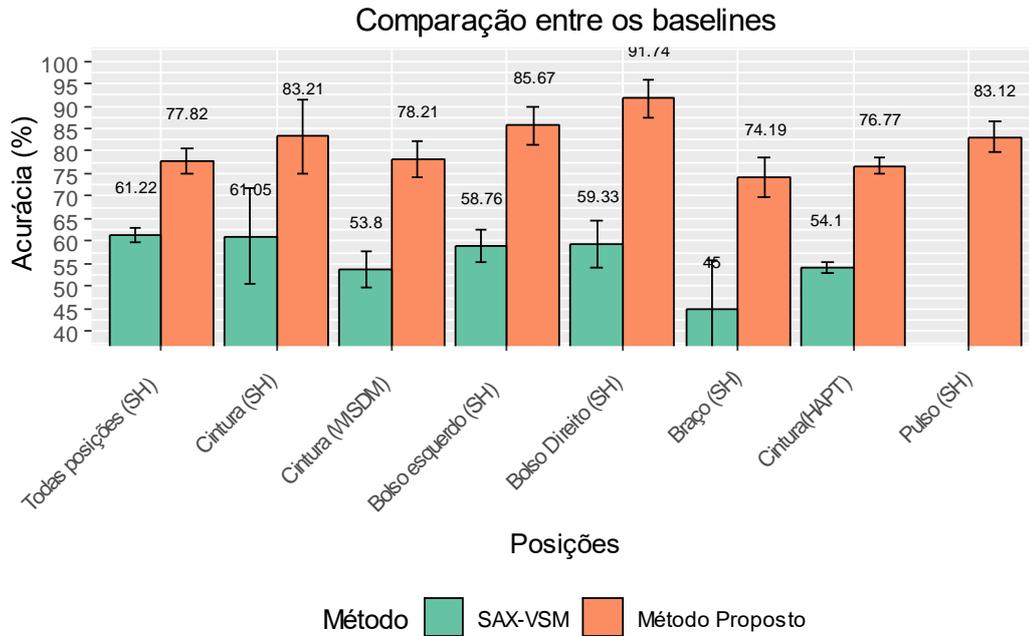


Figura 5-18 Comparação entre o método SAX – VSM e o método proposto HAR-SR. Apenas o sensor do acelerômetro foi utilizado.

O último resultado apresentado na Figura 5-18 mostra que o HAR-SR é melhor no reconhecimento de atividades com ampla margem em relação ao SAX-VSM. Além disso, a diferença de desempenho entre o BOSS-VS e o SAX-VSM já é esperada, conforme apresentada em outros trabalhos (Schäfer, 2016).

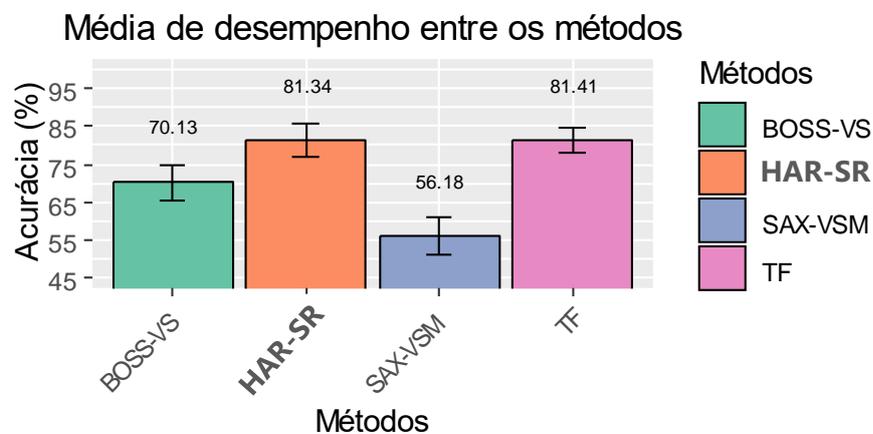


Figura 5-19: Média do desempenho de classificação dos métodos utilizados no cenário E1. Os parâmetros utilizados podem ser encontrados na Tabela 5-13.

A média de desempenho dos trabalhos utilizados no cenário E1 é mostrado na Figura 5-19. Esse resultado mostra a similaridade no desempenho de classificação entre os métodos TF e HAR-SR. O BOSS-VS apresenta um desempenho mediano na classificação de atividades, mas abaixo do desempenho atingido por muitos métodos da literatura. Os resultados obtidos pelo SAX-VSM deixam claro que ele não é a melhor escolha para o contexto RAH, no qual apresentou baixo desempenho em todas as posições avaliadas.

Os resultados obtidos pelo método TF e HAR-SR são muito similares, e como os resultados apresentam sobreposição nos intervalos de confiança, considera-se empate técnico. Por esse motivo, a matriz de confusão será utilizada para apontar em quais classes os algoritmos possuem mais dificuldades na classificação. Essa análise deve mostrar as particularidades desses dois métodos e como novos problemas podem ser tratados no futuro.

5.6.1. Interpretação da matriz de confusão

Esta seção analisa e compara a matriz de confusão dos métodos HAR-SR e TF. Os resultados apresentados são referentes a base de dados SHOAIB, para cada posição.

Tabela 5-14: Parâmetros utilizados na geração dos resultados apresentados na matriz de confusão no cenário E1.

Base de dados	Sensor	Posições
SHOAIB	Acelerômetro	Cintura
		Bolso esquerdo
		Bolso direito
		Braço
		Pulso
		independente

A primeira matriz de confusão mostra o desempenho de classificação do método HAR-SR e do método TF, na posição da cintura. Nessa posição o método HAR-SR apresenta um bom desempenho, com erros de classificação principalmente nas atividades estacionárias², ou seja, atividades estáticas como levantando ou sentando. Um outro

² Não confundir o termo “estacionário” utilizado nesta seção com as séries temporais estacionárias.

tipo de comportamento é observado no método TF, que encontra problemas em diferenciar as atividades não-estacionárias, como andar, correr, subir escadas e descer escadas. Essa diferença entre os tipos de atividades que esses métodos podem classificar já apontam uma grande diferença entre esses dois métodos.

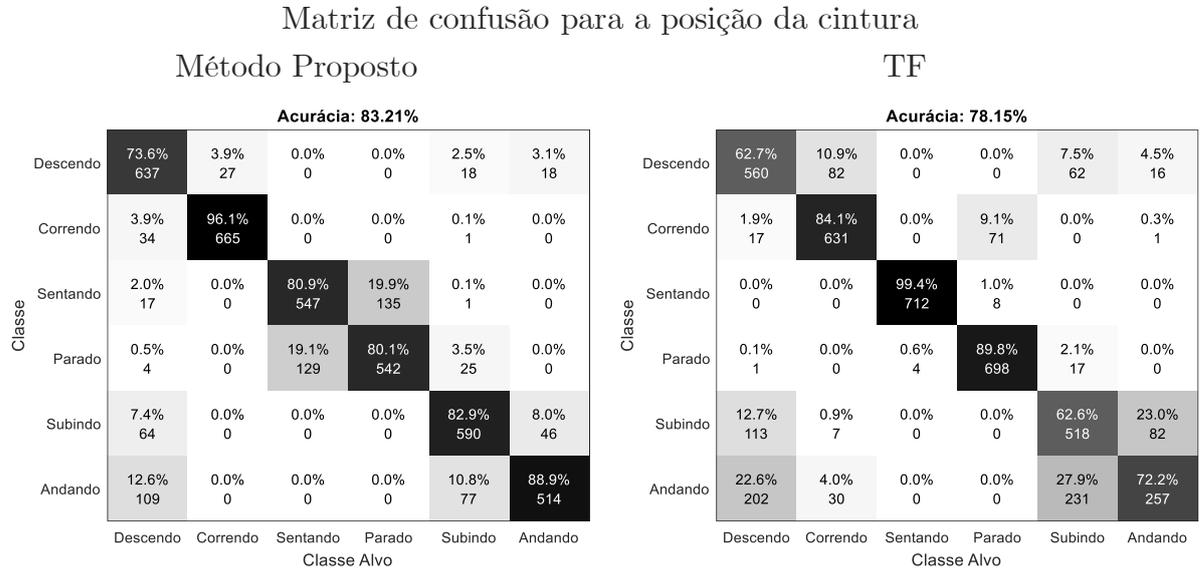


Figura 5-20: Matriz de confusão da base de dados SHOAIB utilizando dados do sensor de acelerômetro na posição da cintura. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, \hat{C} = SAX, ω = 6, α =6).

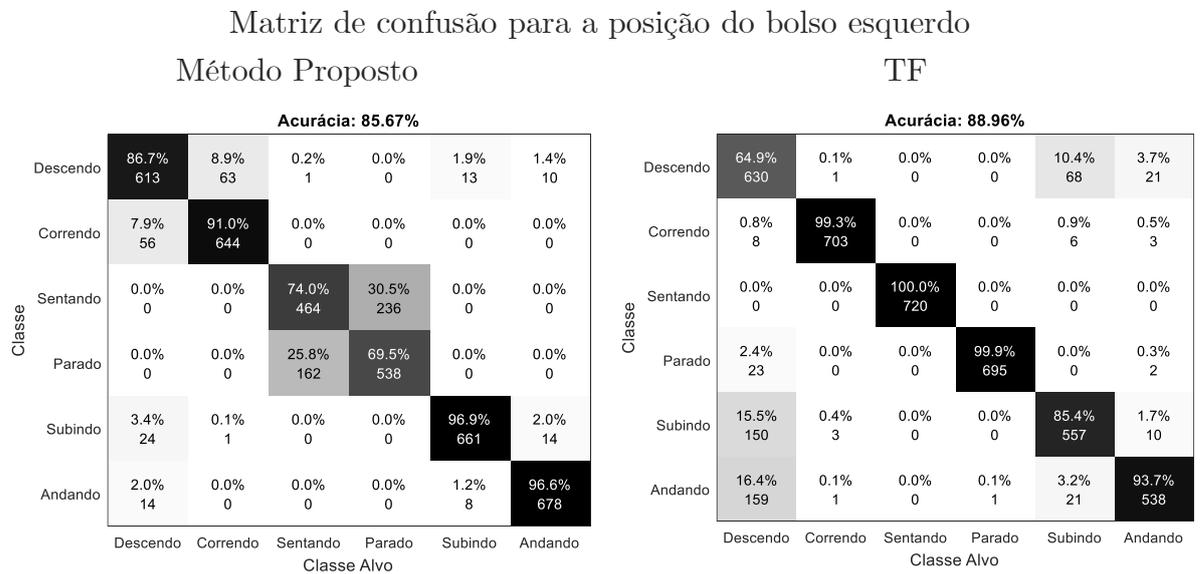


Figura 5-21: Matriz de confusão da base de dados SHOAIB utilizando dados do sensor de acelerômetro com o smartphone localizado na posição do bolso esquerdo. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, \hat{C} = SAX, ω = 6, α =6).

Analisando os resultados obtidos para a posição do bolso esquerdo (Figura 5-21), ambos os métodos apresentam comportamento semelhante ao apresentado pela posição da cintura. O método HAR-SR também mostra dificuldades em diferenciar as atividades estacionárias, parado e sentado, mas em geral, apresenta um bom desempenho na

classificação. O método TF apresenta erros de classificação em atividades não-estacionárias, principalmente em reconhecer a atividade descendo escadas.

Os resultados para a posição do bolso direito (Figura 5-22) são similares aos apresentados para o bolso esquerdo para o método TF. No entanto, houve uma diferença notável no reconhecimento de atividades estacionárias.

Matriz de confusão para a posição do bolso direito

		Método Proposto						TF					
		Acurácia: 91.74%						Acurácia: 87.48%					
Classe		Classe Alvo						Classe Alvo					
		Descendo	Correndo	Sentando	Parado	Subindo	Andando	Descendo	Correndo	Sentando	Parado	Subindo	Andando
Descendo		85.3% 626	4.0% 28	0.0% 0	0.0% 0	5.6% 38	1.1% 8	64.1% 606	2.9% 21	0.0% 0	0.0% 0	10.0% 61	5.4% 32
Correndo		5.2% 38	95.7% 662	0.0% 0	0.0% 0	0.0% 0	0.0% 0	1.1% 10	96.7% 706	0.0% 0	0.0% 0	0.5% 3	0.2% 1
Sentando		0.0% 0	0.0% 0	94.2% 630	9.6% 70	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 720	0.0% 0	0.0% 0	0.0% 0
Parado		0.1% 1	0.0% 0	5.8% 39	90.4% 660	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 719	0.2% 1	0.0% 0
Subindo		6.8% 50	0.3% 2	0.0% 0	0.0% 0	91.0% 617	4.4% 31	20.6% 195	0.3% 2	0.0% 0	0.0% 0	81.0% 494	4.9% 29
Andando		2.6% 19	0.0% 0	0.0% 0	0.0% 0	3.4% 23	94.4% 658	14.2% 134	0.1% 1	0.0% 0	0.0% 0	8.4% 51	89.6% 534

Figura 5-22: Matriz de confusão da base de dados SHOAI B utilizando dados do sensor de acelerômetro com o smartphone localizado na posição do bolso direito. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, \hat{C} = SAX, ω = 6, α = 6).

Uma das hipóteses para este resultado é que essa diferença no desempenho de classificação para as atividades estacionárias pode estar relacionada com a informação de que os indivíduos podem ser destros ou canhotos. Essa diferença entre os indivíduos pode levar a comportamentos involuntários que geram movimentos diferentes. Essa “assinatura” pode ser perceptível pelo método (resultados similares tanto para o SAX quanto para o SFA), e é uma das possíveis explicações para este resultado. A segunda hipótese é que no momento da captura do sinal para essas posições os dispositivos estavam em posições relativamente diferentes, com ângulos diferentes, que possam ter gerado diferenças na captura do sinal e, por esse motivo, ocorreu diferença no desempenho de classificação.

Os resultados da Figura 5-23 mostram a posição do braço é ainda um desafio do método HAR-SR. Nessa posição, o método apresenta baixo desempenho de classificação das atividades estacionárias. Além disso, o método também confunde algumas atividades não-estacionárias. O método TF continua com um bom desempenho nas atividades estacionárias, mas tem problemas no reconhecimento de atividades não-estacionárias.

Matriz de confusão para a posição do braço

		Método Proposto				TF							
		Acurácia: 74.19%						Acurácia: 77.82%					
Classe		Classe Alvo						Classe Alvo					
		Descendo	Correndo	Sentando	Parado	Subindo	Andando	Descendo	Correndo	Sentando	Parado	Subindo	Andando
Descendo		74.2% 554	3.9% 28	0.0% 0	0.0% 0	12.2% 80	5.5% 38	86.2% 569	4.2% 31	0.3% 2	0.0% 0	7.1% 57	9.3% 61
Correndo		2.0% 15	95.9% 682	0.0% 0	0.0% 0	0.5% 3	0.0% 0	1.5% 10	95.8% 709	0.0% 0	0.0% 0	0.0% 0	0.2% 1
Sentando		0.0% 0	0.0% 0	57.0% 366	44.3% 333	0.2% 1	0.0% 0	0.3% 2	0.0% 0	87.0% 640	10.6% 77	0.0% 0	0.2% 1
Parado		0.1% 1	0.0% 0	43.0% 276	55.7% 419	0.6% 4	0.0% 0	0.2% 1	0.0% 0	12.8% 94	84.7% 615	0.1% 1	1.4% 9
Subindo		15.3% 114	0.1% 1	0.0% 0	0.0% 0	78.1% 514	10.3% 71	3.0% 20	0.0% 0	0.0% 0	0.0% 0	58.9% 471	34.7% 229
Andando		8.4% 63	0.0% 0	0.0% 0	0.0% 0	8.5% 56	84.2% 581	8.8% 58	0.0% 0	0.0% 0	4.7% 34	33.8% 270	54.3% 358

Figura 5-23: Matriz de confusão da base de dados SHOAIB utilizando dados do sensor de acelerômetro com o smartphone localizado no braço. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, \hat{C} = SAX, ω = 6, α = 6).

Matriz de confusão para a posição do pulso

		Método Proposto				TF							
		Acurácia: 83.12%						Acurácia: 78.84%					
Classe		Classe Alvo						Classe Alvo					
		Descendo	Correndo	Sentando	Parado	Subindo	Andando	Descendo	Correndo	Sentando	Parado	Subindo	Andando
Descendo		78.5% 578	4.0% 28	0.9% 6	0.0% 0	8.8% 62	3.9% 26	79.1% 576	0.0% 0	0.2% 1	0.1% 1	11.6% 117	5.7% 25
Correndo		3.3% 24	95.9% 674	0.0% 0	0.0% 0	0.3% 2	0.0% 0	0.1% 1	91.0% 719	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Sentando		0.8% 6	0.0% 0	97.3% 674	2.3% 16	0.6% 4	0.0% 0	0.3% 2	9.0% 71	99.8% 646	0.1% 1	0.0% 0	0.0% 0
Parado		0.3% 2	0.0% 0	1.4% 10	97.6% 685	0.4% 3	0.0% 0	0.3% 2	0.0% 0	0.0% 0	96.6% 700	0.5% 5	3.0% 13
Subindo		9.6% 71	0.1% 1	0.4% 3	0.0% 0	61.7% 434	28.8% 191	12.9% 94	0.0% 0	0.0% 0	0.1% 1	49.2% 494	29.8% 131
Andando		7.5% 55	0.0% 0	0.0% 0	0.1% 1	28.2% 198	67.3% 446	7.3% 53	0.0% 0	0.0% 0	1.0% 7	38.7% 389	61.6% 271

Figura 5-24: Matriz de confusão da base de dados SHOAIB utilizando dados do sensor de acelerômetro com o smartphone localizado no pulso do indivíduo. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, \hat{C} = SAX, ω = 6, α = 6).

Na posição do pulso (Figura 5-24) o método HAR-SR apresenta um comportamento diferente dos demais. Ele consegue classificar muito bem as atividades estacionárias, no entanto, apresenta problemas na identificação das atividades não-estacionárias. Para o método TF, o desafio continua sendo em distinguir todas as atividades não-estacionárias. Esse resultado mostra que ambos os métodos precisam encontrar maneiras de distinguir atividades não estacionárias. Esse resultado é importante porque representa toda uma classe de dispositivos: relógios inteligentes ou *smartwatches*.

Diversos estudos são aplicados aos dados obtidos dos *smartwatches*, pois esses dispositivos são muito utilizados por indivíduos quando estão praticando esportes ou até mesmo para medir atividades biológicas, como ciclo de sono e batimentos cardíacos. Identificar corretamente as atividades não-estacionárias nessa posição merece uma atenção especial nos trabalhos futuros.

O último experimento apresentado é independente da posição, ou seja, os rótulos de posições foram removidos da base de dados e os métodos devem encontrar os diferentes tipos de padrões de cada posição. Esse trabalho adicional que os algoritmos de classificação têm para reconhecer os padrões de atividades onde não há nenhuma informação explícita sobre a posição é um cenário compatível com o que acontece no mundo real. Nesse experimento, os resultados mostram que o método HAR-SR apresenta um desempenho abaixo da média dos resultados obtidos em posições específicas. No entanto, ainda mantém uma boa taxa de classificação. O método TF apresenta um desempenho semelhante aos mostrados em todos os cenários específicos, sendo melhor no reconhecimento de atividades estacionárias com erros de classificação nas atividades não-estacionárias.

Matriz de confusão para avaliação independente da posição do smartphone

		Método Proposto						TF					
		Acurácia: 77.82%						Acurácia: 80.72%					
Classe		Descendo	Correndo	Sentando	Parado	Subindo	Andando	Descendo	Correndo	Sentando	Parado	Subindo	Andando
		Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo	Classe Alvo
Descendo		67.0% 2550	6.7% 239	0.1% 4	0.0% 0	13.8% 476	7.4% 241	68.6% 2944	2.9% 107	0.2% 8	0.1% 3	9.8% 382	6.1% 156
Correndo		5.3% 202	93.0% 3293	0.0% 0	0.0% 0	0.4% 13	0.1% 2	1.4% 60	96.4% 3530	0.0% 0	0.0% 0	0.2% 6	0.2% 4
Sentando		0.2% 7	0.0% 1	79.1% 2499	26.0% 1000	0.1% 3	0.0% 0	0.0% 2	0.0% 0	92.3% 3328	7.4% 267	0.0% 1	0.1% 2
Parado		0.1% 5	0.0% 0	20.8% 656	73.9% 2844	0.1% 5	0.0% 0	0.0% 2	0.0% 0	7.0% 254	91.4% 3308	0.4% 14	0.9% 22
Subindo		15.7% 597	0.2% 8	0.0% 1	0.1% 3	73.8% 2549	10.8% 352	13.9% 596	0.4% 16	0.1% 4	0.1% 2	63.7% 2476	19.9% 506
Andando		11.6% 443	0.0% 1	0.0% 1	0.1% 3	11.8% 409	81.7% 2653	16.0% 685	0.2% 8	0.3% 12	1.1% 39	25.9% 1007	72.8% 1849

Figura 5-25: Matriz de confusão da base de dados SHOAI B utilizando dados do sensor de acelerômetro com dados de todas as posições. Os parâmetros utilizados no método HAR-SR foram (φ = concatenação, \hat{C} = SAX, ω = 6, α = 6).

Algumas conclusões podem ser destacadas com base nas informações obtidas a partir de todas as posições. Na primeira, o método HAR-SR, no geral, tem mais dificuldade na classificação de atividades estacionárias quando comparado com o método TF. No entanto, apresenta bons resultados na classificação de atividades não-estacionárias. A exceção para ambos os casos é para a posição do pulso. Nesse caso, ambos os métodos apresentam muitos erros de classificação para as atividades não-estacionárias.

A segunda conclusão é que existe um potencial de evolução para o método HAR-SR. Analisando os pontos fracos do método é possível perceber que a principal dificuldade encontrada está na classificação das atividades estacionárias, como levantando e sentado. Esse problema é menos complexo quando comparado ao reconhecimento de padrões de atividades que envolvem algum tipo de movimentação. As direções futuras envolvem o estudo de novas técnicas de fusão de dados que agregam novas informações para discriminar sinais característicos de atividades estacionárias. Esse novo método permitirá que desempenho do reconhecimento desse grupo de atividades seja maior, tornando o método ainda mais robusto.

5.7. Cenário F1: modelo generalizado vs modelo personalizado

Para avaliar a capacidade de generalização e especialização dos trabalhos, são utilizados dois tipos de modelos de validação: o modelo generalizado e o modelo personalizado. Nos dois casos, o conjunto de dados é separado entre treino e teste. Se existem $D = \{A, B, C\}$ indivíduos, então a primeira interação de experimentos terá $(N - A)$ utilizados para o conjunto de treino, e o indivíduo restante, A , será utilizado no conjunto de teste. O processo se repete para todos os indivíduos N . Os resultados apresentados são baseados na média dos experimentos das N interações.

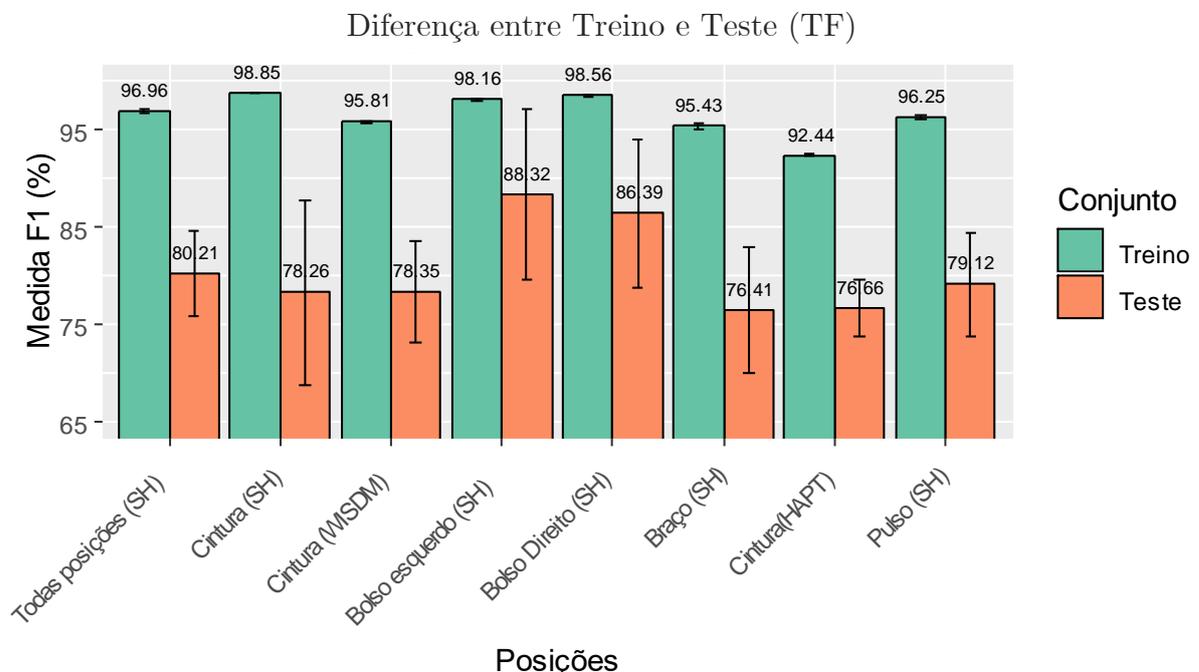


Figura 5-26: Diferença entre a média dos resultados obtidos entre o conjunto de treino (CV) e teste (LOSO) do método TF.

O modelo generalizado utiliza a validação LOSO, ou seja, não mistura dados de um mesmo indivíduo na base de treino e teste. O modelo personalizado realiza a validação

cruzada nos dados que pertencem a um mesmo indivíduo, ou seja, tanto o conjunto de treino quanto o conjunto de teste possuem dados do mesmo indivíduo.

O resultado apresentado na Figura 5-26 mostra que o modelo TF possui um excelente desempenho no conjunto de treino, que utiliza o CV. No entanto, tem seu desempenho degradado no conjunto de testes.

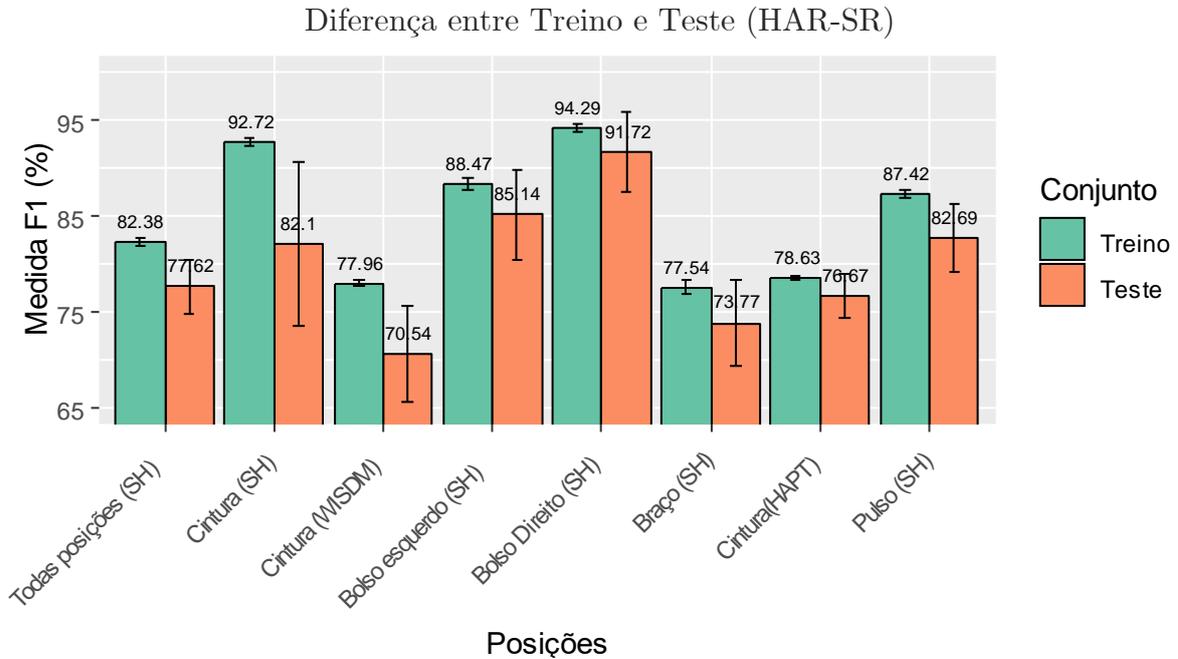


Figura 5-27: Diferença entre a média dos resultados obtidos entre o conjunto de treino (CV) e teste (LOSO) do método HAR-SR.

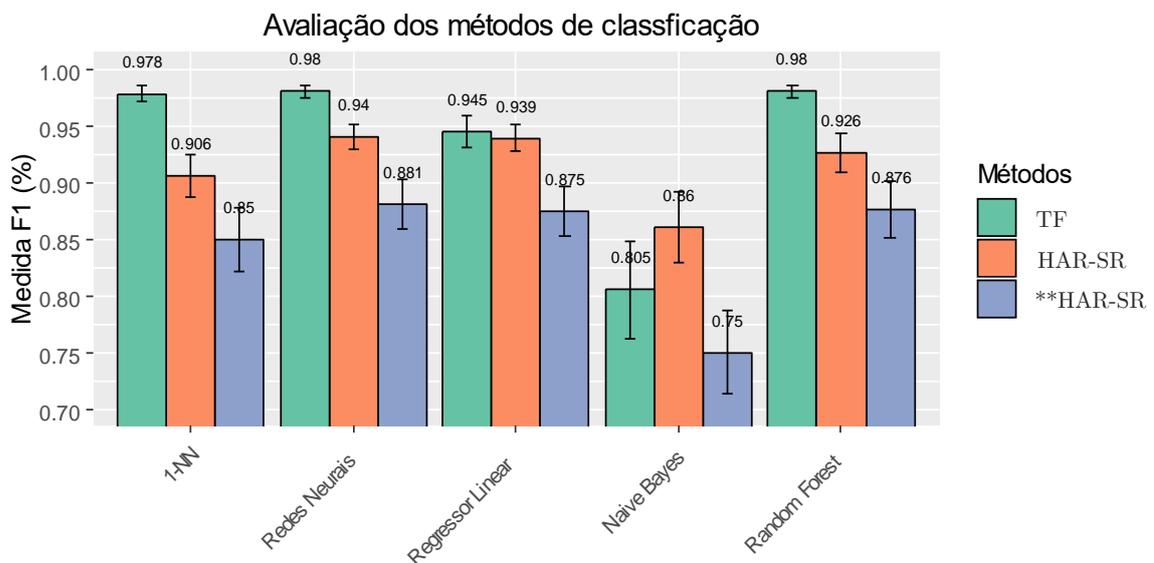


Figura 5-28: Diferença entre a média dos resultados do modelo personalizado utilizando validação cruzada tradicional para diferentes classificadores. O método **HAR-SR utiliza histogramas de referência do mesmo usuário.

A Figura 5-27 mostra os resultados de desempenho para o método HAR-SR. Note que o método HAR-SR apresenta seus resultados entre os modelos de treino e teste muito próximos. Isso significa que o modelo de classificação obtido na etapa de treinamento possui melhor capacidade de generalização, quando comparado com o modelo TF. O modelo TF apresenta margens maiores entre o conjunto de treino e teste. Os valores muito altos na fase de treinamento indicam que esse é um modelo adequado para especialização em um indivíduo, no entanto, se novos indivíduos são inseridos no domínio, o modelo terá dificuldades em reconhecê-los.

O resultado do desempenho de cinco classificadores é apresentado na Figura 5-28. Nesse experimento, foi medido a capacidade de especialização do método TF e HAR-SR. Os dados individuais de cada usuário foram submetidos à validação cruzada e os resultados da média obtida por cada classificador é apresentada.

No modelo personalizado, o método HAR-SR utiliza tanto a validação cruzada com dados de indivíduos diferentes quanto utilizando dados do mesmo indivíduo (**HAR-SR) na geração dos histogramas de referência. O modelo TF apresenta os melhores resultados para o modelo personalizado.

5.8. Considerações do capítulo

O capítulo 5 apresentou os experimentos e os resultados obtidos da avaliação do HAR-SR para reconhecer atividades humanas. Foram utilizados seis cenários de avaliação para observar o comportamento de diferentes etapas do processo de reconhecer atividades humanas e para selecionar o melhor conjunto de parâmetros. Além disso, foram realizadas comparações com métodos encontrados na literatura, o TF, que é baseado em características extraídas manualmente dos sinais no domínio do tempo e frequência, o BOSS-VS e o SAX-VSM, ambos pertencentes ao domínio discreto e que empregam técnicas de *feature learning*. Os resultados mostram que o HAR-SR apresenta uma média de desempenho similar ao método TF, porém utiliza uma quantidade menor de características no modelo de classificação. O método HAR-SR apresentou resultados superiores aos métodos BOSS-VS e SAX-VSM em todos os cenários.

O próximo capítulo apresenta uma nova forma de analisar atividades humanas sob um aspecto visual. O plano Entropia-Complexidade utiliza as medidas estatísticas do método HAR-SR para representar atividades em um plano 2D. Esse plano permite analisar diferentes tipos de comportamentos e interpretá-los sob um aspecto físico, no qual o determinismo e o estocástico podem ser distinguidos.

Análise de atividades utilizando o plano Entropia-Complexidade

Este capítulo apresenta mais detalhes sobre o reconhecimento de atividades a partir do plano de complexidade. O plano entropia-complexidade é obtido através das medidas estatísticas. No contexto RAH, esse é o primeiro estudo a utilizar estes conceitos, com um diferencial: as medidas de complexidade estatísticas foram adaptadas para problemas de classificação. Isso permite que novos valores de complexidade possam ser avaliados. Os resultados apresentados mostram que esses novos valores são indispensáveis para o bom desempenho do método de classificação.

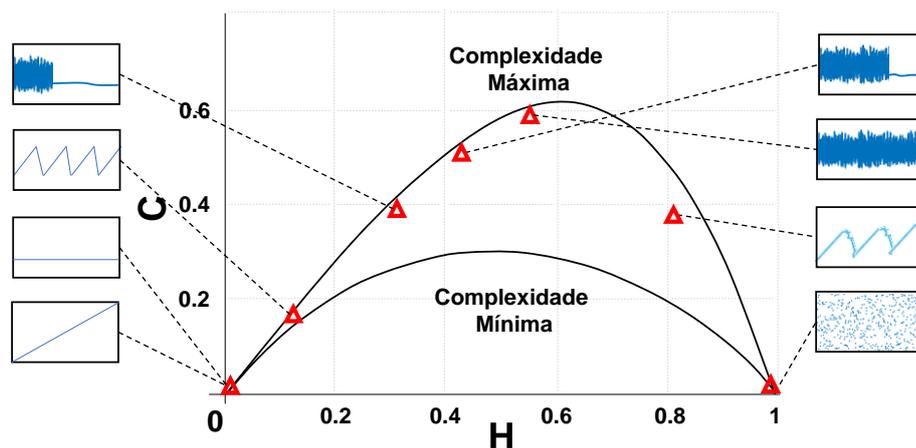


Figura 6-1: A figura mostra diversos exemplos de séries temporais e seu comportamento no plano Entropia-Complexidade.

A medida de complexidade estatística pode ser representada em um plano Entropia-Complexidade (HxC). Esse plano funciona como uma ferramenta de visualização e análise de séries temporais de qualquer tamanho no plano 2D, onde os quantificadores

estatísticos podem distinguir as séries temporais de natureza determinística ou estocástica.

A posição de cada ponto possui uma interpretação física, conforme mostra a Figura 6-1 e Figura 6-2.

- i. Os pontos localizados no canto inferior direito correspondem a uma série temporal aleatória, com comportamento estocástico. Quanto mais à esquerda está o ponto, mais determinístico é seu comportamento.
- ii. Os pontos com valores maiores de complexidade representam as séries muito divergentes da distribuição uniforme.
- iii. A linha contínua consiste na região em que os pares (H, C) estarão presentes. Elas indicam o limite inferior C_{\min} e superior C_{\max} considerando todas as possíveis combinações de histogramas e divergências (Martin, et al., 2006).
- iv. Histogramas uniformes diferentes geram medidas de divergências diferentes. Isso implica em mudanças na posição dos pares (H, C) no plano.
- v. O comportamento dos limites limite inferior C_{\min} e superior C_{\max} muda conforme o histograma de referência uniforme muda.

Neste trabalho, o número de classes utilizados para a construção do plano é referente ao número de atividades que devem ser reconhecidas. No total são seis atividades, conforme destacadas na Figura 6-2. Definido o número de classes do problema novos histogramas de referências podem ser obtidos, um para cada classe.

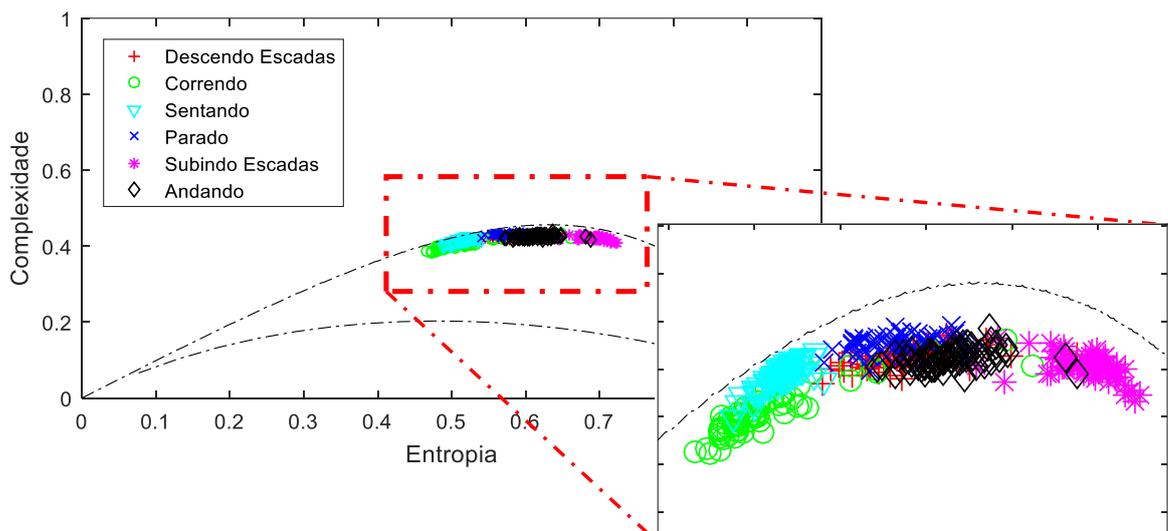


Figura 6-2: Exemplo do plano Entropia-Complexidade de um indivíduo para seis atividades.

Além do histograma uniforme Pe utilizado no cálculo da divergência, são utilizados também os histogramas de referências para cada uma das classes de atividades

existentes. Assim, além das características de entropia, divergência e complexidade geradas a partir do P_e , para cada nova classe existirá uma característica adicional de divergência e complexidade (Tabela 6-1).

Tabela 6-1: Tabela de cada histograma de referência utilizado.

Símbolo	Classe
P_e	Histograma de referência uniforme.
P_D	Histograma de referência para a atividade descendo escadas
P_R	Histograma de referência para a atividade correndo
P_S	Histograma de referência para a atividade sentando
P_P	Histograma de referência para a atividade levantando
P_U	Histograma de referência para a atividade subindo escadas
P_W	Histograma de referência para a atividade andando

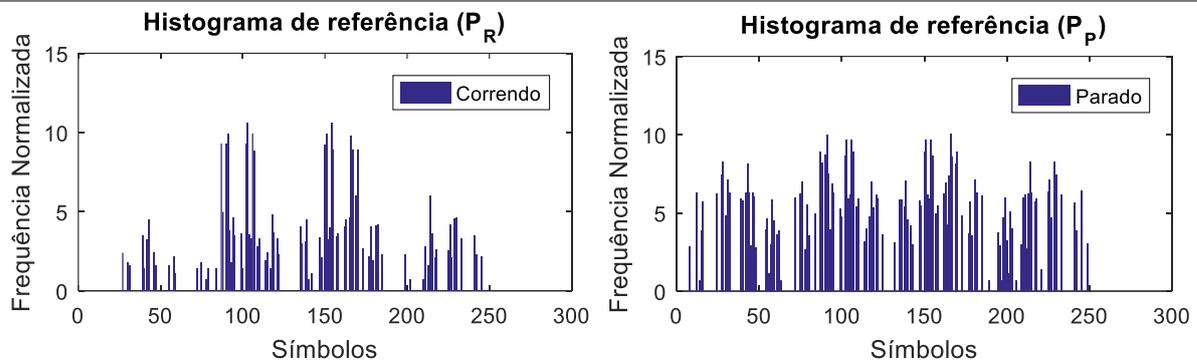


Figura 6-3. histograma de referência P_R e P_P para a classe de atividade ‘correr’ e ‘parado’ respectivamente, em um espaço de 256 símbolos.

A Figura 6-3 mostra um exemplo do histograma de referência P_R e P_P para a classe de atividade ‘correr’ e ‘parado’ respectivamente, em um espaço de 256 símbolos. Note que os histogramas de cada classe de atividades possuem diferenças que podem ser observadas pela distribuição do histograma. Essa diferença será observada também nos valores obtidos pela medida de divergência e de complexidade. A utilização de diferentes tipos de histogramas de referência permite caracterizar melhor o sinal de interesse.

As novas medidas podem ser utilizadas também em um plano Entropia – C_i , , onde C_i são as medidas de complexidade que utilizam os histogramas de referência de cada atividade, conforme mostra a Figura 6-4.

Analisando a Figura 6-4 é possível notar que os valores da medida de complexidade diminuem em relação a uma classe, por exemplo, correndo (P_R), se o histograma de referência utilizado for em relação a mesma classe (Figura 6-5). Essas diferenças entre os histogramas são fundamentais para que os métodos de aprendizagem consigam encontrar padrões e gerar modelos de classificação de atividades com um bom desempenho.

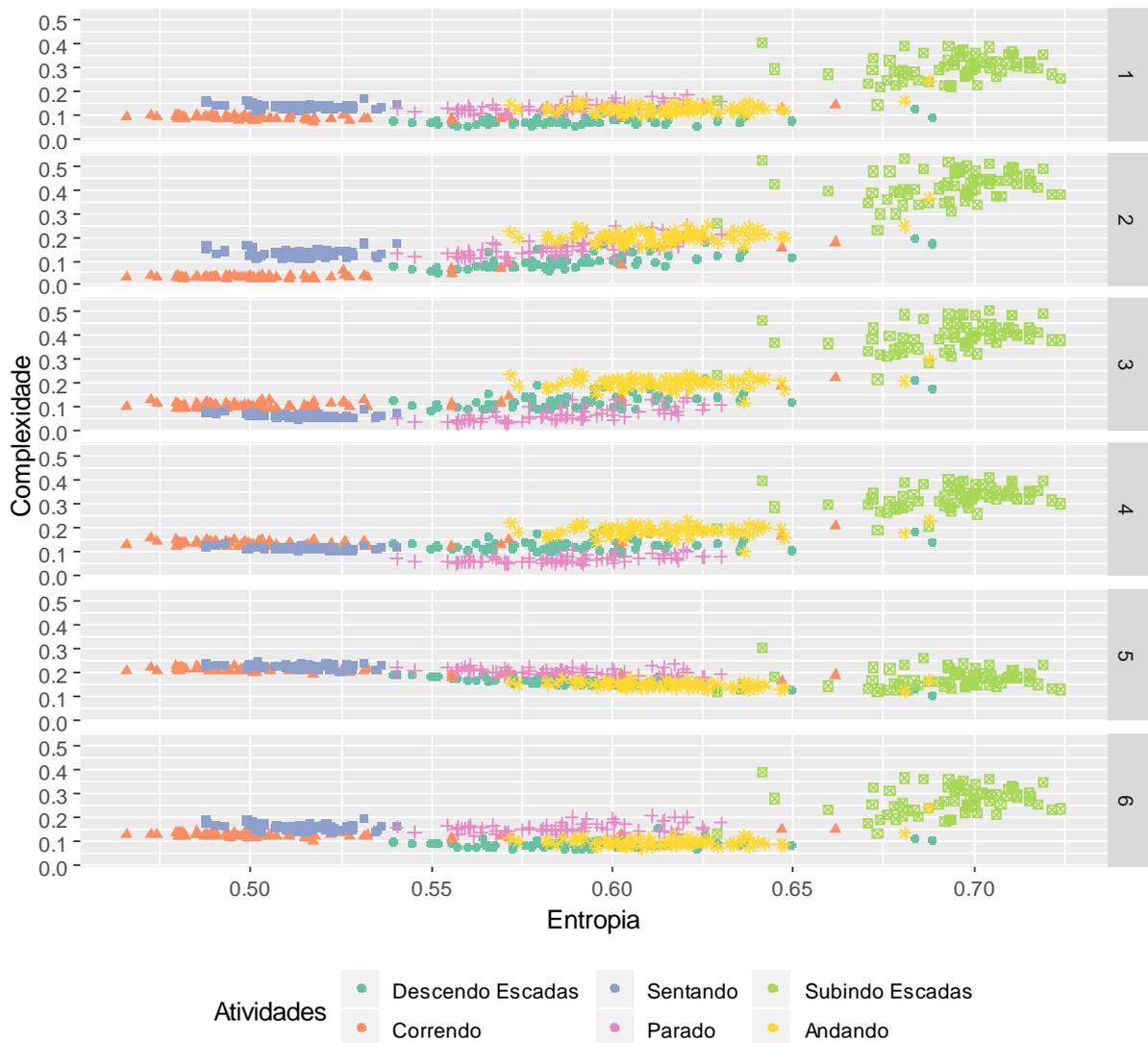


Figura 6-4: Plano Entropia – Complexidade_i, onde C_i são as medidas de complexidade que utilizam os histogramas de referência de cada atividade. Os valores são 1 = P_D, 2 = P_R, 3 = P_S, 4 = P_P, 5 = P_U, 6 = P_W.

Separação de classes entre os diferentes histogramas de referência

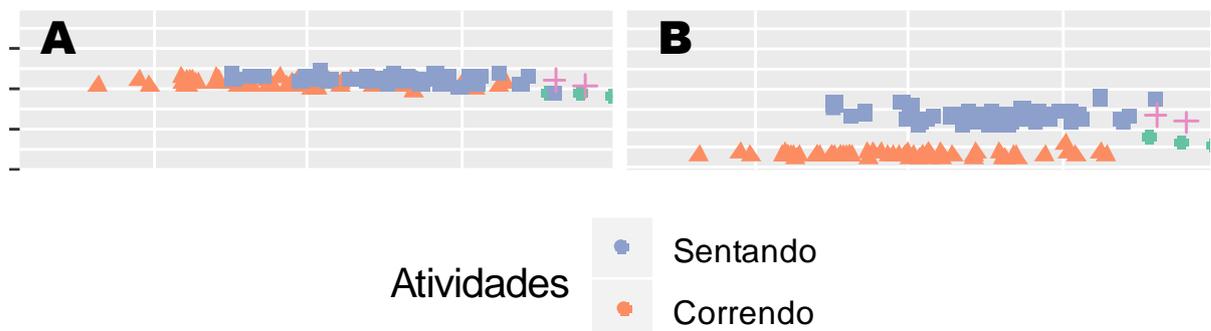


Figura 6-5: Diferenças entre as atividades “correndo” e “sentando” geradas no Plano Entropia-Complexidade para diferentes histogramas de referência.

Uma das principais vantagens da utilização do plano Entropia-Complexidade está na possibilidade de comparação dos indivíduos e seus comportamentos através de uma

representação visual. A Figura 6-6 mostra a comparação entre as atividades executadas por dois indivíduos utilizando o plano Entropia – Complexidade_i ($i = P_R$). A representação mostra que existem diferenças na execução das atividades entre os dois indivíduos. Por exemplo, comparando os indivíduos n.9 e n.15, é possível visualizar padrões diferentes para a atividade andar e correr.

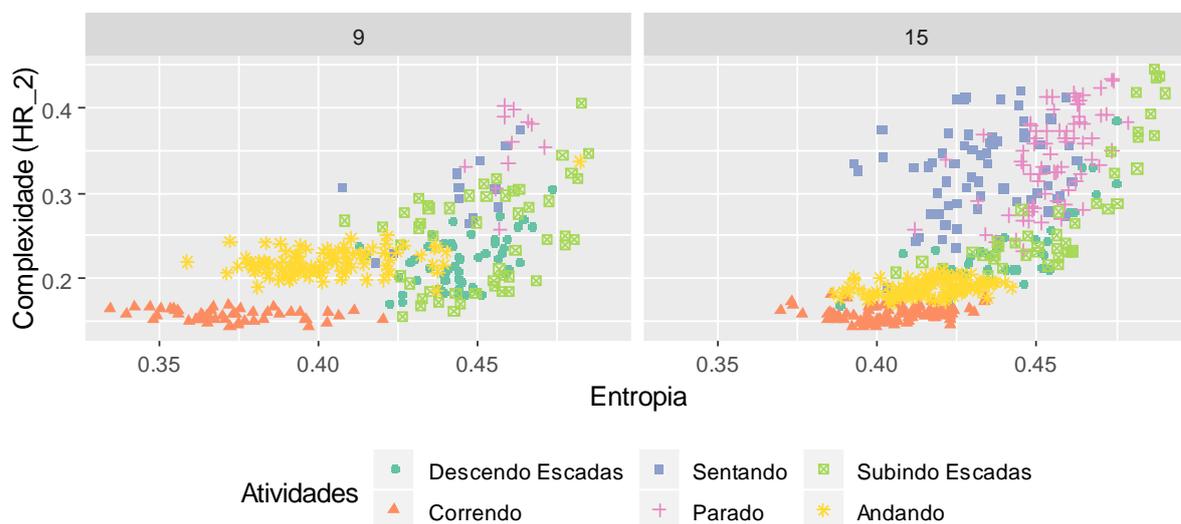


Figura 6-6: Comparação entre dois indivíduos através do plano Entropia – Complexidade_i, onde i utiliza o histograma P_R .

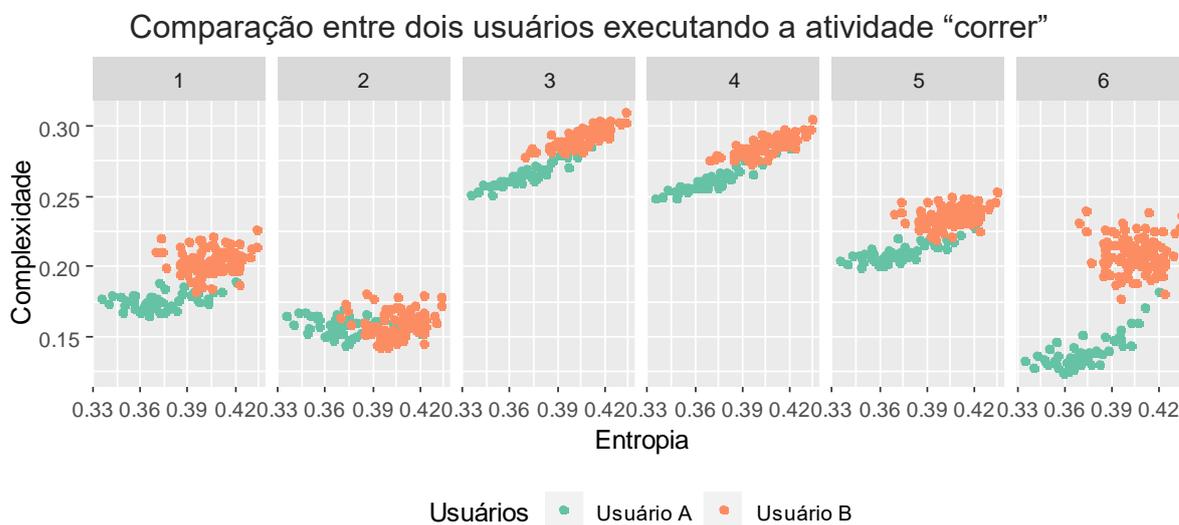


Figura 6-7: Comparação entre dois indivíduos que estão executando a atividade ‘correr’. Histogramas utilizados: 1 = P_D , 2 = P_R , 3 = P_S , 4 = P_P , 5 = P_U , 6 = P_W .

É possível realizar comparações entre atividades específicas, conforme mostra a Figura 6-7. Nesse exemplo, dois indivíduos estão executando a mesma atividade (correr). As diferentes medidas de complexidade (1-6) permitem que diferentes representações possam ser analisadas para determinar o quão diferente são essas atividades, dada a medida de complexidade para cada histograma.

O gráfico de densidade é uma forma de visualização que permite que distribuições possam ser comparadas e analisadas. O ponto mais alto do gráfico mostra onde os valores estão mais concentrados. Esse gráfico pode ser utilizado para analisar as diferenças entre os indivíduos, conforme mostra a Figura 6-8.

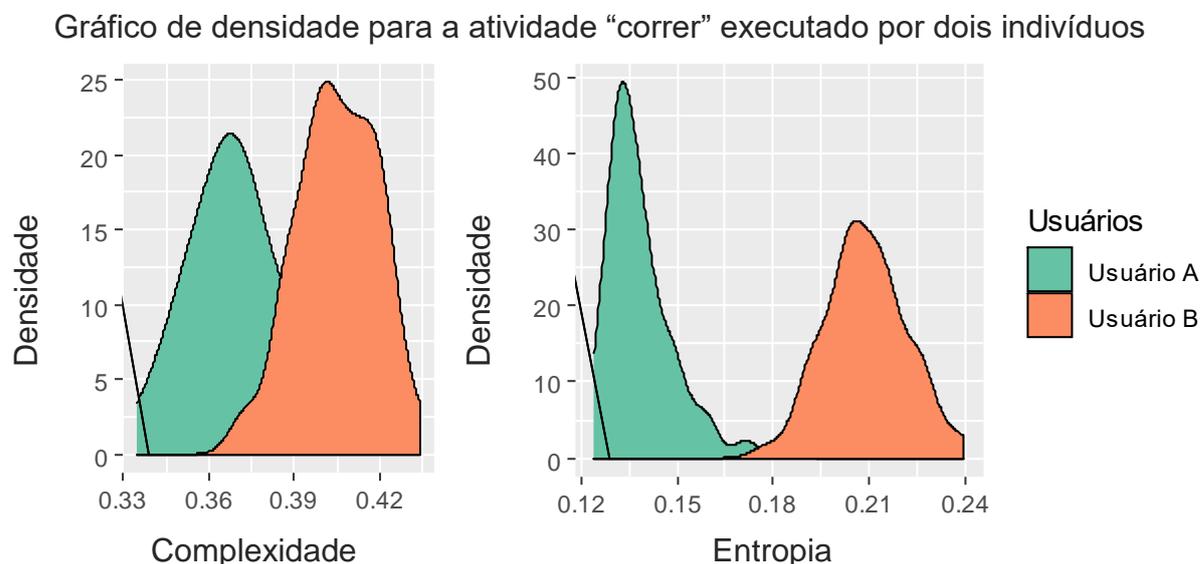


Figura 6-8: O gráfico de densidade é utilizado para comparar as diferenças entre dois indivíduos. No exemplo, dois indivíduos executam a atividade “correr”.

Analisar essas informações pode fornecer diversas informações importantes. O histograma de referência utilizado no cálculo das medidas de divergência e complexidade são representações de um grupo de indivíduo, ou seja, representam a média da população encontrada na base de dados. Dessa forma, é possível descobrir comportamentos que se distanciam muito da média do grupo da qual o indivíduo pertence.

Outra informação pode ser obtida através da análise dos dados de um mesmo indivíduo ao longo do tempo, no qual deve ser possível apontar diferenças ou anomalias em seu comportamento de acordo com as mudanças encontradas no plano Entropia-Complexidade. Essas informações podem ser úteis na criação de sistemas que permitam que seus usuários saibam quando estão se desviando de hábitos saudáveis.

Outra importante aplicação do plano Entropia-Complexidade é que a representação visual gerada pode ajudar na escolha de diversos parâmetros utilizados pelo método HAR-SR. Por exemplo, é possível destacar as diferenças entre os sensores selecionados para a construção do método, as diferenças entre os métodos de fusão de dados, entre os algoritmos de representação simbólica e entre as classes do problema em estudo.

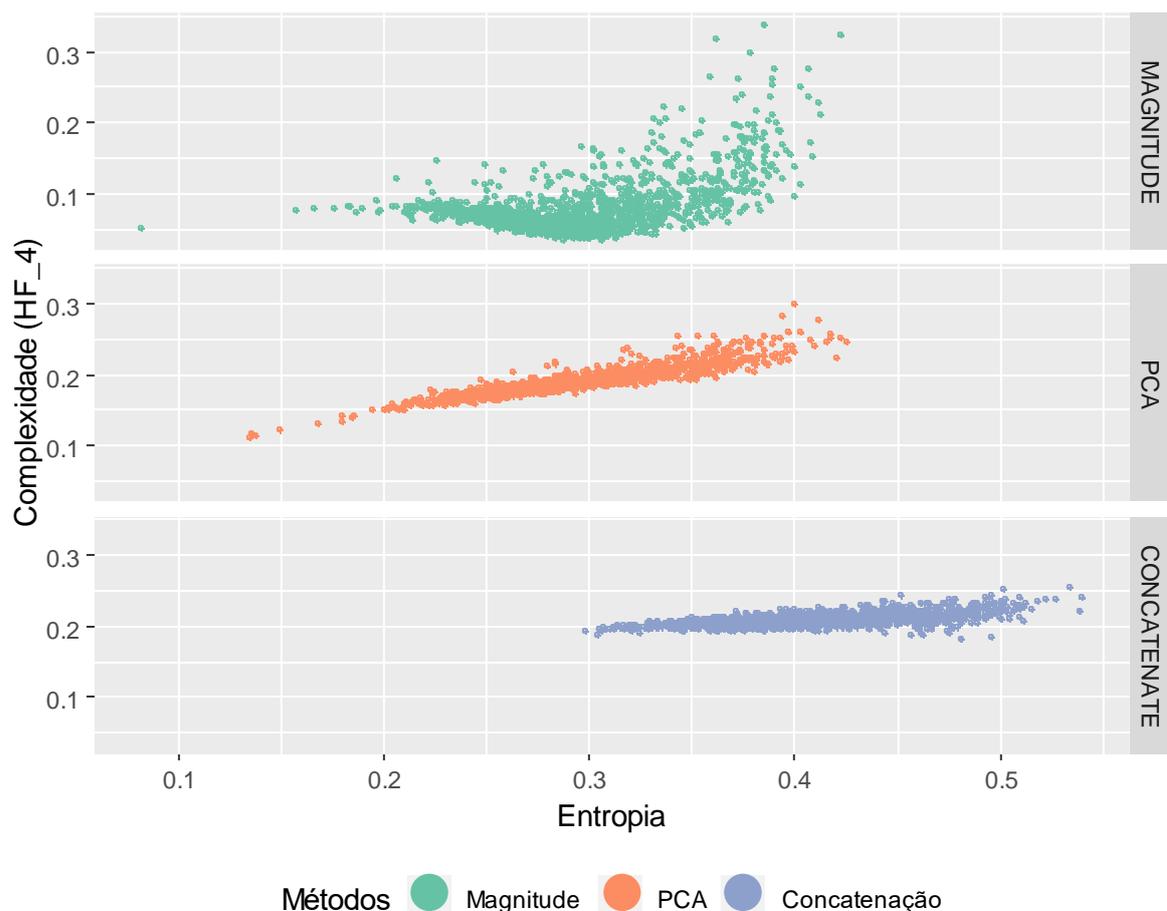


Figura 6-9: Plano Entropia-Complexidade para os métodos de fusão de dados Magnitude, PCA e concatenação simples.

A interpretação física dos valores de entropia e complexidade pode ser utilizada para mostrar quais são os melhores parâmetros. Por exemplo, a Figura 6-9 mostra o plano Entropia-Complexidade para os métodos de fusão de dados Magnitude, PCA e concatenação simples. Como a entropia mínima aponta o total determinismo de uma distribuição, os valores mais baixos de entropia apresentados pelos métodos de Magnitude e PCA na Figura 6-9 revelam que esses sinais possuem padrões mais homogêneos se comparados com a concatenação simples.

A Figura 6-10 mostra o plano de Entropia-Complexidade da base de dados WISDM para a atividade “sentado” com diferentes tamanhos de palavra e alfabeto para os algoritmos de representação simbólica SAX e SFA.

A interpretação visual mostra que valores maiores nos parâmetros de palavra e alfabeto adotados diminuem a entropia da série e aumentam sua complexidade. É um indicativo de que padrões distintos no sinal são encontrados. A implicação dessa interpretação aplicada aos algoritmos de representação simbólica é que esses métodos devem

conseguir encontrar palavras que representam melhor o conjunto de atividades e assim representá-las com mais definição.

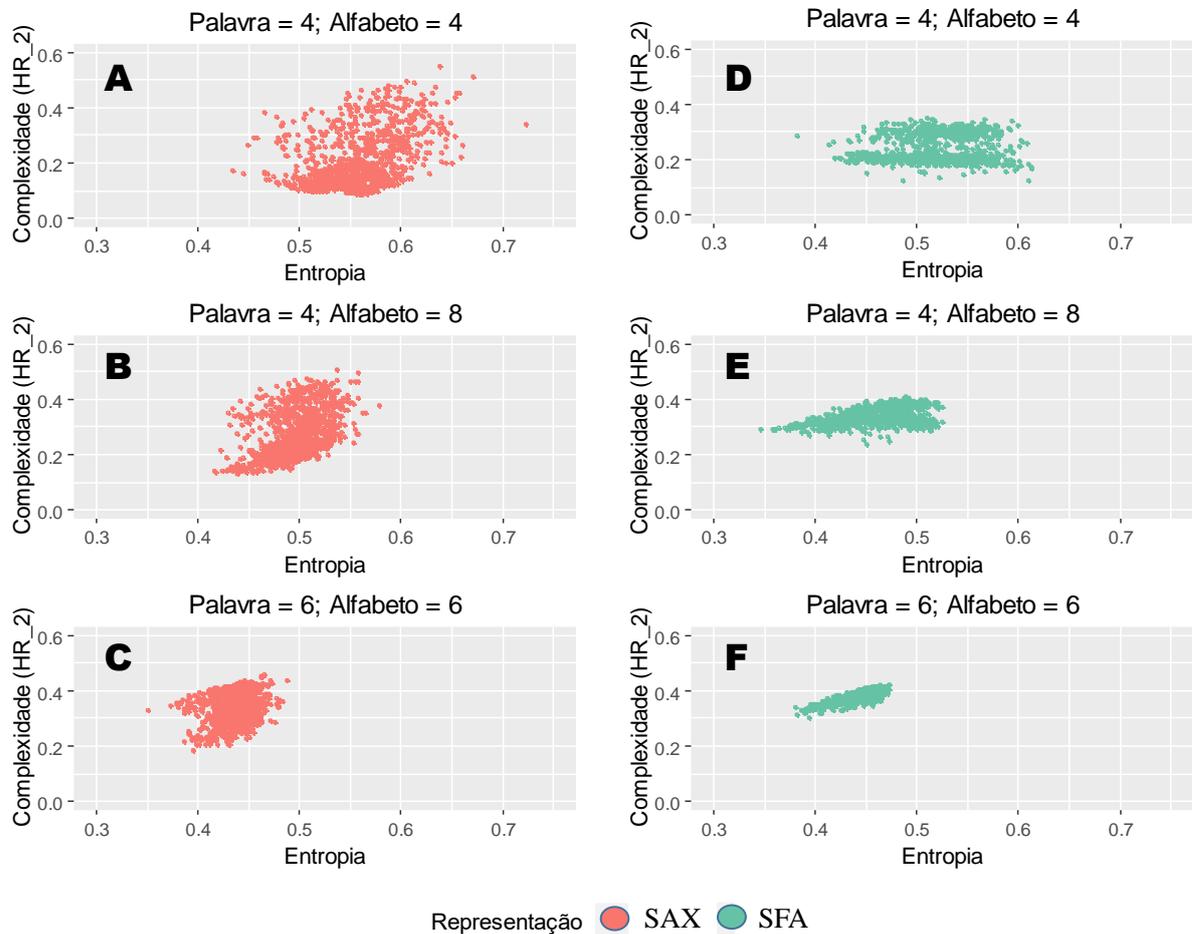


Figura 6-10: Plano de complexidade da atividade “sentando” da base de dados WISDM com os algoritmos de representação simbólica SAX e SFA, variando-se os parâmetros de tamanho de palavra e tamanho do alfabeto.

Pela definição, quanto maior a entropia, maior a desordem de um sistema. Nesse sentido, conforme mostra a Figura 6-10, a variação dos parâmetros e a obtenção de valores de entropia mais baixos estão associados a uma melhor representação dos dados. Os valores da complexidade mais altos, entre 0.3 e 0.4 para os parâmetros maiores, também representam melhorias na qualidade da representação dos sinais no domínio simbólico.

6.1. Considerações do capítulo

O capítulo 6 mostrou uma nova forma de analisar atividades humanas sob um aspecto visual. A utilização das medidas estatísticas de entropia, divergência e complexidade pelo método HAR-SR permitem que atividades sejam representadas em um plano 2D. Esse plano permite analisar diferentes tipos de comportamentos e interpretá-los sob um aspecto físico, no qual o determinismo e o estocástico podem ser distinguidos.

O plano Entropia-Complexidade mostrou que indivíduos executam atividades de formas diferentes, e essa diferença pode ser interpretada como normal de acordo com a média dos indivíduos, dado pelo seu respectivo histograma de referência, ou pode ser vista como uma anomalia. Essa percepção pode direcionar futuras pesquisas para o desenvolvimento de aplicações para pessoas que buscam melhorar seu desempenho em atividades físicas, por exemplo.

Além disso, o plano Entropia-Complexidade é uma boa forma de visualizar o comportamento e a influência da escolha de cada parâmetro do método HAR-SR. A seleção dos melhores parâmetros de acordo com a interpretação dos valores mostrados no plano pode resultar em um modelo de classificação com uma melhor capacidade de generalização no reconhecimento de atividades humanas.

Conclusões e trabalhos futuros

O reconhecimento de atividades humanas é uma área emergente que vem sendo o tema de estudo de muitas pesquisas. A extração automática de características é uma das partes mais desafiadoras, pois influencia no desempenho dos algoritmos de classificação. As pesquisas recentes têm sido direcionadas aos métodos que utilizam redes neurais profundas, que apresentam bons resultados, mas possuem custo computacional elevado para dispositivos móveis. Até o momento, a revisão da literatura mostrou que são poucos os trabalhos que realizam o reconhecimento de atividades no domínio discreto dos sinais. Essa lacuna de poucos trabalhos motivou o desenvolvimento desta pesquisa, que encontrou bons resultados.

7.1. Considerações sobre a extração automática de características

A primeira contribuição desta pesquisa foi a criação de um modelo de reconhecimento de atividades que aprende a extrair características úteis de novas atividades. Esse procedimento é realizado automaticamente porque as informações de cada nova atividade são automaticamente introduzidas no cálculo das características utilizadas pelo método. Através de um processo de aproximação e discretização nos sinais de entrada, o método HAR-SR consegue extrair características discriminantes dos sinais dos sensores automaticamente. Além disso, as características são representadas em um espaço muito menor que os dados de entrada, resultando também em um modelo de classificação mais simples. Os resultados mostram que mesmo com poucas

características o método HAR-SR apresenta um desempenho similar ou superior a muitos trabalhos considerados estado da arte na literatura.

Na introdução deste trabalho foram levantados alguns dos problemas encontrados nas abordagens manuais. O primeiro é o problema da variabilidade intraclasse, no qual uma mesma atividade pode ser executada de formas diferentes por indivíduos diferentes ou um mesmo indivíduo também executar uma mesma atividade de forma diferente. O segundo problema levantado foi a similaridade interclasse, que acontece quando existem classes que são diferentes, mas que possuem características similares nos dados dos sensores.

O problema da variabilidade intraclasse é evidente quando analisado sob o plano Entropia-Complexidade. Esse problema está associado com a capacidade de generalização do modelo. Os resultados obtidos mostram que o método HAR-SR é genérico o suficiente para representar bem grupos normais de indivíduos. No entanto, o plano Entropia-Complexidade mostra que os indivíduos que executam atividades de forma diferente das demais pessoas acabam afetando o desempenho dos algoritmos de classificação. Por exemplo, a atividade andar de um determinado indivíduo pode ser semelhante à atividade correr de outro indivíduo. A diferença entre esses comportamentos pode gerar confusões no modelo de classificação.

A similaridade interclasse é um dos desafios do RAH. Os resultados mostram que é possível criar duas subcategorias para esse tipo de problema. A primeira subcategoria é a similaridade interclasse de atividades estacionárias, como parado, sentado, deitado e a segunda categoria é a similaridade interclasse de atividades não-estacionárias, como correr, andar, subir e descer escadas. Os resultados da matriz de confusão da seção 5.6 mostram o método HAR-SR consegue distinguir bem as categorias de sinais estacionários e não estacionários. Apesar de seu bom desempenho, na identificação das atividades dentro dessa categoria de sinais estacionários, encontra dificuldades em distinguir atividades, como sentar e levantar.

De forma diferente, a abordagem de extração manual de características TF possui dificuldades em classificar as atividades não-estacionárias. Esse resultado mostrou um grande avanço na solução do problema de similaridade interclasse, porque identificar atividades não-estacionárias é uma tarefa mais complexa. Uma das apostas futuras para melhorar o desempenho na classificação de atividades estacionárias está na seleção de um método de fusão de dados que discrimine melhor as características desses sinais e que também possa preservar características globais dos sinais que foram perdidas no processo de normalização aplicado pelos métodos de representação simbólica.

7.2. Considerações sobre a seleção de sensores e fusão de dados

Um problema comum na literatura é que os algoritmos de aprendizagem de máquina tendem a perder desempenho conforme a dimensionalidade de dados aumenta e em contrapartida o custo computacional é maior. Por esse motivo, essa pesquisa avalia três formas diferentes de redução de dimensionalidade do espaço de características utilizadas pelo método HAR-SR. Essa redução possui impacto direto no custo computacional do método (detalhes no Apêndice A).

A primeira forma de reduzir o custo computacional está relacionada com a seleção apropriada dos sensores para reconhecer as atividades, a segunda está relacionada com a aplicação de técnicas de agregação de dados e, por fim, a terceira se refere a um processo de discretização utilizado pelos algoritmos SAX e SFA.

A primeira forma de reduzir a dimensionalidade dos dados busca diminuir a quantidade de dados de entrada através da seleção dos sensores de movimento, sem que haja perda de desempenho do método. Conforme apresentado no Capítulo 5, algumas atividades apresentam os mesmos resultados quando um novo sensor é adicionado e em alguns casos a diferença é muito pequena, o que inviabiliza a utilização do sensor. A escolha errada de um sensor pode adicionar processamento desnecessário ao método, aumentando sua complexidade, e pode resultar em perda de desempenho na classificação. Uma descoberta importante encontrada no cenário C1 é que a posição onde o smartphone está localizado tem relação com quantidade de sensores necessárias para um bom desempenho do método de classificação. Essa análise é importante para que em pesquisas futuras a seleção dos sensores seja realizada conforme a posição do smartphone no indivíduo.

A segunda forma de redução de dimensionalidade foi obtida através dos métodos de Magnitude, PCA e a concatenação simples. Esses métodos possuem a tarefa de transformar os sinais dos sensores em uma forma unidimensional, visto que os dados coletados dos sensores de movimento estão em um plano multidimensional. Esse processo é fundamental para reduzir a complexidade global do algoritmo, além de viabilizar o uso dos métodos de representação simbólica, que só aceitam séries temporais unidimensionais.

A primeira conclusão do estudo envolvendo métodos de fusão de dados é que a qualidade das características geradas pelos métodos de representação simbólica depende da qualidade dos sinais obtidos no processo de fusão de dados. A segunda conclusão com base nos resultados mostra que tanto a Magnitude quanto o PCA são insuficientes

para gerar um sinal representativo para as atividades desse estudo. Isso resultou em um baixo desempenho do algoritmo de classificação nos testes, que em sua maioria não ultrapassa os 60% na medida F1. Uma das explicações possíveis é que tanto a Magnitude quanto o PCA não agregam boas características do sinal e há muita perda de informações durante o processo de fusão de dados.

Por outro lado, o método de concatenação simples apresenta os melhores resultados em todos os cenários, ainda que não seja um método de fusão de dados propriamente dito. A mudança no espaço realizada pelo método de concatenação faz com que as séries temporais multidimensionais sejam convertidas em um vetor de características que é apropriado para o método de discretização. A implicação disso é que o método de discretização terá um segmento de dados maior contendo todos os dados coletados. Esses resultados dos métodos de fusão de dados mostram que os métodos de agregação utilizados nesta pesquisa perdem muita informação acerca dos sinais, e por isso os métodos simbólicos não conseguem extrair representações de qualidade. Por isso, novas investigações ainda podem ser realizadas utilizando métodos mais robustos de fusão de dados, como o filtro de Kalman, por exemplo. O objetivo dessa nova investigação é encontrar métodos mais representativos e que usem um vetor de características menor, se comparados com a concatenação simples, por exemplo.

A terceira forma de fusão de dados é diferente da apresentada pelos métodos de Magnitude e PCA que são aplicadas diretamente nos sinais dos sensores. Sua nova forma é intermediária, a nível de características, que é obtida pelos algoritmos SAX e SFA, cujos resultados são discutidos na próxima seção.

7.3. Considerações sobre os métodos de representação simbólica

A terceira forma de redução de dimensionalidade é intrínseca dos métodos de representação simbólica. Esse processo transforma os dados representados por valores reais em um domínio discreto. Isso permite que diversos métodos de mineração de dados possam ser utilizados, como as características estatísticas utilizadas nesta pesquisa, ou o modelo *tf-idf* utilizado pelos métodos SAX-VSM e BOSS-VS. Os resultados apresentados por esses algoritmos têm apontado novas soluções em diversas áreas no contexto de séries temporais.

A comparação entre os métodos de discretização SAX e SFA deixa indícios de que a utilização de uma tabela de discretização supervisionada (MCB), utilizada pelo algoritmo SFA, pode não ser uma boa escolha para gerar modelos generalistas. Os

melhores resultados obtidos pelo SAX podem ser justificados com base na escolha de sua tabela de discretização, cujos valores são calculados a partir de uma distribuição gaussiana. Por esse motivo, a capacidade de generalização do modelo tende a ser maior que em um modelo que tome com informações somente dos conhecimentos de uma base limitada de dados.

Uma consequência dessa descoberta é que os modelos gerados a partir do SAX são menos complexos se comparados aos do SFA. Isso acontece porque, primeiro, o SAX utiliza uma tabela de pesquisa fixa e pré-definida, como mencionado anteriormente. Segundo, o processo de aproximação realizado pelo PAA, que utiliza a média, é mais simples se comparado ao processo de aproximação do SFA, que utiliza a decomposição de sinais através da transformada de Fourier. Por último, o SFA possui uma etapa extra de treinamento e geração da tabela de pesquisa, chamada MCB.

7.4. Considerações sobre o HAR-SR

Para avaliar o desempenho do método HAR-SR no reconhecimento de atividades foi utilizada a validação cruzada por indivíduo (LOSO), de modo que não fosse possível que o conjunto de treinamento e de testes estivessem contaminados por padrões de um mesmo indivíduo. A utilização do LOSO é importante porque é o indicador do potencial de generalização da aplicação, sem que haja a necessidade de obtenção de dados do indivíduo antes da utilização do sistema.

A utilização dessa metodologia de avaliação mostrou resultados interessantes. Os primeiros resultados mostram uma queda no desempenho de classificação dos trabalhos apresentados na literatura, que até então mostravam altos níveis de acurácia (acima de 90%). No problema RAH, a utilização da validação cruzada tradicional pode não ser uma boa forma de representar o mundo real e, por consequência, os modelos obtidos podem não refletir a realidade.

Por outro lado, os altos níveis de acurácia utilizando a validação tradicional podem ser um bom indicativo de que o modelo de classificação consegue extrair bem os padrões dos dados conhecidos pelo modelo. Isso possibilita a criação de modelos ideais para se utilizar em cenários de classificação personalizados por indivíduo.

Os resultados do método HAR-SR na comparação com os métodos SAX-VSM e BOSS-VS mostraram que o método HAR-SR é superior com uma ampla margem no domínio de RAH. Esse resultado mostra que as características obtidas pelo modelo *tf – idf* são menos representativas se comparadas com as medidas estatísticas utilizadas pelo HAR-SR.

Na comparação do método de características manuais (TF) com o método HAR-SR os resultados são similares. No entanto, analisando alguns aspectos individuais é possível notar diferenças significativas entre os dois métodos. Por exemplo, o HAR-SR apresenta um erro maior na classificação entre as atividades estacionárias, como levantar e sentar. Já o método TF apresenta taxas de erro maiores para o grupo de atividades não-estacionárias, como andar, correr, subir e descer escadas. A exceção para ambos os casos é para a posição do pulso onde, nesse caso, ambos os métodos apresentam as maiores taxas de erro de classificação para as atividades não-estacionárias.

Um ponto positivo para o método HAR-SR é que reconhecer as atividades não-estacionárias é mais complexo, tarefa no qual o método apresentou bons resultados. O método ainda consegue reconhecer com boa taxa de acerto as atividades estacionárias, mas erra na classificação de atividades dentro desse grupo. As informações obtidas a partir da matriz de confusão deixam evidências de que existe um potencial de evolução para o método HAR-SR, já que o método é bom na tarefa de reconhecer as atividades mais complexas, como andar, correr, subir escadas e descer escadas.

Conhecer o funcionamento e as especializações de cada método pode direcionar pesquisas futuras para melhorar a classificação de atividades. Por exemplo, para solucionar o problema de classificação entre as atividades estacionárias, novas técnicas de fusão de dados que agregam novas informações para discriminar sinais característicos de atividades estacionárias podem ser empregadas no processo. Esse novo método permitirá que desempenho do reconhecimento desse grupo de atividades seja maior, tornando o método ainda mais robusto.

Ao se realizar uma comparação com os métodos TF, BOP-SAX e BOSS, o método HAR-SR gera um número significativamente menor de características. Esse conjunto de características é obtido em função do número de classes do problema com base nos valores das medidas estatísticas. Isso pode ser considerado uma vantagem, principalmente porque o modelo de classificação construído é mais simples.

Um dos pontos negativos do método HAR-SR está na seleção de parâmetros ideais para os algoritmos de representação simbólica. A escolha dos valores incorretos para o tamanho de janela, palavra e alfabeto podem resultar numa perda de representação dos dados, cujo impacto é refletido diretamente no desempenho de classificação das atividades. Esse é um problema que o HAR-SR herda dos métodos SAX e SFA, que também é um problema compartilhado por outros métodos de natureza simbólica, como por exemplo, o SAX-VSM, o BOSS-VS. As DNNs, quando analisadas sob o aspecto, possuem diversos parâmetros para controlar sua arquitetura, como por exemplo, tamanho, profundidade e número de filtros, que precisam ser cuidadosamente

selecionados. Além disso, os algoritmos que possuem altas acurácias podem não ser ideais para utilização em um cenário real se forem computacionalmente caros, ou consumirem demasiado tráfego de dados e também possuírem altas latências.

7.5. Considerações sobre o plano Entropia-Complexidade

A introdução do plano de Entropia-Complexidade no contexto RAH utilizando a nova metodologia de extração de características proposta nesta pesquisa é uma inovação em muitos aspectos.

Uma das principais dificuldades da área de RAH consiste em gerar resultados interpretáveis por um ser humano, já que os modelos de classificação propostos na literatura possuem camadas e processos que não são interpretáveis. As abordagens baseadas na extração manual de características, por exemplo, apresentam variáveis estatísticas que são facilmente interpretáveis quando analisadas individualmente, mas são difíceis quando precisam ser observadas em conjunto, principalmente, se o conjunto de característica for grande.

Em outro exemplo está o modelo *tf – idf* utilizado pelos algoritmos SAX-VSM e BOSS-VS, que é extremamente difícil de ser interpretado se o espaço de amostras for relativamente grande. Espaços pequenos podem ser interpretados, mas não são suficientes para extrair padrões significativos para se obter um bom desempenho no processo de classificação.

As pesquisas relacionadas às redes neurais profundas, por exemplo, têm seus métodos utilizados como uma caixa preta, sem a possibilidade de explicar o porquê dos bons resultados ou sem a possibilidade de modificação do método para resolver problemas específicos de classificação. Quando estas redes são aplicadas no contexto RAH, também são utilizadas como caixa-preta e não há uma preocupação com a explicação das variáveis e suas mudanças. Por outro lado, todo o processo descrito pelo HAR-SR, desde a segmentação até a criação do modelo de classificação é interpretável, através da análise das medidas estatísticas no plano Entropia-Complexidade. A introdução das medidas estatísticas resolve esse problema que até então era um grande desafio no RAH.

O plano Entropia-Complexidade permite que os comportamentos associados a cada atividade sejam visualmente interpretados. Os resultados apresentados no capítulo 6 mostraram que é possível, por exemplo, distinguir usuários com padrões de comportamento diferentes ou encontrar padrões de atividades similares. Essa facilidade de interpretação do modelo permitirá que novos estudos sejam direcionados a geração

de modelos de grupos específicos de usuários, por exemplo, na descoberta de anomalias no comportamento por causa de desvios nos padrões de atividades.

7.6. Dificuldades encontradas

As principais dificuldades desta pesquisa estão associadas à aquisição de bases de dados e à metodologia de validação e comparação dos trabalhos encontrados na literatura. A seguir serão apresentados mais detalhes sobre estes problemas.

- i. Pesquisas relacionadas ao RAH ainda possuem um número de rótulos de atividades limitadas e faltam atividades mais complexas. A dificuldade de criação de bases de dados grandes está principalmente no processo de rotulagem de atividades, que são realizados de forma manual através da análise dos dados brutos. Esse tipo de rotulagem, em geral, requer indivíduos especialistas, consome muito tempo e uma grande quantidade de participantes pode tornar o processo complexo. Eliminar essa limitação é importante porque não é possível medir a escalabilidade dos métodos propostos se os conjuntos de dados é limitado a apenas poucas atividades ou até mesmo a poucas instâncias. Dessa maneira, medir o desempenho de um método conforme a complexidade das atividades ou conforme cresce o número de atividades é um item desejável, mas indisponível.
- ii. Existem poucas bases de dados públicas com quantidade de dados suficientemente grande. Esse aspecto também é um fator de limitação na apuração dos resultados e testes dos métodos literatura. Muitos dessas bases possuem poucas instâncias de atividades e muitas delas são geradas em laboratórios. Assim, modelos gerados com dados que fogem do cotidiano podem sofrer uma degradação acentuada no seu desempenho quando submetidos aos testes do mundo real.

7.7. Direções futuras

O método HAR-SR se mostrou promissor, não só no contexto de RAH, mas no cenário de classificação de séries temporais como um todo. Para continuar com a evolução do método, existem algumas direções principais a serem seguidas:

- Métodos eficientes para realizar a fusão dos dados dos sensores. O capítulo 5 de resultados mostrou que uma das principais causas de erros de classificação do método HAR-SR são as atividades estacionárias, como levantar e sentar. Por isso, novos estudos devem ser direcionados em busca de novas técnicas de fusão de dados que agreguem novas informações para discriminar sinais característicos de atividades estacionárias, permitindo que o erro de classificação de atividades seja reduzido. Esse processo pode tornar o método ainda mais robusto.
- Necessidade de um estudo que selecione os sensores dinamicamente, conforme a posição e atividade que está sendo executada (Figura 7-1). Como visto, alguns sensores são melhores para reconhecer atividades em determinada posição que outros sensores. O resultado desse estudo pode levar a construção de métodos com maiores taxas de acerto.

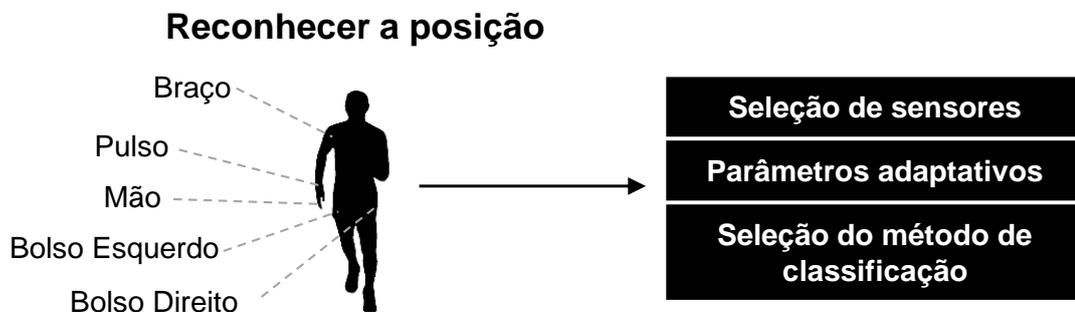


Figura 7-1: O método de classificação pode reconhecer a posição do smartphone e a partir disso, selecionar os sensores apropriados para o reconhecimento de atividades.

- Metodologia de classificação independente. Essa nova abordagem extrai características de cada sinal do sensor de forma independente. Dessa forma, cada um dos eixos x, y, z possuirá seu próprio conjunto de características obtidos através das medidas estatísticas, conforme apresentado na Figura 7-2.

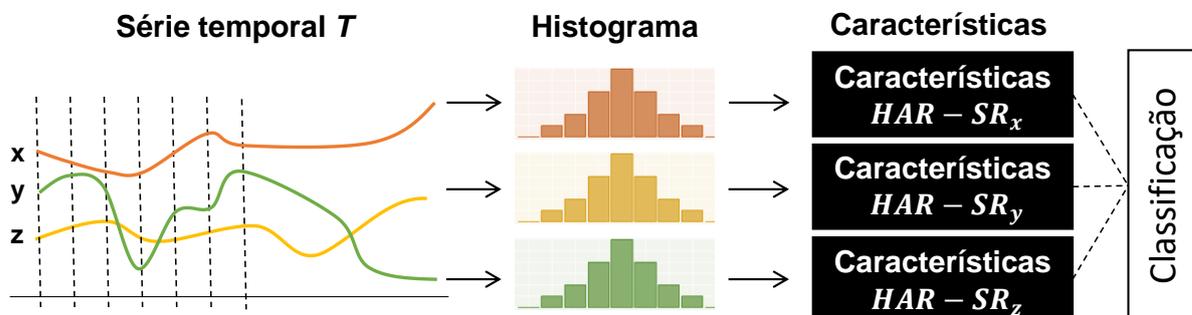


Figura 7-2: Uma nova metodologia pode realizar a extração de características a partir de cada sinal do sensor para gerar o conjunto de treinamento.

- Segurança e privacidade. As atividades executadas por indivíduos podem significar uma assinatura de seu comportamento. Os resultados mostraram que

existem diferenças entre a execução de atividades pelos indivíduos. Nesse sentido, uma nova proposta pode ser relacionada a detecção e autenticação de um indivíduo baseada na sua assinatura comportamental.

Referências

- Anguita, D. et al., 2013. A Public Domain Dataset for Human Activity Recognition Using Smartphones. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 24-26.
- Antonelli, M., De Micco, L., Larrondo, H. & Rosso, O. A., 2018. Complexity of Simple, Switched and Skipped Chaotic Maps in Finite Precision. *Entropy*, Volume 20, pp. 1-24.
- Bagnall, A. et al., 2016. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, pp. 1-55.
- Bagnall, A., Lines, J., Hills, J. & Bostrom, A., 2015. Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles. *IEEE Transactions on Knowledge and Data Engineering*, Volume 27, pp. 2522-2535.
- Bandt, C. & Pompe, B., 2002. Permutation entropy - a complexity measure for time series. *Physical Review Letters*, Volume 88, p. 174102.
- Banos, O. et al., 2014. Window size impact in human activity recognition.. *Sensors (Basel, Switzerland)*, Volume 14, pp. 6474-99.
- Banovic, N. et al., 2016. Modeling and Understanding Human Routine Behavior. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pp. 248-260.
- Bao, L. & Intille, S. S., 2004. Activity Recognition from User-Annotated Acceleration Data. *Pervasive Computing*, pp. 1-17.
- Barber, D., 2012. *Bayesian reasoning and machine learning*. s.l.:Cambridge University Press.
- Bersch, S. D. et al., 2014. Sensor data acquisition and processing parameters for human activity classification. *Sensors (Basel, Switzerland)*, Volume 14, pp. 4239-4270.
- Bulling, A., Blanke, U. & Schiele, B., 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, Volume 1, pp. 1-33.

- Butler, M. & Kazakov, D., 2015. SAX Discretization Does Not Guarantee Equiprobable Symbols. *IEEE Transactions on Knowledge and Data Engineering*, Volume 27, pp. 1162-1166.
- Cichosz, P., 2014. *Data Mining Algorithms: Explained Using R*. s.l.:John Wiley & Sons.
- Colonna, J. G., Cristo, M. A. P., Rosso, O. A. & Nakamura, E. F., 2014. Permutation Entropy Applied to Bioacoustic Signal Segmentation. *Conference on Nonequilibrium Statistical Mechanics and Nonlinear Physics*, Volume XVIII, pp. 16-17.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C., 2009. *Introduction to Algorithms, Third Edition*. 3rd ed. s.l.:The MIT Press.
- Dernbach, S. et al., 2012. Simple and Complex Activity Recognition through Smart Phones. *2012 8th International Conference on Intelligent Environments (IE)*, pp. 214-221.
- Duda, R. O., Hart, P. E. & Stork, D. G., 2000. *Pattern Classification (2Nd Edition)*. s.l.:Wiley-Interscience.
- Figo, D., Diniz, P. C., Ferreira, D. R. & Cardoso, J. M. P., 2010. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, Volume 14, pp. 645-662.
- Gravina, R., Alinia, P., Ghasemzadeh, H. & Fortino, G., 2017. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, Volume 35, pp. 1339-1351.
- Gu, Y. et al., 2016. We know what you are doing or going to do: Towards accurate human activities sensing. *2016 25th International Conference on Computer Communications and Networks, ICCCN 2016*.
- Gu, Y., Liu, W. & Song, J., 2015. Can Activities of Human Daily Life be Recognized and Predicted?. *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 127-132.
- Harari, G. M. et al., 2016. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, Volume 11, pp. 838-854.
- He, K. & Sun, J., 2014. Convolutional Neural Networks at Constrained Time Cost. *CoRR*, Volume abs/1412.1.

- Ignatov, A., 2018. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing Journal*, Volume 62, pp. 915-922.
- Incel, O. D., Kose, M. & Ersoy, C., 2013. A Review and Taxonomy of Activity Recognition on Mobile Phones. *BioNanoScience*, Volume 3, pp. 145-171.
- Khan, W. Z., Xiang, Y., Aalsalem, M. Y. & Arshad, Q., 2013. Mobile Phone Sensing Systems: A Survey. *IEEE Communications Surveys & Tutorials*, Volume 15, pp. 402-427.
- Khusainov, R., Azzi, D., Achumba, I. E. & Bersch, S. D., 2013. Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations.. 13(10), pp. 12852-12902.
- Kolosnjaji, B. & Eckert, C., 2015. *Neural Network-Based User-Independent Physical Activity Recognition for Mobile Devices*. s.l.:s.n.
- Kwapisz, J. R., Weiss, G. M. & Moore, S. a., 2011. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, Volume 12, p. 74.
- Lara, O. D. & Labrador, M. a., 2013. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys & Tutorials*, Volume 15, pp. 1192-1209.
- Larsen, R. J., Marx, M. L. & others, 1986. *An introduction to mathematical statistics and its applications*. s.l.:Prentice-Hall Englewood Cliffs, NJ.
- Li, F. et al., 2018. Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors. *Sensors*, Volume 18, p. 679.
- Lima, W. S., Bragança, H. L. S., Quispe, K. G. M. & Souto, J. P., 2018. Human Activity Recognition based on Symbolic Representation Algorithms for Inertial Sensors. *Sensors*, Volume 18, pp. 1-26.
- Lin, J., Keogh, E., Lonardi, S. & Chiu, B., 2003. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. *SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pp. 2-11.
- Lin, J., Keogh, E., Wei, L. & Lonardi, S., 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Min Knowl Disc*, Volume 15, pp. 107-144.

- Lin, J., Khade, R. & Li, Y., 2012. Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems*, Volume 39, pp. 287-315.
- Manning, C. D., Raghavan, P. & Schütze, H., 2008. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press.
- Martin, M. T., Plastino, A. & Rosso, O. A., 2006. Generalized statistical complexity measures : Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, Volume 369, pp. 439-462.
- Micucci, D., Mobilio, M. & Napoletano, P., 2017. UniMiB SHAR: a new dataset for human activity recognition using acceleration data from smartphones.
- Mikolov, T., Chen, K., Corrado, G. & Dean, J., 2013. Distributed Representations of Words and Phrases and their Compositionality. *Nips*, pp. 1-9.
- Miller, G., 2012. The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*, Volume 7, pp. 221-237.
- Morales, J. & Akopian, D., 2017. Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomedical Engineering*, Volume 37, pp. 388-400.
- Murphy, K. P., 2012. *Machine Learning: A Probabilistic Perspective*.
- Nakamura, E. F., Loureiro, A. a. F. & Frery, A. C., 2007. Information fusion for wireless sensor networks. *ACM Computing Surveys*, Volume 39, pp. 9----es.
- Nweke, H. F., Teh, Y. W., Al-garadi, M. A. & Alo, U. R., 2018. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, Volume 105, pp. 233-261.
- Nweke, H., Wah, T., Mujtaba, G. & Ali Al-garadi, M., 2018. Data Fusion and Multiple Classifier Systems for Human Activity Detection and Health Monitoring: Review and Open Research Directions. *Information Fusion*, 6. Volume 46.
- Pires, I. M., Garcia, N. M., Pombo, N. & Fl??rez-Revuelta, F., 2016. From data acquisition to data fusion: A comprehensive review and a roadmap for the identification of activities of daily living using mobile devices. *Sensors (Switzerland)*, Volume 16.
- Raví, D. et al., 2017. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, Volume 21, pp. 4-21.

- Ravi, D., Wong, C., Lo, B. & Yang, G.-Z., 2016. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 71-76.
- Riedl, M. & Wessel, N., 2013. Practical considerations of permutation entropy. *The European Physical Journal Special Topics*, Volume 262, pp. 249-262.
- Ronao, C. A. & Cho, S.-b., 2016. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, Volume 59, pp. 235-244.
- Rosso, O. A. et al., 2010. Generalized Statistical Complexity Measure. *International Journal of Bifurcation and Chaos*, Volume 20, p. 775.
- Rosso, O. A. et al., 2007. Distinguishing noise from chaos. *Physical Review Letters*, Volume 99, pp. 1-4.
- Schäfer, P., 2015. The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, Volume 29, pp. 1505-1530.
- Schäfer, P., 2016. Scalable time series classification. *Data Mining and Knowledge Discovery*, Volume 30, pp. 1273-1298.
- Schäfer, P. & Höggqvist, M., 2012. SFA: a symbolic fourier approximation and index for similarity search in high dimensional datasets. *Data Mining and Knowledge Discovery*, Volume 29, pp. 228-242.
- Senin, P. & Malinchik, S., 2013. SAX-VSM: Interpretable time series classification using sax and vector space model. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 1175-1180.
- Shannon, C. E. & Weaver, W., 1963. *A Mathematical Theory of Communication*. s.l.:University of Illinois Press.
- Shirahama, K. & Grzegorzec, M., 2017. On the Generality of Codebook Approach for Sensor-Based Human Activity Recognition. *Electronics*, Volume 6, pp. 1-18.
- Shoaib, M. et al., 2014. *Fusion of smartphone motion sensors for physical activity recognition*. s.l.:s.n.
- Shoaib, M. et al., 2016. Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors (Switzerland)*, Volume 16, pp. 1-24.

- Shoaib, M. et al., 2015. A Survey of Online Activity Recognition Using Mobile Phones. *Sensors*, Volume 15, pp. 2059-2085.
- Shoaib, M., Scholten, H. & Havinga, P. J. M., 2013. *Towards Physical Activity Recognition Using Smartphone Sensors*. s.l., s.n., pp. 80-87.
- Siirtola, P. et al., 2011. Improving the classification accuracy of streaming data using SAX similarity features. *Pattern Recognition Letters*, Volume 32, pp. 1659-1668.
- Sun, L. et al., 2010. *Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations*. s.l.:s.n.
- Sze, V., Chen, Y. H., Yang, T. J. & Emer, J. S., 2017. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, Volume 105, pp. 2295-2329.
- Wang, J. et al., 2018. Deep learning for sensor-based activity recognition: A Survey. *Pattern Recognition Letters*.
- Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J., 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. 4 ed. Burlington(MA): Morgan Kaufmann.
- Wu, W. et al., 2012. Classification accuracies of physical activities using smartphone motion sensors. *Journal of Medical Internet Research*, Volume 14.

Apêndice A – Discussão de complexidade

Ao desenvolver um algoritmo umas das questões que devem ser respondidas é sobre o quão eficiente um algoritmo é ao executar determinada tarefa. Isso nem sempre é fácil, mas existem algumas abordagens que ajudam a responder essa pergunta.

Uma abordagem muito utilizada é baseada em medições empíricas dos resultados dos algoritmos. Essa é a forma mais simples de se medir a eficiência de um algoritmo. Em resumo, basta medir quanto tempo uma implementação leva para terminar sua execução, dada uma entrada de dados, e então comparar o resultado obtido com outros algoritmos, dada a mesma entrada de dados.

Um dos problemas dessa abordagem é que não há garantias de que o algoritmo executará rapidamente em entradas de dados relativamente maiores. O segundo problema está relacionado aos detalhes de implementação. Utilizar métricas que são dependentes de dispositivos ou arquiteturas de processadores específicos, que podem ser mascaradas por características específicas dessas plataformas, ou até mesmo depende das capacidades de algum compilador pode não ser uma boa ideia (Figo, 2010).

A solução para problemas mais complexos está na busca de uma forma mais universal para determinar a eficiência de um algoritmo. Essa nova forma deve analisar os aspectos de execução do algoritmo sem que haja a necessidade de executar o algoritmo e ainda assim pode dizer como o algoritmo irá se comportar, não importando o tamanho dos dados de entrada. Essa métrica mede a complexidade do algoritmo baseando-se, por exemplo, no número de repetições das funções ou quantidade de operações em um cálculo. Essa abordagem utiliza a notação O para medir o tempo de execução de um algoritmo para o pior caso, de acordo com uma entrada de dados (Cormen, et al., 2009).

Complexidade computacional das abordagens HCFM

Começando pelas abordagens HCFM, (FIGO, 2010) (Shoib, et al., 2014) apresentam uma análise da complexidade computacional das principais características utilizadas por essas abordagens no domínio do tempo, frequência e discreto. Para ajudar a distinguir o custo entre as diversas características analisadas, o autor atribui quatro níveis complexidade a cada característica: muito baixo, baixo, médio e alto.

Os níveis *muito baixo* e *baixo* requerem um número de operações que é linear em n , cujas operações consistem em soma, subtração, multiplicação e divisão. O nível médio inclui as métricas que são quadráticas para entradas de tamanho n . E no nível *alto* estão as técnicas que requerem um número de operações maior que o limite quadrático.

A maior parte das características analisadas possui custo computacional muito baixo ou baixo, no entanto algumas características possuem custo no nível médio e a única característica categorizada como nível alto é a entropia, que pode ser otimizada através da utilização de uma tabela hash, sendo rebaixada ao nível médio. Dessa forma, a conclusão de seu trabalho mostra que a maior parte das características podem ser obtidas em tempo linear, para uma dada entrada de tamanho n . No entanto, vamos adotar o pior cenário de características que são utilizadas para avaliar o custo computacional: o nível *médio* com custo $O(n^2)$.

Caso a técnica de janela deslizante de tamanho w seja utilizada na série T_n , a complexidade de uma subsequência será de $O(w^2)$, e para todas as janelas $n - w + 1$ terá complexidade computacional?

$$\begin{aligned} O(nw^2 - w^3 + w^2) \\ = O(nw^2). \end{aligned}$$

Assumindo que a variável c representa o número de características utilizado pelas abordagens HCFM e que N seja o total de instâncias do conjunto D , é possível obter a complexidade computacional através da seguinte equação:

$$O(N \cdot c \cdot n \cdot w^2)$$

Onde tipicamente $n \ll N$, e o maior valor de c encontrado é apresentado por (Anguita, 2013), cujo $c = 563$. Nesse cálculo, a quantidade de sensores utilizados não é considerada.

Análise de complexidade do método HAR-SR

O modelo HAR-SR é baseado no modelo BOP-SAX (Lin, 2012) e BOSS (schaeffer, 2012). Esses modelos extraem subestruturas como características de alto nível de uma série temporal. As subestruturas são representadas por um histograma para cada instância do conjunto.

O que todos esses métodos possuem em comum é o método de transformação simbólica de uma série, que podem ser discretizadas tanto pelo algoritmo SAX quanto pelo algoritmo SFA. Esses métodos utilizam técnicas diferentes, que impactam tanto na

robustez do modelo de classificação em termos de acurácia, quando no tempo de execução do algoritmo. A próxima seção apresenta detalhes da complexidade computacional do SAX e do SFA.

Complexidade do SAX

O algoritmo SAX é baseado em cálculos de média e por isso sua implementação é muito simples. Possui apenas duas etapas: a aproximação e a discretização. O processo de aproximação utiliza o algoritmo PAA para dividir a série T em ω partes. O processo de quantização consiste numa pesquisa na tabela de quantização β , que é fixa e pré-definida para todos os domínios. Por esse motivo, a complexidade do SAX é dominada pela complexidade do algoritmo PAA, que é linear em n (Lin, et al., 2003) (Schäfer & Höggqvist, 2012)(Senin & Malinchik, 2013):

$$\text{SAX}(T_n) = O(n).$$

Complexidade do SFA

O SFA é um algoritmo mais robusto e por isso possui uma fase adicional: a fase de treinamento, que tem por objetivo obter uma tabela de discretização β . Essa tabela, chamada de MCB é dinâmica e utiliza o conjunto de dados disponíveis para ser criada. Somente após a fase de treinamento é que uma série temporal T pode ser discretizada.

A fase de treinamento consiste no processo de aproximação de todas as séries temporais do conjunto D e a partir desse resultado, permite a criação da tabela MCB. O processo de aproximação visa transformar uma série T_n de tamanho n em uma nova série T_m , cujo tamanho é m , onde $m \ll n$. O processo de aproximação do SFA utiliza a Transformada Discreta de Fourier (DFT) e a Transformada Momentânea de Fourier (MFT).

O processo de aproximação do SFA é dominado pela complexidade computacional de uma única DFT para uma série T_n , resultando na complexidade $O(n \log n)$. Caso a técnica de janela deslizante de tamanho w seja utilizada na série T_n , a complexidade de uma subsequência, utilizando a DFT, será de $O(w \log w)$, e para todas as janelas $n - w + 1$ terá complexidade computacional $O(nw \log w)$. No entanto, otimizações no algoritmo SFA permitem que a MFT seja utilizada para lidar com o problema da sobreposição significativa dos cálculos das janelas w . Dessa forma, a DFT é aplicada somente na primeira janela ($O(w \log w)$) e todo o restante das janelas ($n - w$) são

calculadas através da MFT ($O(\omega)$). Dessa forma, a nova complexidade computacional do método é reduzida para:

$$\text{MFT}(T_{sw}) = O(\omega(n - w) + w \log w)$$

Para o conjunto D, a complexidade computacional é $O(N(\omega n + w \log w))$. A aplicação da MFT em todo o conjunto finaliza o processo de aproximação. Na sequência, o processo de quantização consiste na utilização da tabela MCB como tabela de pesquisa, que contém intervalos de quantização. Primeiro, o conjunto D é utilizado, representado por uma matriz contendo todas as N transformações de Fourier de tamanho ω . Então, dada essa matriz, as colunas são ordenadas e particionadas em intervalos equidistantes. A ordenação das ω colunas com N valores possui complexidade $O(\omega N \log N)$, utilizando o merge sort, por exemplo. Os intervalos equidistantes aplicados às colunas ordenadas requerem uma busca em cada coluna, adicionando uma complexidade de $O(\omega N)$. Dessa forma, a complexidade total do processo de aproximação e criação da tabela MCB é definido:

$$\begin{aligned} \text{MCB (FT)} &\in \text{ordenação (TF}_n) + \text{quantização(FT}_n) + \text{SFA(FT}_n) \\ &\in O(\omega N \log N) + O(\omega N) + O(Nn \log n) \end{aligned}$$

Utilizando a DFT:

$$\begin{aligned} \text{MCB (FT}_n) &= O(\omega N \log N + \omega N + Nn \log n) \\ &= O(N (\omega \log N + \omega + n \log n)) \end{aligned}$$

Utilizando a MFT:

$$\begin{aligned} \text{MCB (FT}_{sw}) &= O(\omega N \log N + \omega N + Nn + w \log w) \\ &= O(N(\omega \log N + \omega + n) + w \log w) \end{aligned}$$

O resultado da complexidade computacional mostra que a etapa de discretização (cálculo da tabela MCB) da fase de pré-processamento é dominada pela transformação de Fourier de tamanho N e pela ordenação para se obter os intervalos equidistantes.

A fase de transformação do SFA, que acontece depois da fase de pré-processamento, consiste discretização de uma série T utilizando a tabela MCB. Então, para transformar uma série T usando a MFT, uma palavra SFA de tamanho ω requer ω buscas na tabela MCB que contém α intervalos. Isso resulta em $O(\omega \log \alpha)$ operações (utilizando pesquisa binária). Assim, a complexidade total no processo de transformação do SFA para uma serie T, dada a tabela MCB é: $\text{SFA}(T) \in \text{SFA}(T) + \text{pesquisa}(T)$.

Transformação SFA	
Método	Complexidade
DFT_T	$SFA(T_n) = O(n \log n + \omega \log \alpha)$
DFT_{FT}	$SFA(FT_n) = O(N(n \log n + \omega \log \alpha))$
MFT_T	$SFA(T_{sw}) = O(\omega + w \log w + \omega \log \alpha)$
MFT_{FT}	$SFA(FT_{sw}) = O(N(\omega n + w \log w + \omega \log \alpha))$

Assim, o processo de transformação de SFA é dominado pela transformação de Fourier, pelo impacto da palavra de tamanho l , pelo tamanho da janela w e pelo alfabeto c .

Complexidade da extração de subestruturas de uma série

O HAR-SR funciona de forma semelhante ao modelo BOP e ao modelo BOSS. Esses modelos (*structure-based*) extraem um vetor de características de alto nível para construir um modelo de classificação a partir de algoritmos de aprendizagem, como KNN, árvore de decisão e SVM.

O processo de extração de subestruturas é simples. Aplica-se a técnica de janela deslizante em uma série temporal T extrair w subsequências. Em seguida, cada subsequência é discretizada através do SAX ou SFA, cujas complexidades já são conhecidas. Por isso, a complexidade do HAR-SR para extrair subestruturas depende do método de discretização utilizado. O algoritmo 1 apresenta um método para extrair subestruturas de uma série temporal T que utiliza o método SAX.

Algoritmo 1: *buildClassifierBOP* (lista de n casos de tamanho m , $T = (X, y)$)

Parâmetros: tamanho da palavra ω , tamanho do alfabeto α , e tamanho da janela w

```

1: seja  $\mathbf{P}$  a lista de  $n$  histogramas  $(p_1, p_2, \dots, p_n)$ 
2: for  $i = 1: n$ 
3:   for  $j=1: m - w + 1$ 
4:      $q = x_{i,j}, \dots, x_{i,j+w-1}$ 
5:      $r = SAX(q, \omega, \alpha)$ 
6:     if  $(r \neq p)$ 
7:       //a função index determina a localização da palavra  $r$  na matriz  $h_j$ 
8:        $pos = \mathbf{index}(r)$ 
9:        $p_{i,pos} = p_{i,pos+1}$ 
10:       $p = r$ 
11:    end
12:  end
13: end

```

Em ambos os casos, para transformar uma série T usando a tabela de pesquisa β , uma palavra de tamanho ω requer ω buscas que contém α intervalos. Isso resulta em $O(\omega \log \alpha)$ operações (utilizando pesquisa binária). Senso assim, é possível calcular a

complexidade adotada no método HAR-SR caso o algoritmo de discretização seja o SAX através da equação:

Complexidade da extração de subestruturas	
Método	Complexidade
SAX (\mathbf{FT}_{sw})	$O(Nn + \omega \log \alpha)$
SFA (\mathbf{FT}_{sw})	$O(N(\omega n + w \log w + \omega \log \alpha))$

Complexidade das medidas estatísticas

As medidas estatísticas são utilizadas como novas características obtidas a partir do histograma P_η , onde η é o tamanho do espaço de amostras. O algoritmo 2 apresenta um método para calcular a entropia de um histograma P .

Algoritmo 2: *entropiaHistograma* (\mathbf{P}, η)

Parâmetros: histograma $\mathbf{P} = (p_1, p_2, \dots, p_\eta)$, tamanho do espaço η

```

1  Hmax = log( $\eta$ );
2  for i = 1:  $\eta$ 
3      if  $p_i > eps$ 
4          sum =  $p_i * \log(p_i)$ ;
5      else
6          sum = 0;
7      end
8  H = H - sum;
9  end
10 H = H/Hmax;
```

O cálculo de entropia, conforme apresentado no Algoritmo 2, pode ser calculado em tempo linear em η , e assim, sua complexidade computacional é:

$$\text{entropia}(P) = O(\eta)$$

Algoritmo 3: *divergênciaHistograma* ($\mathbf{P},$)

Parâmetros: histograma $\mathbf{P} = (p_1, p_2, \dots, p_\eta)$, tamanho do espaço η

```

1  beta=0.5;
2  jensen0 = -2*((( $\eta+1$ )/  $\eta$ )*log( $\eta+1$ )-2*log(2*  $\eta$ ) + log( $\eta$ ))(-1));
3  Hmax = log( $\eta$ );
4  for i = 1:  $\eta$ 
5      prob = (beta *  $p_i$ ) + (1-beta) * (1/  $\eta$ );
6      if prob > eps
7          sumj = prob * log(prob);
8      else
9          sumj = 0;
10     end
```

```

11   JS = JS - sumj;
12   end
13   JS = jensen0*(JS-beta*entropiaHistograma (P,eta)*Hmax-(1-beta)*Hmax);

```

O cálculo de divergência, conforme apresentado no Algoritmo 3, pode ser calculado em tempo linear em η , e como a entropia já havia sido calculada anteriormente, sua complexidade computacional é:

$$\text{divergencia}(P) = O(\eta) + O(\eta) = O(\eta)$$

O cálculo de complexidade consiste na relação entre as duas medidas, com custo $O(1)$ caso as medidas de entropia e complexidade tenham sido calculadas previamente.

Análise computacional do processo de classificação

O processo de classificação do HAR-SR consiste em associar uma série T_u não rotulada usando um conjunto de treinamento. Para esta tarefa, o HAR-SR utiliza o algoritmo KNN para executar uma pesquisa entre a nova série T_u e o conjunto de treinamento DS cujo objetivo é encontrar o rótulo para a série T_u . Então, a série T_u é associada a classe da série T que minimize o valor da métrica de distância adotada.

$$\text{rótulo}(T_u) = \text{argmin}_{T \in DS}(T_u, T)$$

Esse procedimento possui complexidade $O(N(k+n))$, onde N é o tamanho do conjunto de treinamento, n é o tamanho de cada instância do conjunto DS e k é o parâmetro do KNN. Cada cálculo de distância requer $O(n)$ execuções, e o custo desse cálculo em todo o conjunto DS é $O(Nn)$. Para selecionar os k valores mais próximos são necessários mais $O(kn)$ buscas.

Para finalizar, a complexidade total do método HAR-SR desde a transformação da série temporal até o processo de classificação é sumarizada na tabela:

	SAX	SFA
Pré-processamento (β)	-	$O(\omega N(\log N + 1 + n) + w \log w)$
Transformação (T)	$O(n)$	$O(\omega(n + \log \alpha) + w \log w)$
Extração de subestruturas (D)	$O(Nn)$	$O(N(w \log w) + N\omega(n - w))$
Classificação	$O(N(k+n))$	$O(N(k+n))$

Comparação com outros métodos			
Métodos	Complexidade no treino	Complexidade de teste	
		Limite inferior	Limite superior
SAX-SVM	$O(Nn^3)$	$O(n)$	
1-NN BOSS	$O(N^2n^2 + Nn^2 \log n)$	$= O(Nn + w \log w)$	
1-NN BOSS VS	$O(Nn^{\frac{3}{2}} \log n)$	$= O(n + w \log w)$	
SVM	$O(N^2n) \text{ à } O(N^3n)$	$O(n)$	

Onde: n =tamanho da série, w =tamanho da janela (tipicamente $w \ll n \ll N$), N =número total de séries (Bagnall, et al., 2016) (Schäfer, 2016).

Discussão sobre a complexidade computacional de redes neurais

As redes neurais têm apresentado bons resultados em praticamente todos os cenários de sua utilização. Para cada problema, existe um subtipo de algoritmos com modificações em suas estruturas internas que permitem detectar os mais variados tipos de padrões. No entanto, todo esse poder e flexibilidade possui um custo associado (He & Sun, 2014). Diversos trabalhos da literatura têm mostrado o poder de representação das redes neurais e o quão bons esses algoritmos são para obter as melhores acurácias dentre diversos algoritmos mais simples (Sze, et al., 2017).

Esta seção faz uma discussão do porquê esses métodos não são apropriados para dispositivos ou smartphones com baixo poder computacional (CPU e GPU de baixo desempenho), onde o desempenho das aplicações é limitada, principalmente em cenários reais de reconhecimento (He & Sun, 2014) (Ravi, et al., 2016)

O cálculo de complexidade de uma rede neural pode ser uma tarefa extremamente difícil por diversos fatores:

- i. A complexidade depende da arquitetura da rede.
- ii. Para cada arquitetura, existem diferentes tipos de parâmetros, como por exemplo, em uma CNN é preciso analisar as operações de convolução, tamanho de imagem, número de *kernels*, entre outros.
- iii. O processo de treinamento possui um parâmetro chamado *épocas* que é um potencial multiplicador da complexidade geral do método.
- iv. Muitas operações podem ser otimizadas quando utilizadas em hardware especiais, como GPUs. Tais recursos nem sempre estão disponíveis em dispositivos móveis.

- v. Pequenas modificações na arquitetura podem significar grandes mudanças na complexidade.

Por esses motivos, as complexidades dos diferentes tipos de redes neurais são difíceis de serem estimadas. Para analisar alguns aspectos da complexidade desses algoritmos, essa seção se baseia nos dois modos de operações mais básicos:

- A inferência (*forward propagation*), que é basicamente o conjunto de multiplicação de matrizes, cuja complexidade pode ser comparada com a complexidade encontrada no cálculo da multiplicação de matrizes.
- Treinamento (*backpropagation*)

A multiplicação de matrizes

A matriz de multiplicação é a operação fundamental no cálculo do *forward* e *backward* utilizando o *backpropagation*. Analisar o tamanho das estruturas que compõem essas redes é um ponto importante porque o custo de processamento escala conforme as matrizes se tornam maiores.

Algoritmo 4: multiplicaçãoMatriz(a, b, n)

Parâmetros: matriz a , b com tamanho n

```
1  for  $i = 1:n$ 
2    for  $j = 1:n$ 
3      for  $k = 1:n$ 
4         $c_{i,j} = c_{i,j} + a_{i,k} * b_{k,j}$ 
5      end
6    end
7  end
```

O algoritmo mais simples para se calcular a multiplicação de matrizes, como por exemplo o Algoritmo, possui complexidade computacional $O(n^3)$, assumindo-se que as matrizes possuem o mesmo número de linhas e colunas. Existem otimizações para o algoritmo de multiplicação de matrizes (algoritmo de *Coppersmith-Winograd*) que diminui a complexidade computacional para $O(n^{2.3737})$.

Complexidade de uma rede neural profunda

A complexidade de uma DNN pode ser representada por três fatores.

- i. A matriz de multiplicação, como o número de multiplicações e o tamanho das matrizes.

- ii. A transformação não-linear, que permite que essas redes aprendam funções não lineares.
- iii. O vetor de pesos compartilhados, é uma forma de otimizar o número de atualizações no *backpropagation*.

Esses três fatores devem dar uma boa estimativa para calcular a complexidade de uma rede (He & Sun, 2014), que pode ser estimada através da complexidade:

$$\text{CNN} = O\left(\sum_{l=1}^d n_{l-1} \cdot s_l^2 \cdot n_l \cdot m_l^2\right)$$

Onde l é o índice da camada de convolução, d é a profundidade (número de camadas de convolução), n_l é o número de filtros (largura) na l -ésima camada, n_{l-1} é o número de canais de entrada na l -ésima camada, s_l é o tamanho espacial (comprimento) do filtro e m_l é o tamanho espacial do mapa de saída.

A complexidade dos métodos TF, HAR-SR e CNN são apresentados na Figura abaixo. Para simplificar os cálculos, no método CNN todos os parâmetros possuem valores iguais.

