

Lara Mondini Martins

***Activity Recording*: um aplicativo para
reconhecimento de atividade humana usando
sensores de acelerômetro**

Uberlândia, Brasil

2019

Lara Mondini Martins

***Activity Recording*: um aplicativo para reconhecimento de atividade humana usando sensores de acelerômetro**

Trabalho de Conclusão de Curso submetido à Universidade Federal de Uberlândia como parte dos requisitos necessários para a obtenção do Grau de Bacharel em Sistemas de Informação. Sob a orientação da Professora Doutora Elaine Ribeiro de Faria Paiva.

Universidade Federal de Uberlândia

Faculdade de Computação

Bacharel em Sistemas de Informação

Orientador: Prof^a. Dr^a. Elaine Ribeiro de Faria Paiva

Uberlândia, Brasil

2019

Lara Mondini Martins

***Activity Recording*: um aplicativo para reconhecimento de atividade humana usando sensores de acelerômetro**

Trabalho de Conclusão de Curso submetido à Universidade Federal de Uberlândia como parte dos requisitos necessários para a obtenção do Grau de Bacharel em Sistemas de Informação. Sob a orientação da Professora Doutora Elaine Ribeiro de Faria Paiva.

Trabalho aprovado. Uberlândia, Brasil, 10 de julho de 2019:

Prof^a. Dr^a. Elaine Ribeiro de Faria Paiva
Orientador

Prof. Dr. Humberto Luiz Razente

Prof. Dr. Paulo Henrique Ribeiro Gabriel

Uberlândia, Brasil
2019

Agradecimentos

Agradeço a Deus por ter me dado força, saúde e sabedoria para superar os desafios. Minha orientadora, pelas dicas, conhecimento compartilhado, dedicação e compreensão durante o desenvolvimento deste trabalho. Aos meus pais e irmão, pelo amor, apoio, paciência e compreensão na reta final do curso. Ao meu namorado, pela paciência, amor, carinho e ajuda para resolver os bugs do aplicativo. Aos meus amigos, pelos momentos de distração e diversão. Aos familiares, pelo apoio. Aos professores que compartilharam seus conhecimentos e experiências durante um período muito importante em minha vida. Ao Programa de Educação Tutorial de Sistemas de Informação, PET-SI, pelos projetos de ensino, pesquisa e extensão que me ajudaram a conhecer amplamente a área acadêmica. À Universidade Federal de Uberlândia por ter me dado as ferramentas e incentivos necessários para concluir a graduação com qualidade. Muito obrigada!

A informática está interligada ao mundo sobre as reações intergalaxias!

Bill Gates

Resumo

A inatividade física é o quarto maior fator de risco de morte no mundo. Uma das ações que pode incentivar as pessoas a praticarem mais exercícios é o monitoramento de suas atividades. Sensores de acelerômetro, juntamente com técnicas de aprendizado de máquina, podem identificar a atividade desempenhada pelo usuário. Essa técnica é conhecida na literatura como Reconhecimento de Atividade Humana (RAH). Em cenários de RAH, é ideal que os dados sejam tratados como um Fluxo Contínuo de Dados (FCDs), porque os dados coletados do sensor de acelerômetro chegam de forma contínua e em alta velocidade, e são potencialmente infinitos. Este trabalho apresenta o aplicativo *Activity Recognition*, desenvolvido para o sistema operacional *Android*, que realiza a coleta de dados do sensor de acelerômetro do *smartphone* do usuário e identifica a atividade desempenhada pelo mesmo, usando técnicas de aprendizado de máquina. Além disso, o aplicativo disponibiliza uma funcionalidade destinada para pesquisadores da área de RAH que permite obter os dados parciais de cada etapa do Reconhecimento de Atividade Humana, desde a coleta até o reconhecimento da atividade. O aplicativo também permite a visualização gráfica das atividades identificadas no monitoramento do usuário. O aplicativo desenvolvido possui potencial para ser disponibilizado em um serviço de distribuição digital de aplicativos.

Palavras-chaves: Reconhecimento de Atividade Humana, Fluxo Contínuo de Dados, monitoramento de atividade, acelerômetro.

Lista de ilustrações

Figura 1 – Conjunto de tecnologia. Extraído de Silva, Botelho, Santos e Sanchez (2015)	15
Figura 2 – Aceleração dos sensores de acelerômetro	16
Figura 3 – Fluxo classificação de dados	17
Figura 4 – Teorema de Bayes	19
Figura 5 – <i>K-Nearest Neighbors</i> , $K=4$	20
Figura 6 – Exemplo árvore de decisão	20
Figura 7 – Diagrama de Caso de Uso Principal	28
Figura 8 – Diagrama de Caso de Uso Treinamento	29
Figura 9 – Diagrama de Caso de Uso Monitoramento	29
Figura 10 – Tela para coleta de dados do APP	30
Figura 11 – Tipos Janela	31
Figura 12 – Pré-processamento	33
Figura 13 – Treinamento	34
Figura 14 – Classificação de Dados	35
Figura 15 – Monitoramento em Tempo Real	37
Figura 16 – Finalizar Monitoramento em Tempo Real	38
Figura 17 – Sumariação Atividades Arquivo	39
Figura 18 – Sumariação Atividades do Monitoramento	40
Figura 19 – Teste Monitoramento Em Tempo Real	40

Lista de tabelas

Tabela 1 – Fórmula Atributos	32
--	----

Lista de abreviaturas e siglas

ADL	<i>Activities of Daily Living</i>
APP	Aplicativo
AR	<i>Activity Recognition</i>
ARFF	<i>Attribute-Relation File Format</i>
CSV	<i>Comma Separated Values</i>
FCDs	Fluxo Contínuo de Dados
GPS	<i>Global Positioning System</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
KDD	<i>Knowledge-Discovery in Databases</i>
KNN	<i>K-Nearest Neighbor</i>
LR	<i>Logistic Regression</i>
MLP	<i>Multi-layer Perceptron</i>
MOA	<i>Massive Online Analysis</i>
OMS	Organização Mundial da Saúde
RF	<i>Random Forest</i>
RAH	Reconhecimento de Atividade Humana
SVM	Máquinas de Vetor Suporte
VIGITEL	Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

Sumário

1	Introdução	11
1.1	Contextualização	11
1.2	Motivação	12
1.3	Objetivo Geral	12
1.4	Objetivos Específicos	13
1.5	Organização da Monografia	13
2	Revisão Bibliográfica	14
2.1	Introdução	14
2.2	Computação Ubíqua e Computação Vestível	14
2.3	Monitoramento de Saúde e <i>Fitness</i>	15
2.3.1	Acelerômetros e Monitoramento de Saúde	16
2.4	Etapas do Reconhecimento de atividade Humana	16
2.4.1	Coleta e Pré-Processamento de Dados	17
2.4.2	Extração de Atributos	18
2.4.3	Algoritmos de classificação	18
2.4.3.1	Naive Bayes	18
2.4.3.2	<i>K-Nearest Neighbors</i>	19
2.4.3.3	Árvores de decisão	20
2.5	Trabalhos Relacionados	21
2.5.1	RAH usando técnicas <i>batch</i> de aprendizado de máquina	21
2.5.2	RAH usando técnicas <i>online</i> de aprendizado de máquina	22
2.5.3	Aplicativos para monitoramento de atividades e saúde humana	23
2.6	Considerações Finais	24
3	Aplicativo: <i>Activity Recording</i>	26
3.1	Introdução	26
3.2	Visão Geral do Funcionamento do Aplicativo	26
3.2.1	Diagramas de Caso de Uso	27
3.3	Desenvolvimento do Aplicativo	28
3.3.1	Coleta de dados	29
3.3.2	Pré-processamento	30
3.3.3	Treinamento de Dados	32
3.3.4	Classificação de Dados	34
3.3.5	Monitoramento em Tempo Real	35
3.3.6	Sumariação de dados	36

3.4	Testando o Aplicativo	38
3.5	Considerações Finais	39
4	Conclusão	41
	Referências	42

1 Introdução

1.1 Contextualização

A prática de atividades físicas no Brasil ainda é insuficiente. Segundo a Organização Mundial da Saúde (OMS) (SAUDE, 2014) a inatividade física é o quarto mais importante fator de risco de morte no mundo. No Brasil, 62,1% das pessoas com 15 anos ou mais não praticaram qualquer tipo de esporte ou atividade física em 2015 (IBGE, 2015). A inatividade física pode levar a doenças crônicas como obesidade, diabetes e hipertensão.

Os exercícios físicos são muito importantes para manter a saúde e qualidade de vida das pessoas (DELUNARDO; RIBEIRO, 2018). Além de praticar exercícios, é importante dimensionar a quantidade de atividade realizada. Para isso, o monitoramento das atividades de um usuário é uma tarefa importante, que permite identificar as atividades mais frequentes, bem como quanto de atividades de repouso e movimento o usuário executa. Uma das formas de monitorar a quantidade de atividade desenvolvida é por meio de entrevistas e questionários, que são formatos subjetivos e que dependem da avaliação do usuário.

Estudos estão sendo desenvolvidos para ajudar a população a intensificar e monitorar a prática de atividades físicas. Um dos estudos desenvolvidos objetiva o uso de sensores de acelerômetros para reconhecer a atividade desenvolvida pelo usuário, o qual é um método quantitativo e mais preciso que métodos subjetivos (SILVA, 2015). Os acelerômetros estimam a aceleração, velocidade e deslocamento nos eixos X, Y e Z (RAVI, 2005). Esse sensor pode ser encontrado em *smartphones* e *smartwatches*, tornando-o acessível e de fácil manuseio.

A partir dos dados coletados dos sensores de acelerômetro e utilizando técnicas de aprendizado de máquina, é possível identificar a atividade desempenhada pelo usuário. Os trabalhos (BAYAT, 2014; RAVI, 2005; SUN, 2010) são exemplos de pesquisas que desenvolveram um estudo de pré-processamento dos dados brutos obtidos pelo sensor e posteriormente diferentes técnicas de classificação atividades foram aplicadas. Esses trabalhos alcançaram altas taxas de acurácia na identificação da atividade desempenhada.

Existem muitos estudos que mostram as diferentes etapas de Reconhecimento de Atividade Humana (RAH) desde a coleta até a classificação, porém poucos são os aplicativos desenvolvidos para tal tarefa. Além disso, muitos trabalhos executam a tarefa de classificação em um cenário *batch*, também chamado estático, e apenas poucos trabalhos consideram esses dados com um fluxo contínuo de dados (FCDs), que precisa ser monitorado de forma *online*.

Neste trabalho, foram utilizadas as técnicas de RAH, comumente usadas na literatura, para o desenvolvimento de um aplicativo (APP) *Android* nativo para RAH. O aplicativo poderá ser usado tanto por pessoas que desejem monitorar suas atividades, quanto por pesquisadores da área de RAH, que desejem montar seus próprios experimentos e treinar o modelo de classificação de atividades. O aplicativo ainda permite a visualização gráfica de um sumário de quais atividades foram desempenhadas durante um período de tempo.

1.2 Motivação

Com o objetivo de monitorar o desenvolvimento de atividades, vários trabalhos, como, (BAYAT, 2014; RAVI, 2005; SUN, 2010; MARTINS, 2018; GIACOMELLI, 2018a) desenvolveram um estudo teórico sobre Reconhecimento de Atividade Humana. Esses trabalhos atuam nas etapas de coleta de dados brutos, pré-processamento e classificação das atividades.

As pesquisas de (GIACOMELLI, 2018a) e (MARTINS, 2018) testaram o uso de diferentes técnicas de pré-processamento dos dados. Já os trabalhos de (BAYAT, 2014) e (RAVI, 2005) testaram diferentes classificadores na tarefa de RAH. Outro ponto estudado por alguns trabalhos é a importância da posição do sensor de acelerômetro na identificação da atividade (SUN, 2010).

Recentes estudos, como (TWOMEY, 2018), mostram a importância de tratar os dados dos acelerômetros como um fluxo contínuo de dados, o que exige o constante monitoramento das atividades desempenhadas pelo usuário, bem como uma constante atualização do modelo que classifica essas atividades.

Por outro lado, trabalhos como o de (WEISS, 2016) atuaram no desenvolvimento de um aplicativo para reconhecer as atividades que os usuários desempenharam enquanto estavam usando seus *smartphones*. Em ferramentas como o *Google Play*, é possível encontrar aplicativos que fazem o monitoramento da saúde do usuário.

Apesar de haver vários estudos e trabalhos teóricos de técnicas de pré-processamento e classificação de dados, poucos são os trabalhos que desenvolveram aplicativos de fato para o problema. Em especial, uma ferramenta que permite os pesquisadores montarem cenários personalizados para melhor realização de experimentos em seus estudos e que faça a classificação da atividade desempenhada pelo usuário.

1.3 Objetivo Geral

O objetivo desse trabalho foi desenvolver um aplicativo capaz de captar os dados do acelerômetro de um *smartphone Android*, pré-processar os dados e classificar as atividades

em tempo real.

1.4 Objetivos Específicos

- Criar uma funcionalidade para coletar dados do sensor de acelerômetro de *smartphones Android* e armazená-los em arquivo;
- Criar uma funcionalidade para converter os dados brutos do sensor de acelerômetro para dados pré-processados, extraíndo medidas estatísticas a partir de janelas de dados;
- Implementar um algoritmo de classificação no sistema operacional *Android* para que ele possa identificar as atividades desenvolvidas pelo usuário, a partir dos dados brutos;
- Implementar uma funcionalidade que permita que o usuário treine o modelo de classificação de dados;
- Sumariar em um gráfico as atividades desempenhadas pelo usuário em um período de tempo;
- Implementar uma funcionalidade especial para pesquisadores da área de RAH, que permite a execução das etapas de RAH separadamente.

1.5 Organização da Monografia

A monografia está organizada da seguinte forma:

- Capítulo 1: apresentou a importância da atividade física na vida das pessoas, a usabilidade dos acelerômetros para o Reconhecimento de Atividade Humana, a motivação e os objetivos deste trabalho.
- Capítulo 2: explica os conceitos, técnicas e referências bibliográficas que serão utilizadas para o desenvolvimento do trabalho.
- Capítulo 3: descreve o funcionamento do aplicativo *Activity Recording*, detalhando as funcionalidades e técnicas utilizadas para o desenvolvimento do APP.
- Capítulo 4: relata as conclusões do trabalho.

2 Revisão Bibliográfica

2.1 Introdução

Este capítulo apresenta os conceitos e técnicas utilizados na literatura para o RAH, bem como os algoritmos de classificação de dados usados para tal tarefa. A seção 2.2 apresenta os termos de computação ubíqua, vestível, pervasiva e móvel. Na seção 2.3 é mencionado a importância dos acelerômetros e dos aplicativos que fazem o monitoramento de saúde e *fitness*. A seção 2.4 descreve as etapas a serem seguidas para o reconhecimento da atividades. E por fim a seção 2.5, apresenta os trabalhos relacionados á área de RAH, FCDs e algoritmos de classificação.

2.2 Computação Ubíqua e Computação Vestível

Os conceitos de computação Ubíqua, Pervasiva e Móvel estão relacionados entre si, e existem diferença entre esses conceitos (SILVA, 2015). A Figura 1 mostra a relação entre computação Ubíqua, Pervasiva e Móvel, que apesar de serem conceitos diferentes, eles complementam um ao outro. Esses três termos podem ser definidos da seguinte forma:

- **Computação Ubíqua:** define interação humano-computador no qual o processamento da informação é completamente integrado aos objetos e atividades cotidianas (ZHAO; WANG, 2011). Exemplo: iluminação de um imóvel controlada por meio de controle remoto.
- **Computação Pervasiva:** é o conceito mais abrangente que define que a tecnologia está presente naquele lugar, mas é invisível aos humanos (SILVA, 2015). Um exemplo de Computação pervasiva é o uso de sensores em casas inteligentes.
- **Computação Móvel:** define que a computação pode ser levada a qualquer lugar, com isso o acesso a informação pode ser feito a qualquer momento (SILVA, 2015). *Smartphones* e *notebooks* são exemplos de Computação Móvel.

Outro assunto recente na área de computação, relacionado a esses três termos, é a Computação Vestível ou *Wearable Computing*. Computadores vestíveis são pequenos dispositivos que podem estar no corpo do usuário, em outras palavras o usuário "veste" a tecnologia. Um exemplo de computação vestível são os *smartwatches*, que são relógios inteligentes. Starner (2002) diz que a computação vestível permite a exploração de princípios da Computação Pervasiva, porque não precisa esperar pela implementação de uma

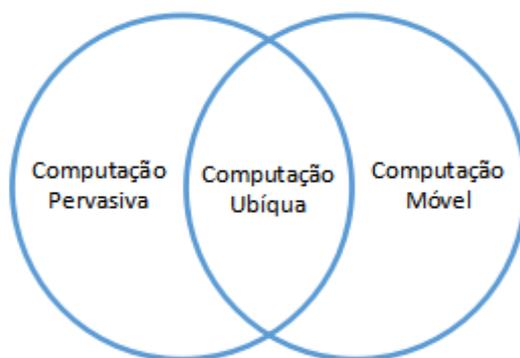


Figura 1: Conjunto de tecnologia. Extraído de [Silva, Botelho, Santos e Sanchez \(2015\)](#)

infraestrutura frequentemente associada a um espaço físico. É comum o uso da Computação Vestível para monitoramento de pessoas, por exemplo, para monitoramento cardíaco e monitoramento de atividades.

O aplicativo desenvolvido neste trabalho usa todos os conceitos apresentados. A computação móvel será usada porque o sistema será implementado para dispositivos com sistema operacional *Android*. Computação Ubíqua e Pervasiva, pois terá o uso do sensor de acelerômetro do dispositivo do usuário, assim a tecnologia estará presente, sem que o usuário perceba. E por fim, o uso de *Wearable Computing*, também por meio dos sensores de acelerômetro, que podem estar disponíveis em dispositivos vestíveis como *smartwatches*.

2.3 Monitoramento de Saúde e *Fitness*

A prática de atividades físicas traz muitos benefícios para a saúde, prevenindo doenças como pressão arterial, diabetes, gordura no fígado, colesterol alto e obesidade. Entretanto, 53,8% dos brasileiros estão acima do peso ([BRASIL, 2016](#)). Para ajudar as pessoas a monitorarem a saúde e a prática de atividades físicas foram desenvolvidos aplicativos com esse objetivo, como o *Samsung Health*, que mostra a quantidade de passos diários, tempo em que o usuário praticou alguma atividade física, calorias perdidas etc.

Uma pesquisa realizada pela *Mobiquity* mostrou que 70% das pessoas usam aplicativos em seus *smartphones* para monitoramento da saúde ([REPORTER, 2014](#)). Essa pesquisa mostrou que o uso de aplicativos que monitoram a saúde das pessoas é predominante.

De acordo com essa pesquisa, os aplicativos mais usados são os que monitoram ingestão de calorias, monitoramento de peso e atividades físicas. As pessoas relataram que se sentiam mais saudáveis usando aplicativos de monitoramento de saúde e *fitness*

(GORDILHO, 2014). Além disso o uso desse tipo de aplicativo incentiva a prática de atividades físicas e o acompanhamento da saúde.

Alguns aplicativos relacionados ao monitoramento de saúde e *fitness* serão citados e descritos com detalhes na seção 2.5.3.

2.3.1 Acelerômetros e Monitoramento de Saúde

Acelerômetros são sensores capazes de captar a aceleração nos eixos X, Y e Z, conforme ilustrado na Figura 2. Aceleração é quem determina a taxa de variação da velocidade em função do tempo. Esses sensores são encontrados em *smartphones* e *smartwatches*.

Os sensores de acelerômetros podem ser usados em diversas aplicações, algumas delas são: alarmes de carros, medição de inclinação, forças inerciais e choque ou vibração (BUDZINSKI, 1999). Os acelerômetros também são usados para fazer o monitoramento de idosos e monitoramento de atividades físicas.

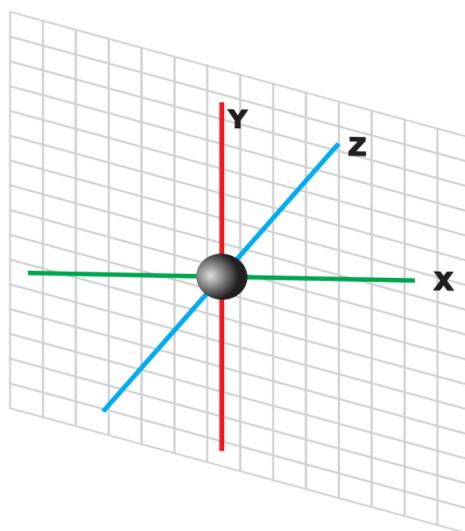


Figura 2: Aceleração dos sensores de acelerômetro

Muitos aplicativos *fitness* que existem atualmente (RUNTASTIC, 2018), (ELECTRONICS, 2018) usam sensores de acelerômetros e de GPS para monitorar as atividades de seus usuários. Neste trabalho o sensor de acelerômetro de *smartphones Android* será usado para o monitoramento de atividades do usuário.

2.4 Etapas do Reconhecimento de atividade Humana

O Reconhecimento de Atividade Humana se resume em três etapas (SUN, 2010): (1) coletar e pré-processar os dados; (2) extrair atributos; e (3) aplicar algoritmos de classificação previamente treinados.

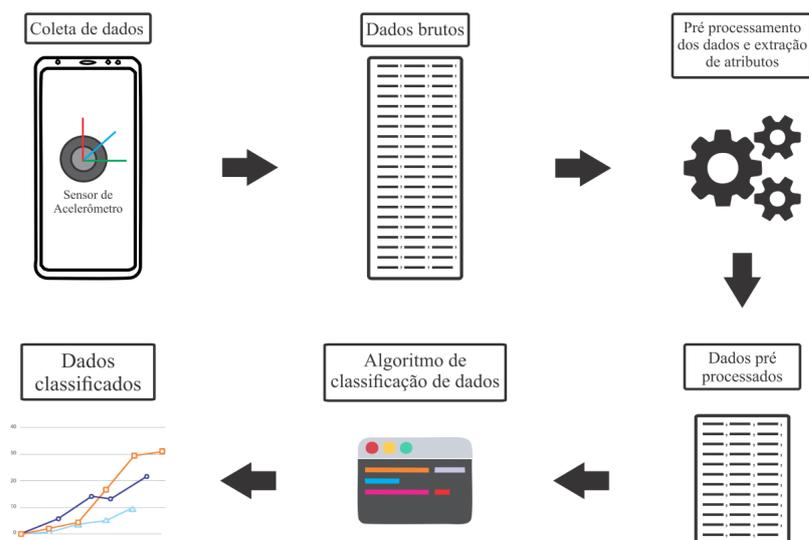


Figura 3: Fluxo classificação de dados

O fluxo de classificação de dados está representado na Figura 3. As seções 2.4.1, 2.4.2, 2.4.3 explicam como cada etapa do RAH ocorre.

2.4.1 Coleta e Pré-Processamento de Dados

A primeira etapa de classificação de dados é a coleta de dados, que é feita por meio de um sensor de acelerômetro. Esse tipo de sensor é, normalmente, encontrado em dispositivos inteligentes, como *smartphones* e *smartwatches*. Os sensores de acelerômetro captam a aceleração nos eixos X, Y e Z, como representado na Figura 2. A coleção de dados gerada pelo sensor de acelerômetro, em geral, é armazenada em um arquivo do tipo CSV.

Geralmente os dados coletados pelos acelerômetros são: *timestamp*, aceleração nos eixos X, Y e Z. Alguns trabalhos também coletam dados do sensor de giroscópio (WANG, 2019), (BULLING, 2014) e (HA; CHOI, 2016).

Algumas bases de dados brutos foram disponibilizadas publicamente, como por exemplo, a WISDM (KWAPISZ, 2011). Nessa base, os dados coletados pelos autores foram: código do usuário, atividade, *timestamp* e aceleração nos eixos X, Y e Z.

Após a coleta de dados, é feito o pré-processamento desses dados, que consiste na exclusão de dados que não foram coletados corretamente na etapa anterior (GIACOMELLI, 2018b).

As duas fases descritas estão representadas nos dois primeiros passos da Figura 3.

2.4.2 Extração de Atributos

Os dados brutos, capturados na primeira etapa, precisam ser organizados para que seja possível aumentar a capacidade de informação. Diante disso, na segunda etapa de RAH acontece a extração de atributos, no qual o arquivo de dados brutos é dividido em janelas de tamanhos pré-definidos. Após a divisão dessas janelas são extraídas medidas de posição (média e mediana) e medidas de dispersão (desvio padrão e variância), bem como outras medidas tais como distribuição binomial, energia e aceleração média. Essa fase está representada no terceiro passo da Figura 3.

Depois da coleta de dados, pré-processamento e extração de atributos, os dados são colocados num formato para execução dos algoritmos de classificação, por exemplo o *Attribute-Relation File Format* (ARFF), que apresenta uma lista de instâncias que compartilham um conjunto de atributos (WIKI, 2018). Esse tipo de arquivo é usado nos *frameworks* WEKA (HALL, 2009) e MOA (BIFET, 2010). Essa fase está representada no quarto passo da Figura 3.

2.4.3 Algoritmos de classificação

A última etapa da classificação dos dados é a execução do algoritmo de classificação. Em geral, os algoritmos de classificação precisam de um treinamento, no qual treinam um modelo de decisão usando um conjunto de instâncias rotuladas, ou seja, instâncias que possuem o rótulo da atividade. Uma vez treinado, o algoritmo pode ser usado para classificar novas instâncias.

Os algoritmos descritos nas próximas subseções são os mais utilizados na literatura para a classificação de dados. Existem duas formas de treinar um algoritmo de classificação: (1) *offline*: o algoritmo aprende uma única vez como classificar as instâncias e a chegada de novos dados, não muda o modelo usado para classificação; (2) *online*: considera a chegada de novas instâncias de dados para atualizar o modelo usada para classificação.

Após essa etapa, com os dados já classificados, é possível analisar e sumariar o desempenho de atividades realizado pelo usuário. Essas duas fases estão apresentadas na Figura 3 nos passos cinco e seis, respectivamente.

2.4.3.1 Naive Bayes

O algoritmo Naive Bayes é um modelo probabilístico baseado no Teorema de Bayes. Esse teorema calcula a probabilidade de um evento ocorrer dado uma condição (CORREA, 2003). Observando a Figura 4, o Teorema de Bayes calcula a probabilidade do evento $B1$ ocorrer sabendo que existe A .

Segundo Tan, Steinbach, Karpatne e Kumar (2013), o algoritmo *Naive Bayes* estima a probabilidade condicional da classe, assumindo que os atributos são indepen-

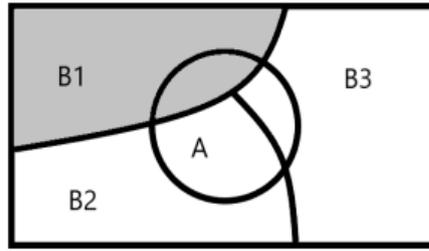


Figura 4: Teorema de Bayes

dentos, dado o valor do atributo classe. Assim, seja X um conjunto de d atributos $X = X_1, X_2, \dots, X_d$, para classificar uma nova instância, o classificador Naive Bayes calcula a probabilidade posterior para cada classe Y , de acordo com a Equação 2.1.

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y) \quad (2.1)$$

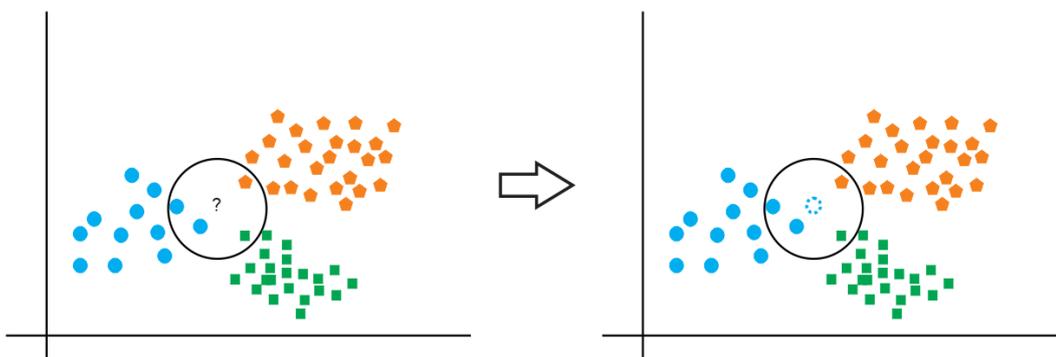
O algoritmo Naive Bayes constrói uma tabela de probabilidade dado um atributo e seu valor e assim que uma nova instância é rotulada, ela é inserida nessa tabela de probabilidades, portanto, o algoritmo é incremental (TAN, 2013).

2.4.3.2 *K-Nearest Neighbors*

O algoritmo de classificação dos K vizinhos mais próximos, *K-Nearest Neighbors* (KNN), calcula a distância Euclidiana dos seus K vizinhos mais próximos e classifica a instância desconhecida com a classe da maioria desses K vizinhos. Os K vizinhos mais próximos são obtidos a partir de uma base de dados de treinamento, ou seja, uma base de dados rotulada.

KNN é um algoritmo não paramétrico, ou seja, não faz qualquer suposição sobre a distribuição dos dados. A Figura 5 mostra o funcionamento do KNN, considerando a constante K igual a 4. Primeiro, o algoritmo calcula a distância de todos os pontos e encontra os k -vizinhos mais próximos da instância a ser classificada. A seguir, rotula a instância com a mesma classe da maioria desses k vizinhos.

O algoritmo KNN pode ser executado tanto na versão *batch* quanto na *online*. Na versão *batch*, considera-se que uma vez obtido a base de treinamento, esta não muda com o passar do tempo. Na versão *online*, considera-se uma janela deslizante, usada para selecionar o conjunto de treinamento, de modo a usar instâncias mais recentes de dados rotulados para classificar novos dados do fluxo.

Figura 5: *K-Nearest Neighbors*, $K=4$

2.4.3.3 Árvores de decisão

As árvores de decisão são algoritmos de aprendizado de máquina que usam a estrutura de uma árvore para classificação de dados. A abordagem usada por esse algoritmo é *dividir para conquistar* (GARCÍA-MARTÍN, 2017). A árvore de decisão contém nós pai, filhos e as folhas. O nó pai dá origem aos nós filhos e as folhas são os nós sem filhos.

Na Figura 6 é mostrado um exemplo do funcionamento desse algoritmo. A árvore de decisão ilustrada tem o objetivo de decidir se uma pessoa vai à praia dado as condições climáticas. O nó pai é o tempo, que pode ser ensolarado, tempestade ou nublado. Se o tempo for ensolarado, então a pessoa deve ir à praia, se o tempo for tempestade a pessoa não deve ir à praia. E se o tempo for nublado, é necessário analisar se está frio ou calor, no caso de calor, deve-se ir à praia, no caso de frio não.

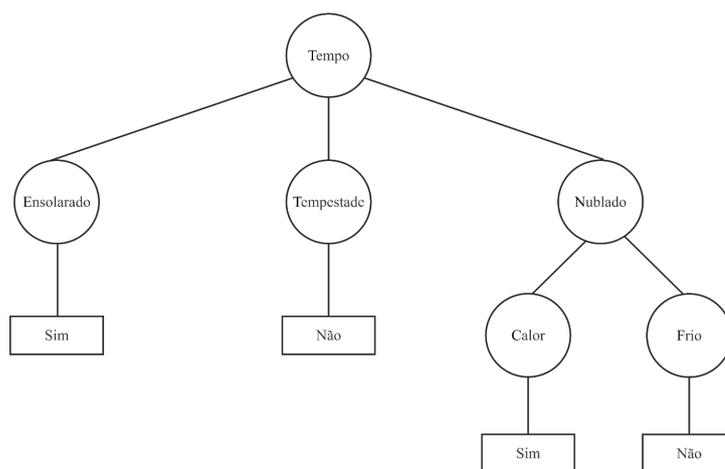


Figura 6: Exemplo árvore de decisão

As árvores de decisão podem ser usadas para resolver diversos tipos de problemas relacionados a regressão ou a classificação. Os problemas de regressão, estão relacionados a classe de dados numéricas, e os de a classificação estão relacionados a classe de dados nominais (GARCÍA-MARTÍN, 2017). Portanto, o exemplo da Figura 6 é um problema

de classificação.

As árvores de decisão também podem ser usadas na versão *online* para cenários de Fluxo Contínuo de Dados. Nessa versão, a árvore é construída de forma incremental, de acordo com a chegada de dados (GARCÍA-MARTÍN, 2017).

2.5 Trabalhos Relacionados

Essa seção descreve os trabalhos relacionados a área de pesquisa de Reconhecimento de Atividade Humana que usaram sensores de acelerômetros. Na seção 2.5.1 é apresentado os trabalhos que usaram algoritmos de classificação estáticos (*batch*). Já na seção 2.5.2 os trabalhos que usaram algoritmos de classificação *online* são apresentados. Por fim, na seção 2.5.3 foi retratado aplicativos que estão presentes no *GooglePlay* que fazem o monitoramento de saúde.

2.5.1 RAH usando técnicas *batch* de aprendizado de máquina

Os trabalhos relacionados aos métodos de classificação estáticos, supõe que o número de atividades desempenhadas pelo usuário é previamente conhecida e que exemplos de todas essas classes de atividades estão disponíveis para treinamento do classificador. Além disso, esses trabalhos também supõem que a forma como um usuário desempenha uma atividade não muda ao longo do tempo e portanto, o treinamento do modelo de classificação ocorre uma única vez.

O trabalho de (KWAPISZ, 2011) descreve um sistema que usa acelerômetros, presentes em *smartphones*, para Reconhecimento de Atividade Humana. O sistema operacional usado pelos autores foi o *Android*. Os atributos extraídos dos dados brutos do sensor foram: média, desvio padrão, diferencial absoluto médio, aceleração média resultante, tempo entre picos e distribuição dos dados em *bins*. Para a classificação das atividades os autores usaram o *framework* WEKA com as configurações padrão e os classificadores árvore de decisão (J48), regressão logística e redes neurais multicamadas. Para a validação foi utilizada o método de validação cruzada de 10 pastas e a medida acurácia. Nesse trabalho, as atividades desempenhadas por 29 usuários foram: andar, correr, subir escadas, sentar e permanecer em pé.

Em (RAVI, 2005), foi realizada uma comparação entre o desempenho de classificadores de nível base e de nível meta para o RAH. Um único sensor de acelerômetro foi usado e estava localizado na região pélvica do corpo. Os atributos extraídos dos dados brutos foram: energia, correlação, média e desvio padrão. Para o Reconhecimento de Atividade Humana foi usado o *framework* WEKA e os classificadores de nível base: árvore de decisão (C4.5), tabelas de decisão, SVM (Máquinas de Vetor Suporte), Naive Bayes e

KNN. As atividades desempenhadas por duas pessoas foram: andar, ficar de pé, subir e descer escadas, escovar os dentes, passar aspirador, sentar e levantar.

No estudo de (SUN, 2010) foi usado um sensor de acelerômetros para coleta de dados de atividades físicas desempenhadas diariamente. O sensor usado no trabalho estava presente no dispositivo Nokia N97s e ele foi colocado em seis posições diferentes no bolso do usuário. Os atributos extraídos foram: média, variação, correlação, energia e *Frequency-Domain Entropy*. Para o reconhecimento das atividades foi usado o classificador SVM, o método de validação cruzada de 10 pastas e medidas de acurácia, precisão e revocação. As atividades desempenhadas por sete pessoas foram: andar, correr, andar de bicicleta, permanecer parado, dirigir, subir e descer escadas.

O trabalho de (GIACOMELLI, 2018a) propôs um estudo comparativo entre as principais ferramentas, métodos e bases de dados para RAH. Nesse estudo, os classificadores árvore de decisão, Naive Bayes e KNN foram comparados usando duas bases de dados públicas WISDM e CHEST. Além disso, foi realizado um estudo comparativo para analisar o impacto do uso de diferentes conjuntos de atributos no desempenho dos classificadores. Os resultados indicaram o KNN como o melhor classificador.

2.5.2 RAH usando técnicas *online* de aprendizado de máquina

A forma que um humano desempenha uma atividade pode mudar ao longo do tempo, o que torna necessário atualizar o modelo de decisão que descreve a forma como um indivíduo desempenha uma atividade (IGLESIAS, 2010). Dessa forma novos estudos foram e estão sendo realizados para que os classificadores se adaptem a mudanças. Os trabalhos de RAH que usam técnicas online de aprendizado de máquina apresentam estudos de classificadores que se adaptam a modificações e aparecimento de novas classes de atividades.

O trabalho de (IGLESIAS, 2010) é uma abordagem baseada em sistemas difusos que criam modelos dinâmicos que evoluem ao longo do tempo. A abordagem proposta é aplicada em sistema de reconhecimento de atividades da vida diária (*Activities of Daily Living [ADLs]*), que incluem atividades como lavar as mãos e fazer uma chamada telefônica. Os dados usados na tarefa de classificação representam leituras de sensores coletadas em um pequeno apartamento para testes. Cada atividade é descrita por uma ou mais regras difusas, que evoluem ao longo do tempo. Os resultados indicam que o algoritmo proposto obtém resultados similares ao de técnicas de classificação *offline* com apenas um passo sobre os dados.

Os trabalhos de (ABDALLAH, 2016), (NGUYEN, 2015) apontam que no domínio de RAH é importante lidar como modelos de decisão que são atualizados ao longo do fluxo de dados, de forma a tratar atividades que estão em evolução, ou seja, que apa-

recem e desaparecem ao longo do fluxo, e tratar atividades anormais, como uma queda súbita. Além disso, é importante que o modelo seja capaz de reconhecer novas classes de atividades, por exemplo dirigir, quando exemplos da classe dirigir não estão disponíveis na etapa de treinamento.

Abdallah, Gaber, Srinivasan e Krishnaswamy (2016) propõe uma técnica para detecção de novidade em fluxos contínuos de dados aplicando-o à tarefa de Reconhecimento de Atividade Humana. A técnica é capaz de detectar a emergência de novos conceitos e usa aprendizado ativo a fim de escolher os melhores exemplos a serem rotulados pelo especialista para atualização do modelo de decisão. Para os experimentos, foram usadas duas bases de dados públicas, OPPORTUNITY e WISDM e diferentes medidas de validação como acurácia, precisão, revocação, especificidade e taxa de alarme falso. No entanto, os resultados do trabalho não foram comparados com outras propostas da literatura.

Em (NGUYEN, 2015) foi estudado o problema de reconhecer novas classes de atividade usando uma quantidade limitada de dados rotulados. A ideia básica do trabalho é considerar o significado semântico de novas atividades e seu relacionamento com atividades existentes, compensando assim a falta de dados rotulados. O algoritmo proposto foi aplicado em três bases de dados públicas, *Mhealth*, *DailyAndSport*, e *RealDisp*, com dados coletados de 10, 8 e 17 usuários, respectivamente. As atividades presentes nas bases são atividades diárias, como andar e sentar, e atividades físicas, como jogar basquete e andar de bicicleta. Para o pré-processamento foi usado janela deslizante com sobreposição de 50% e os atributos mínimo, máximo, média e correlação. As medidas usadas para avaliação foram precisão e revocação.

O objetivo do trabalho de (TWOMEY, 2018) é estudar o RAH com as configurações do mundo real. Os autores dizem que em vez da classificação ocorrer instância por instância, é melhor que haja uma janela de dados, assim o custo computacional será menor. Foram usadas três bases de dados públicas HAR, USCHAD e PAMAP2, e os seguintes algoritmos de classificação: *Random Forest (RF)*, *Logistic Regression (LR)*, e *Multi-layer Perceptron (MLP)*. Uma das conclusões dos autores é que uma grande janela de dados para a extração de atributos pode ajudar os algoritmos na classificação da instância, porque uma proporção maior do todo é conhecida.

2.5.3 Aplicativos para monitoramento de atividades e saúde humana

Existem muitos aplicativos que fazem monitoramento de atividades e/ou saúde, *Google Fit*, *Samsung Health*, *SportsTracker*, etc. Eles têm objetivo de monitorar as atividades desenvolvidas pelo usuário ao longo do dia. Normalmente, eles fazem comunicação com outros APPs relacionados a saúde, que coletam frequência cardíaca, localização em que as atividades foram desempenhadas, peso, calorias gastas etc. Esses aplicativos ajudam as pessoas a acompanharem a saúde de uma maneira simples e rápida, usando um

smartphone.

Google Fit (GOOGLE, 2018) é um aplicativo desenvolvido pela Google que tem o objetivo de monitorar a saúde dos usuários. O lema do aplicativo é "*Coaching you to a healthier and more active life*". Esse APP tem parceria com a Organização Mundial da Saúde, que recomenda dois pontos importantes que impactam a saúde de uma pessoa: atividades físicas e batimentos cardíacos. Esses dois pontos são destacados no Google Fit como: (1) *Move Minutes*: que capta as atividades que foram desenvolvidas pelo usuário; (2) *Health Points*: capta os pulsos do coração do usuário. O aplicativo também apresenta o resumo e o histórico dos dados captados. O APP é compatível com *smartphones* com sistema operacional *Android* e o *smartwatch Wear OS by Google* e está disponível na *Google Play*. O monitoramento de atividades é feita em *background*.

Samsung Health (ELECTRONICS, 2018) é um aplicativo que pertence a Samsung que tem o objetivo de monitorar a saúde de seus usuários. O aplicativo já vem instalado por padrão nos dispositivos *Samsung* e funciona apenas no sistema operacional *Android* e no *smartwatch* da *Samsung*. O lema do *Samsung Health* é "Sua saúde em suas mãos". O APP permite que o usuário controle suas atividades, alimentação, dados de estresse, frequência cardíaca e sono, e oferece o resumo e histórico dos dados do usuário. O monitoramento das atividades é feita em *background*, assim todas as atividades que o usuário desempenhar o APP consegue captar. O *Apple Health* (APPLE, 2018) é um aplicativo da *Apple* e é parecido com o *Samsung Health*, e está disponível apenas para *smartphones* com sistema operacional *IOS* e para o *smartwatch Apple Watch*. Seu lema é "Uma visão mais completa da sua saúde". Ambos aplicativos permitem a comunicação com aplicativos externos.

Sports Tracker (TECHNOLOGIES, 2018) é um aplicativo que tem o objetivo de monitorar as atividades que o usuário desempenha. Diferente dos demais aplicativos apresentados, para que o APP capte as atividades que são desempenhadas, o usuário precisa iniciar uma atividade pré-rotulada, assim ele obtém os dados de duração, distância, velocidade, velocidade média, ritmo de passos, energia (calorias), altitude e passos. O *Sports Tracker* mostra os dados em tempo real em números e em gráficos. No fim da atividade é apresentado um resumo dos dados captados, mapa do local de onde a atividade foi desempenhada e a quilometragem percorrida. Além disso, é mostrado o sumário das atividades feitas nos últimos trinta dias. O aplicativo está disponível para *smartphones* com sistema operacional *Android* e *IOS*. O APP também faz a comunicação com o *smartwatch Suunto Spartan* e o *Smart Sensor* que é uma cinta que é colocado na barriga do usuário.

2.6 Considerações Finais

Neste capítulo foi discutido os termos computação ubíqua, vestível, pervasiva e móvel. Além disso, foi abordado a importância das atividades físicas e aplicativos de

monitoramento de saúde e *fitness* para manter e aumentar a qualidade de vida e saúde da população.

Este capítulo também apresentou os principais trabalhos relacionados da área de Reconhecimento de Atividade Humana, bem como alguns dos principais algoritmos de classificação *batch* e *online* usados na área. Por fim, este capítulo descreveu o funcionamento de alguns aplicativos que usam RAH para monitorar a saúde dos usuários.

Grande parte dos trabalhos relacionados ao Reconhecimento de Atividade Humana estudam a parte teórica do RAH e os experimentos realizados com os algoritmos de classificação utilizam bases de dados públicas. As bases de dados públicas nem sempre retratam a realidade e os testes realizados pelos pesquisadores, na maioria das vezes, tratam de um cenário irreal. Além disso, os testes dos algoritmos de classificação podem ficar viciados, já que as mesmas bases públicas são utilizadas em muitos trabalhos relacionados a tarefa de RAH.

Dado isso, este trabalho desenvolveu um aplicativo que realiza todas as etapas do RAH e o monitoramento de atividades de um usuário. No próximo capítulo serão apresentados os detalhes de desenvolvimento do aplicativo *Activity Recording*.

3 Aplicativo: *Activity Recording*

3.1 Introdução

O Reconhecimento de Atividade Humana tem o objetivo de identificar qual atividade foi desempenhada por um humano. O RAH pode ser utilizado em aplicativos que realizam o monitoramento das atividades de um usuário e isto pode incentivar o usuário a praticar mais atividades físicas.

O Capítulo 2 discutiu os conceitos, técnicas e trabalhos relacionados ao RAH apresentados na literatura. Além disso, mostrou exemplos de aplicativos que usam Reconhecimento de Atividade Humana.

Este capítulo descreve com detalhes o funcionamento do aplicativo *Activity Recording*, bem como as técnicas e algoritmos usadas em cada etapa. O APP foi desenvolvido para *smartphones* com o sistema operacional *Android* e faz o uso da metodologia e dos algoritmos comumente usados no processo de descoberta do conhecimento, do inglês *Knowledge Discovery in Databases* (KDD), em base de dados de acelerômetros.

A seção 3.2 apresenta em linhas gerais o funcionamento do aplicativo. A seção 3.3 descreve com detalhes as técnicas utilizadas para o desenvolvimento do APP. Por fim, a seção 3.5 apresenta as considerações finais do capítulo.

3.2 Visão Geral do Funcionamento do Aplicativo

O aplicativo *Activity Recording* tem o objetivo de coletar dados do sensor de acelerômetro, identificar qual atividade o usuário está desempenhando e gerar gráficos que sumariem as atividades realizadas em período de tempo. O APP pode ajudar pesquisadores da área de RAH a coletar os dados do sensor de acelerômetro para montagem de cenários de teste personalizados. Além disso, o aplicativo pode ser utilizado por usuários que desejam monitorar suas atividades utilizando o seu *smartphone*.

O fluxo principal de dados seguido no aplicativo está representado na Figura 3. A seguir são descritos com detalhes os passos seguidos pelo APP.

1. Coleta de dados: é feita por meio do sensor de acelerômetro que está presente no *smartphone* do usuário. Os dados coletados pelo sensor de acelerômetro são chamados de dados brutos;
2. Pré-processamento: os dados brutos, coletados na etapa anterior, são utilizados para extração de atributos, tais como, média, desvio padrão, maior, menor, energia, desvio

absoluto médio, variância, correlação, aceleração média e binomial. Assim, os dados brutos são pré-processados utilizando técnicas para extração de atributos e geram os chamados dados pré-processados.

3. Classificação dos dados: os dados pré-processados obtidos na etapa anterior são classificados pelo algoritmo KNN;
4. Sumariação: o resultado da classificação das atividades é sumariado em um gráfico para que o usuário possa visualizar com clareza as atividades que foram desempenhadas em um dado período.

O aplicativo possui dois módulos distintos: (1) destinado aos usuários que desejarem monitorar suas atividades e (2) destinado aos pesquisadores da área de Reconhecimento de Atividade Humana.

O módulo destinado aos usuários que desejarem monitorar suas atividades possui as funcionalidades de treinamento, monitoramento de atividades e sumariação das atividades desempenhadas ao longo do tempo. As telas do aplicativo destinadas a esse módulo são apresentadas nas Figuras 13, 15 e 18.

A funcionalidade de treinamento faz todas as etapas do RAH e salva os dados em um arquivo armazenado no *smartphone* do usuário. A funcionalidade de monitoramento de atividades faz todas as etapas de RAH em tempo real, ou seja, os dados brutos coletados do acelerômetro, são pré-processados, classificados e sumariados no momento em que a atividade está sendo feita pelo usuário.

O módulo destinado aos pesquisadores da área de Reconhecimento de Atividade Humana permite que o usuário faça separadamente a coleta de dados brutos, pré-processamento dos dados, classificação e sumariação das atividades classificadas em um gráfico. Uma tela do aplicativo é destinada a esse módulo, e ela está representada nas Figuras 10, 12 e 17.

As funcionalidades de coleta de dados brutos, pré-processamento e classificação destinadas aos pesquisadores de RAH armazenam os dados em arquivos em estão localizados no *smartphone* do usuário.

3.2.1 Diagramas de Caso de Uso

Os diagramas de caso de uso representam a visão de alto nível do sistema. Eles são utilizados para explicitar as principais funcionalidade do sistema e a interação dessas funcionalidades com o usuário (RIBEIRO, 2012).

A figura 7 mostra o diagrama de caso de uso principal do aplicativo *Activity Recording*. O usuário pesquisador tem acesso as funcionalidades de coleta de dados, pré-

processamento dos dados, classificação dos dados, treinamento e sumariação. E o usuário final tem acesso as funcionalidades de treinamento, monitoramento em tempo real e sumariação.

O caso de uso 'Normalizar Dados' tem como pré-requisito 'Pré-Processar Dados', ou seja, para que os dados sejam normalizados é obrigatório o pré-processamento dos dados. A mesma situação acontece para os casos de uso 'Classificar Dados' e 'Treinamento'. O caso de uso 'Coletar Dados Brutos' utiliza um sensor de acelerômetro para realizar a coleta.

O caso de uso 'Treinamento' utiliza as funcionalidades de coleta de dados brutos e pré-processamento dos dados. Veja figura 8.

O caso de uso 'Monitoramento Tempo Real' utiliza as funcionalidades de coleta de dados brutos, pré-processamento dos dados e classificação dos dados. Ao final da classificação dos dados, o rótulo da atividade é armazenado no banco de dados. A figura 8 ilustra o diagrama desse caso de uso.

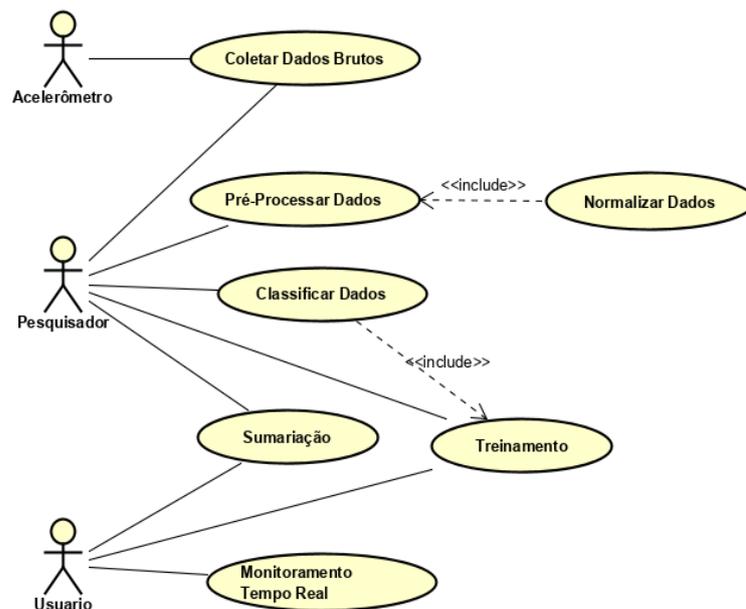


Figura 7: Diagrama de Caso de Uso Principal

A seção 3.3 descreve com detalhes a parte técnica utilizada para o desenvolvimento do aplicativo. Além disso mostra as telas respectivas de cada funcionalidade.

3.3 Desenvolvimento do Aplicativo

O aplicativo *Activity Recording* foi desenvolvido para o sistema operacional *Android* e usou a linguagem de programação Java. A ferramenta utilizada para o desenvolvimento do aplicativo foi o *Android Studio* (DEVELOPERS, 2019a).

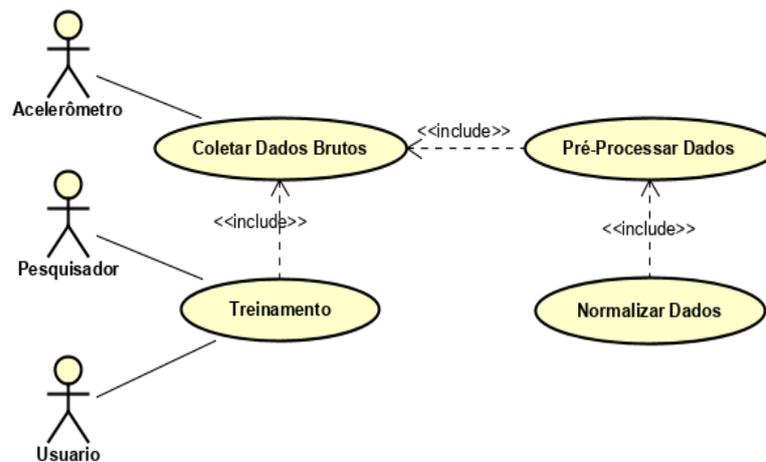


Figura 8: Diagrama de Caso de Uso Treinamento

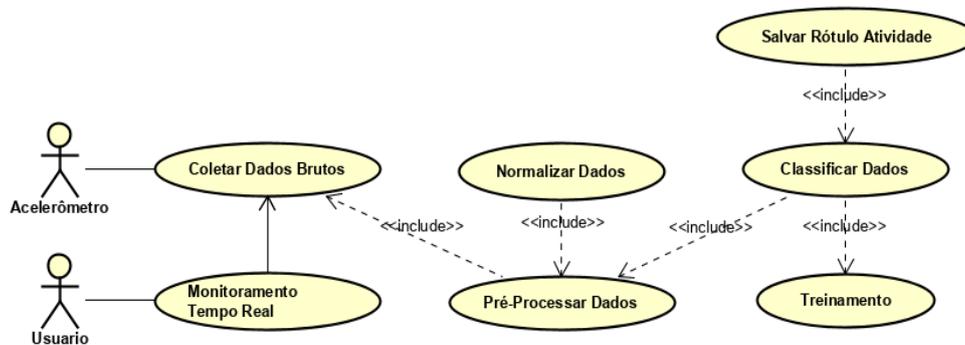


Figura 9: Diagrama de Caso de Uso Monitoramento

As próximas seções descrevem como foi implementado cada uma das etapas do sistema e as telas relacionadas a cada funcionalidade.

3.3.1 Coleta de dados

A coleta de dados consiste em usar o sensor de acelerômetro do *smartphone* do usuário para coletar a aceleração nos eixos X, Y e Z e o *timestamp*. O *timestamp* de um acelerômetro é coletado como a quantidade de segundos desde primeiro de janeiro de mil novecentos e setenta em Londres, na Inglaterra.

Para a coleta de dados, o sistema verifica se existe um sensor de acelerômetro no *smartphone* do usuário. Se ele não existir, uma mensagem de alerta é emitido e os próximos passos do fluxo não são executados. Ao finalizar a coleta de dados, é gerado um arquivo de texto com os dados brutos. Essa base de dados será identificada por Base de Dados Brutos.

A Figura 10 mostra a tela do sistema responsável por coletar os dados brutos. Para iniciar a coleta de dados o usuário deve clicar no botão 'Gravar' e para finalizar a

coleta de dados brutos o usuário deve clicar no botão 'Exportar'. Essa funcionalidade é destinada ao usuário pesquisador da área de RAH.



Figura 10: Tela para coleta de dados do APP

3.3.2 Pré-processamento

O pré-processamento de dados consiste em extrair atributos dos dados brutos coletados do acelerômetro. A extração de atributos é feita apenas usando os atributos X, Y e Z.

Para extrair os atributos são usadas janelas de dados. Existem dois tipos de janelas (1) fixa: tem pré-definido o tamanho da janela N e cada bloco de dados tem N registros; (2) deslizante: tem pré-definido o tamanho da janela N , pré-definido uma sobreposição S e cada bloco de dados tem N registros, porém a partir do segundo bloco a leitura de dados acontece a cada $(K * N) - S$ registros, onde K é o número da janela menos 1. A Figura 11 ilustra os dois tipos de janelas. O *Activity Recording* irá usar por padrão a janela fixa de tamanho 200.

As medidas desvio padrão, maior, menor, média, desvio médio absoluto, variância, energia, correlação, binomial e aceleração média são, em geral, os atributos mais utilizados na literatura de RAH. Este trabalho optou por usar estas medidas como os atributos a serem extraídos dos dados brutos, na fase de pré-processamento dos dados.

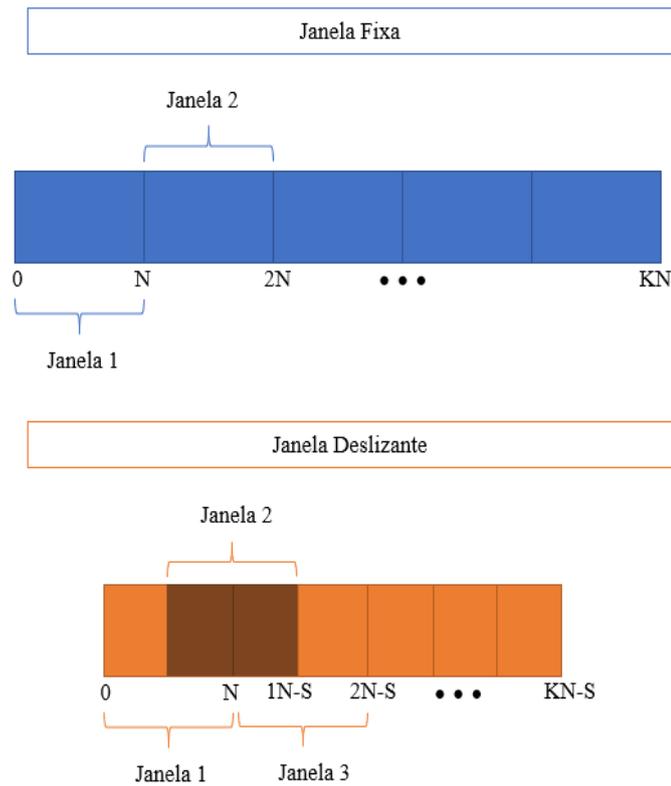


Figura 11: Tipos Janela

No fim do pré-processamento, os dados são armazenados em um arquivo do tipo ARFF. A tabela 1 detalha as fórmulas utilizadas para o cálculo de cada atributo na etapa de pré-processamento.

Após o pré-processamento dos dados, eles são re-escalados (processo aqui denominado de normalização), ou seja, todos os dados de cada atributo do arquivo são convertidos em um intervalo entre 0 e 1. A equação 3.1 é utilizada para normalizar o arquivo, onde x é a instância a ser normalizada, $min(x)$ é o menor valor possível do atributo e $max(x)$ é o maior valor possível do atributo.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \quad (3.1)$$

O maior valor do sensor de acelerômetro é extraído da função *getMaximumRange()* (DEVELOPERS, 2019b) e o menor valor é extraído da função *getMaximumRange()* multiplicado por -1. Essa função determina o alcance máximo do sensor de acelerômetro do *smartphone* e quando multiplicada por -1 determina o alcance mínimo do sensor.

O cálculo do maior e menor valor do atributo é feito utilizando o maior e menor valor do sensor de acelerômetro, ou seja, para cada fórmula apresentada na tabela 1 é feito um cálculo utilizando o maior e menor valor do sensor de acelerômetro. Logo, para

Tabela 1: Fórmula Atributos

Atributo	Fórmula
Desvio Padrão	$S = \sqrt{\frac{\sum_{i=1}^n (x_i - media)^2}{n-1}}$
Maior	$if(x > maior) maior = x;$
Menor	$if(x < menor) menor = x;$
Média	$x = \frac{\sum x_i}{n}$
Desvio Médio Absoluto	$\frac{1}{n} \sum_{i=1}^n (x_i - media(x))$
Variância	$S^2 = \frac{\sum_{i=1}^n (x_i - media)^2}{n-1}$
Energia	$E = \frac{\sum_{i=1}^n x_i^2 }{n}$
Correlação	$corr(x, y) = \frac{cov(x,y)}{\Theta_x \Theta_y}$
Binomial	$B = \frac{maior - menor}{10}$
Aceleração Média	$AcM = \frac{\sqrt{(x_i^2 + y_i^2 + z_i^2)}}{n}$

descobrir o maior valor possível da média, por exemplo, o cálculo apresentado na equação 3.2 é feito.

$$MaiorMedia = \frac{\sum getMaximumRange()}{n} \quad (3.2)$$

Ao finalizar a normalização do arquivo pré-processado um novo arquivo do tipo ARFF é gerado com os dados pré-processados normalizados.

A Figura 12 mostra a tela do sistema responsável por pré-processar o arquivo de dados brutos coletados e normalizá-los. Essa funcionalidade é destinada ao usuário pesquisador da área de RAH.

3.3.3 Treinamento de Dados

Em geral, a tarefa de classificar dados exige o treinamento de um classificador, ou seja, a criação de um modelo de decisão à partir de uma base de dados rotulada, chamada de base de treinamento. O aplicativo proposto neste trabalho utiliza para a classificação de dados o algoritmo KNN, que pertence à classe dos algoritmos *lazy*, que trabalha diretamente sobre os dados, não necessitando da criação de um modelo. Ele exige apenas que uma base de dados de treinamento esteja disponível, ou seja, uma amostra dos



Figura 12: Pré-processamento

dados com seus respectivos rótulos (classes) para a classificação de uma nova instância.

Para a criação dessa amostra de treinamento o APP disponibiliza uma função chamada Treinamento. Ao usar essa funcionalidade, o usuário escolhe uma atividade, para a qual ele deseja treinar o classificador e desempenha essa atividade por três minutos. O treinamento faz a coleta de dados brutos, pré-processamento e normalização dos dados utilizando o rótulo informado pelo usuário. Esses dados coletados são armazenados como instâncias do treino. O usuário pode treinar o classificador para diferentes atividades.

Cada atividade treinada será armazenada em um arquivo diferente com o mesmo sufixo, por exemplo *andando_training_normalize_file.arff*. Os arquivos gerados no treino serão utilizados para classificação das atividades do usuário. Se o usuário treinar o sistema com uma atividade já treinada, o arquivo de treinamento da atividade será sobrescrito, ou seja, o treinamento anterior será substituído pelo atual.

A captura dos dados brutos do treinamento é iniciada após dez segundos de início do cronômetro e finalizada dez segundos antes de fim do cronômetro. Essa medida foi adotada para não coletar os dados de início e fim da atividade, que poderiam atrapalhar no treinamento do classificador.

A Figura 13 mostra a tela do APP responsável por treinar o sistema. Essa funcionalidade é destinada para qualquer usuário.

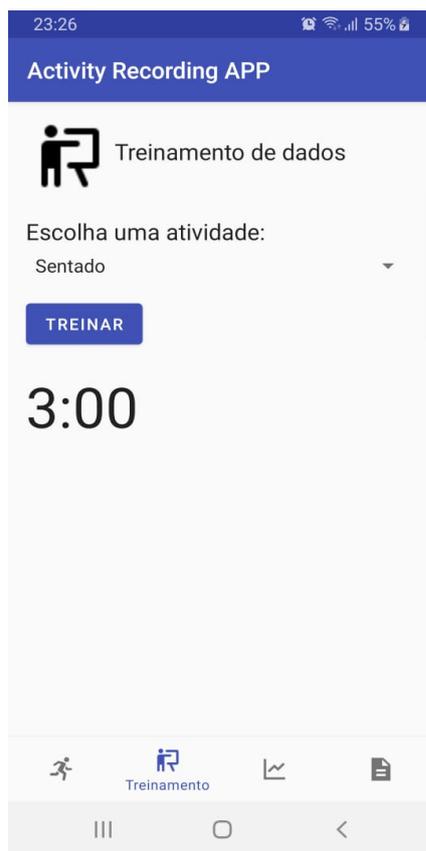


Figura 13: Treinamento

3.3.4 Classificação de Dados

O algoritmo de classificação de dados no *Activity Recording* é o KNN. A Figura 5 ilustra o funcionamento do KNN. Esse algoritmo de classificação foi escolhido, porque de acordo com os experimentos feitos nos trabalhos (MARTINS, 2018) e (GIACOMELLI, 2018b), o algoritmo KNN foi o que obteve melhores resultados na maioria dos testes de classificação de dados para o uso de acelerômetros no cenário de Reconhecimento de Atividade Humana.

A classificação utiliza os arquivos de treinamento de dados e o arquivo dos dados pré-processados normalizados.

Para classificação, o sistema faz a leitura de todos os dados da base de treino B e guarda na memória e depois faz a leitura linha a linha da base de teste. Para cada instância A da base de teste é feito o cálculo da distância euclidiana de A com todos os dados da base de treino B . Esses dados são guardados em uma lista de objetos L que contém a distância Euclidiana e o rótulo da instância de B .

Depois disso, a lista L é ordenada e o algoritmo seleciona o rótulo da atividade predominantemente entre os K primeiros itens da lista, ou seja, seleciona os K rótulos que possuem a menor distância Euclidiana com A , depois escolhe o rótulo predominante e

classifica a instância *A*. Esse passo se repete até classificar todas as instâncias da base de teste.

O rótulo e seus respectivos dados pré-processados normalizados são armazenados em um arquivo do tipo ARFF. Esse arquivo será usado para sumariar os dados. Veja a subseção 3.3.6.

A Figura 14 mostra a tela do APP responsável por classificar os dados coletados na etapa de pré-processamento. Essa funcionalidade é destinada para o usuário pesquisador da área de RAH.



Figura 14: Classificação de Dados

3.3.5 Monitoramento em Tempo Real

O monitoramento em tempo real faz todas as etapas de RAH em tempo real, ou seja, faz a coleta de dados brutos utilizando o acelerômetro, depois pré-processa e normaliza esses dados, posteriormente classifica o dado e por fim, armazena a classificação do dado em um banco de dados. Essa informação será usada para sumariação dos dados.

Para que o monitoramento ocorra é necessário pressionar o botão 'Iniciar', veja Figura 15. Assim que o botão for pressionado, o sistema inicia a coleta de dados brutos até que a quantidade de linhas coletadas seja igual a 200, ou seja, até que a quantidade de

linhas coletadas seja igual ao tamanho da janela utilizada. Após a coleta das 200 linhas de dados brutos, o sistema irá pré-processar e normalizar esses dados gerando uma instância I .

O próximo passo é fazer a leitura de todos os arquivos de treinamento e guardar na memória, ou seja, guarda na memória as instâncias rotuladas R . Após isso, o cálculo da distância euclidiana de I com todas as instâncias R são armazenadas em uma lista que posteriormente será ordenada.

Por fim, o algoritmo seleciona o rótulo da atividade predominantemente entre os K primeiros itens da lista, ou seja, seleciona os K rótulos que possuem a menor distância euclidiana com I , escolhe o rótulo predominante e classifica a instância I . O rótulo da instância I é armazenado no banco de dados do *smartphone* do usuário. Esses passos descritos são feitos até que o usuário selecione o botão 'Finalizar', veja Figura 16.

O banco de dados *SQLite* é utilizado pelo APP e o modelo lógico do APP é composto por uma tabela chamada de *activity_time*. Essa tabela possui as seguintes colunas: (1) *id* que é gerado de forma incremental, (2) *dt_inicial* é a data de início da atividade, (3) *dt_final* representa a data de fim da atividade e (4) *activity* que é o rótulo da atividade. Para cada atividade identificada é verificado se a última atividade gravada no banco é igual a atual. Se elas forem iguais é feito a atualização da data de fim da atividade. Se elas forem diferentes um novo registro é adicionado na tabela.

Os dados armazenados no banco de dados serão utilizados para a sumariação, que possibilita que o usuário veja quais atividades foram desempenhadas ao longo do tempo. Veja a seção 3.3.6.

3.3.6 Sumariação de dados

A sumariação dos dados consiste em apresentar ao usuário quais atividades foram classificadas pelo sistema em um gráfico. O gráfico apresenta a atividade e a quantidade de minutos que aquela atividade foi desempenhada.

No APP é possível visualizar a sumariação das atividades na tela voltada para o usuário pesquisador. Os dados mostrados nesse gráfico estão no arquivo que foi gerado na etapa de classificação. Para visualizar o gráfico das atividades classificadas o usuário deve pressionar o botão 'Gráfico'. A Figura 17 mostra as telas dessa funcionalidade.

Também é possível visualizar a sumariação das atividades que foram desenvolvidas ao longo do tempo pelo usuário. Os dados mostrados nesse gráfico foram obtidos no monitoramento em tempo real e estão salvos no banco de dados do *smartphone*. O usuário pode visualizar as atividades que foram feitas no dia, na semana, no mês ou todas as atividades que já foram desempenhadas. A Figura 18 mostra essa tela.

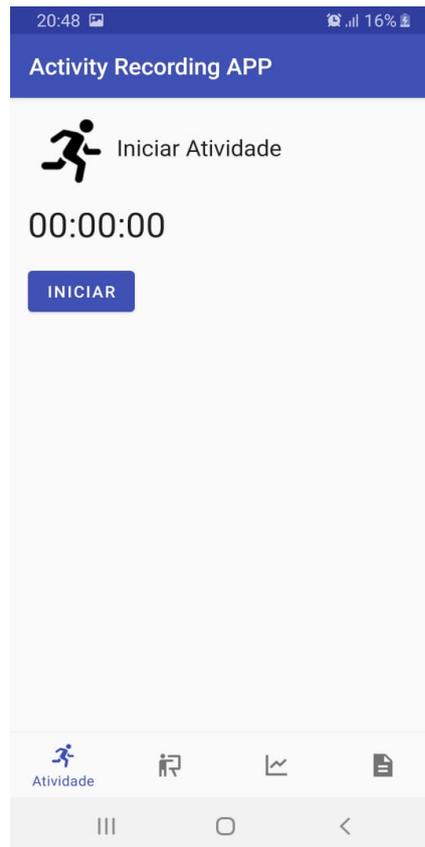


Figura 15: Monitoramento em Tempo Real

Para sumariar os dados obtidos no monitoramento em tempo real a consulta apresentada no trecho de código 3.1 é feita no banco de dados. A função *JULIANDAY* é uma função do *SQLite* utilizada para fazer cálculos com data. A variável *@day* é utilizada para selecionar as atividades que foram desempenhadas em um determinado período de tempo. Por exemplo, se o usuário desejar visualizar as atividades desempenhadas na semana, o valor da variável *@day* seria sete. A consulta 3.1 tem como resultado as atividades e a quantidade de segundos que cada atividade foi desempenhada. Os segundos são transformados em minutos na apresentação do gráfico.

Listing 3.1: Consulta Utilizada para Sumariação

```

1  SELECT
2  A.activity,
3  SUM(CAST((
4      (JULIANDAY(CAST(A.dt_final AS VARCHAR(30)))
5      - JULIANDAY(CAST(A.dt_inicial AS VARCHAR(30))))
6      * 24 * 3600) AS INTEGER)) AS seconds
7  FROM ACTIVITY_TIME A
8  WHERE JULIANDAY(CAST(A.dt_inicial AS VARCHAR(30)))
9      BETWEEN JULIANDAY('now', '-' + @day + ' days')
10      AND JULIANDAY('now')
11  GROUP BY A.activity;

```



Figura 16: Finalizar Monitoramento em Tempo Real

3.4 Testando o Aplicativo

Essa seção descreve um cenário real de uso do aplicativo *Activity Recording*. Para o teste do aplicativo as funcionalidades de treinamento e monitoramento em tempo real foram utilizadas.

Para que a classificação da atividade ocorra o aplicativo precisa ser treinado. A tela apresentada na figura 13 foi utilizada para o treinamento e as atividades treinadas foram andando, sentado e em pé.

Após o treinamento o monitoramento em tempo real foi iniciado. A figura 15 mostra a tela do monitoramento. As atividades desenvolvidas no monitoramento em tempo real foram andando e sentado. Essas duas atividades foram realizadas por 6 minutos, sendo 3 minutos andando e 3 minutos sentado.

A figura 19 mostra o resultado da sumariação das atividades desenvolvidas. O aplicativo reconheceu a atividade sentado por 2,48 minutos, andando por 1,83 minutos e em pé por 0,32 segundos. Apesar da atividade em pé não ter sido desempenhada, ela é próxima da atividade andando, portanto houve um desvio da classificação da atividade andando para em pé.

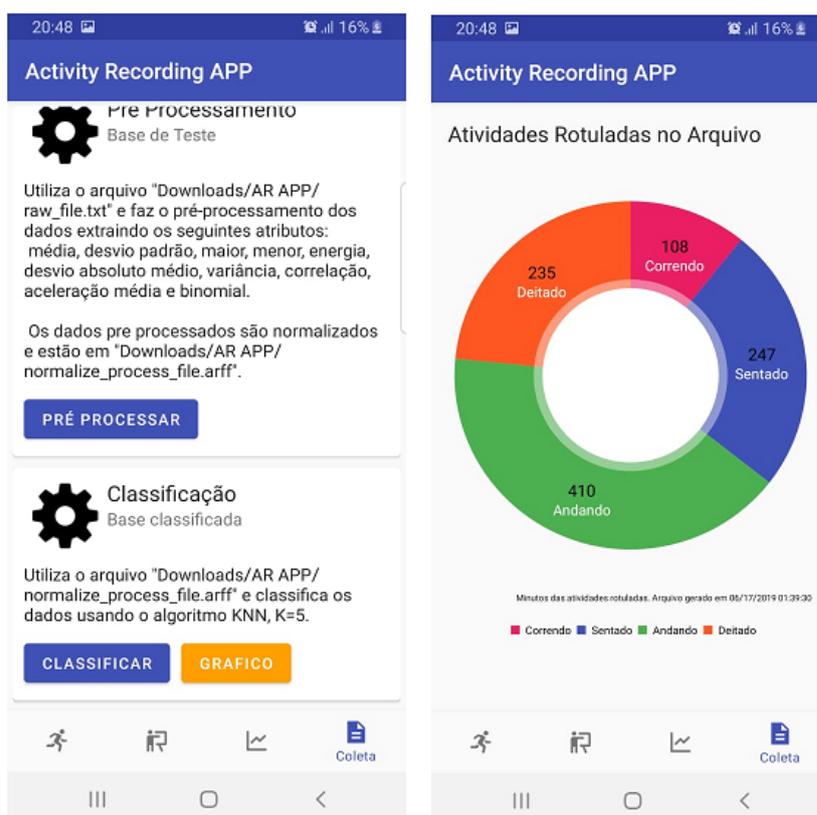


Figura 17: Sumariação Atividades Arquivo

3.5 Considerações Finais

Nesse capítulo foi descrito o funcionamento geral e técnico do aplicativo *Activity Recording*. Além disso, mostrou a interface e a utilização do sistema.

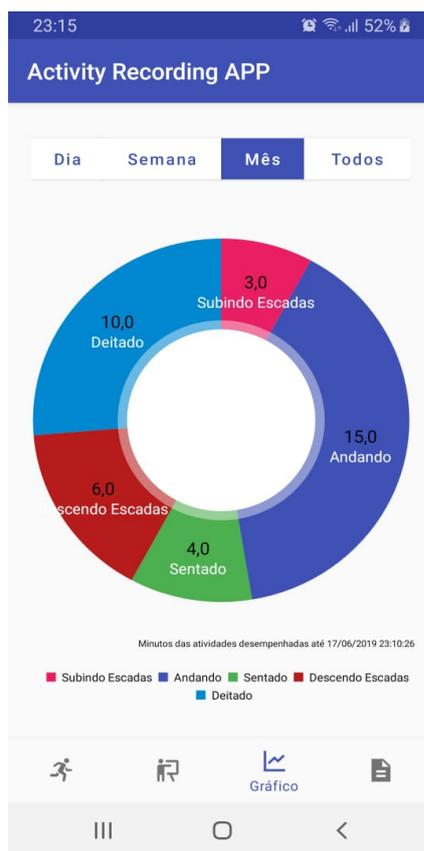


Figura 18: Sumariação Atividades do Monitoramento

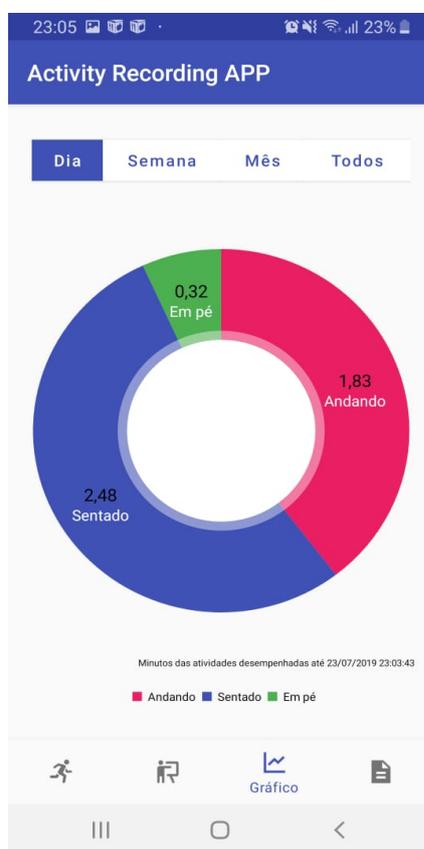


Figura 19: Teste Monitoramento Em Tempo Real

4 Conclusão

O monitoramento de atividades pode incentivar as pessoas a praticarem mais atividades físicas e conseqüentemente a melhorarem a qualidade de vida. Um fator crucial para o monitoramento de atividades é o estudo do RAH, que consiste em identificar quais atividades foram desempenhadas por uma pessoa.

Utilizando as bases teóricas apresentadas neste trabalho e presentes na literatura, o aplicativo *Activity Recording* foi desenvolvido para monitorar as atividades dos usuários e ajudar os pesquisadores da área a criarem bases de dados personalizadas e executar as etapas de RAH separadamente.

O desenvolvimento do aplicativo foi concluído com sucesso. Ele contempla todas as etapas de um sistema de RAH. Além disso, ele pode ser usado tanto por usuários finais, que pretendem monitorar suas atividades, quanto por usuários pesquisadores, que pretendem estudar técnicas e algoritmos para RAH. O aplicativo encontra-se em fase final de testes e pode ser, em breve, disponibilizado em serviços de distribuição digital de aplicativos, tais como o *Google Play*.

Os trabalhos futuros a serem feitos dizem respeito a (i) melhoria da implementação do algoritmo de classificação KNN; (ii) publicação do aplicativo na *Google Play*; (iii) desenvolvimento do aplicativo para outros sistemas operacionais; (iv) implementação de outros algoritmos de classificação; (v) melhoria da interface gráfica do aplicativo; (vi) estudo sobre migrar o processamento do aplicativo para um servidor na nuvem; (vii) geração de mais relatórios e estudo sobre outros gráficos para sumariação dos dados.

Referências

ABDALLAH, Z. S.; GABER, M. M.; SRINIVASAN, B.; KRISHNASWAMY, S. Anyonovel: detection of novel concepts in evolving data streams. *Evolving Systems*, Springer, v. 7, n. 2, p. 73–93, 2016. Citado 2 vezes nas páginas 22 e 23.

APPLE. *Uma visão mais completa da sua saúde*. 2018. Disponível em: <<https://www.apple.com/br/ios/health/>>. Acesso em: 10/10/2018. Citado na página 24.

BAYAT, A.; POMPLUN, M.; TRAN, D. A. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, Elsevier, v. 34, p. 450–457, 2014. Citado 2 vezes nas páginas 11 e 12.

BIFET, A.; HOLMES, G.; KIRKBY, R.; PFAHRINGER, B. MOA: massive online analysis. *Journal of Machine Learning Research*, v. 11, p. 1601–1604, 2010. Disponível em: <<http://portal.acm.org/citation.cfm?id=1859903>>. Citado na página 18.

BRASIL, V. *Hábitos dos brasileiros impactam no crescimento da obesidade e aumenta prevalência de diabetes e hipertensão*. 2016. Disponível em: <<http://portalarquivos.saude.gov.br/images/pdf/2017/abril/17/Vigitel.pdf>>. Acesso em: 02/07/2019. Citado na página 15.

BUDZINSKI, J. A. S. *Acelerômetros/Aplicações*. 1999. Disponível em: <<http://www-eletrica.ufpr.br/edu/Sensores/1999/joao/aplicacoes.htm>>. Acesso em: 01/10/2018. Citado na página 16.

BULLING, A.; BLANKE, U.; SCHIELE, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, ACM, v. 46, n. 3, p. 33, 2014. Citado na página 17.

CORREA, S. M. B. B. Probabilidade e estatística. In: _____. 2ª. ed. [S.l.]: PUC Minas Virtual, 2003. Citado na página 18.

DELUNARDO, C. S.; RIBEIRO, P. F. E. Educação física escolar, atividade física e qualidade de vida: a visão de alunos universitários. *Cadernos UniFOA*, v. 12, n. 35, p. 111–122, 2018. Citado na página 11.

DEVELOPERS, G. *Instalar o Android Studio*. 2019. Disponível em: <<https://developer.android.com/studio/install?hl=pt-br>>. Acesso em: 20/05/2019. Citado na página 28.

DEVELOPERS, G. *Sensor*. 2019. Disponível em: <<https://developer.android.com/reference/android/hardware/Sensor>>. Acesso em: 17/06/2019. Citado na página 31.

ELECTRONICS, S. *Samsung Health: Sua saúde em suas mãos*. 2018. Disponível em: <<https://www.samsung.com/br/apps/samsung-health/>>. Acesso em: 01/10/2018. Citado 2 vezes nas páginas 16 e 24.

- GARCÍA-MARTÍN, E. *Extraction and Energy Efficient Processing of Streaming Data*. Tese (Doutorado) — Blekinge Tekniska Högskola, 2017. Citado 2 vezes nas páginas 20 e 21.
- GIACOMELLI, D. F. *Estudo e caracterização das ferramentas para o reconhecimento de atividade humana usando sensores de acelerômetro*. 45 p. Monografia (Trabalho de Conclusão de Curso) — Universidade Federal de Uberlândia, Uberlândia, 2018. Citado 2 vezes nas páginas 12 e 22.
- GIACOMELLI, E. R. d. F. D. Estudo e caracterização das principais ferramentas para reconhecimento de atividade humana usando sensores de acelerômetro. In: *Simpósio Brasileiro de Sistemas de Informação*. [S.l.: s.n.], 2018. Citado 2 vezes nas páginas 17 e 34.
- GOOGLE. *Google Fit: Coaching you to a healthier and more active life Get the app*. 2018. Disponível em: <<https://www.google.com/fit/>>. Acesso em: 10/10/2018. Citado na página 24.
- GORDILHO, R. *Médicos precisam começar a recomendar apps para saúde e fitness*. 2014. Disponível em: <<https://saudebusiness.com/noticias/medicos-precisam-comecar-recomendar-apps-para-saude-e-fitness/>>. Acesso em: 28/09/2018. Citado na página 16.
- HA, S.; CHOI, S. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In: IEEE. *2016 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2016. p. 381–388. Citado na página 17.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA data mining software: An update. *SIGKDD Explorations*, v. 11, n. 1, p. 10–18, 2009. Citado na página 18.
- IBGE. *Prática de esportes e atividade física*. 2015. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv100364.pdf>>. Acesso em: 02/07/2019. Citado na página 11.
- IGLESIAS, J. A.; ANGELOV, P.; LEDEZMA, A.; SANCHIS, A. Human activity recognition based on evolving fuzzy systems. *International journal of neural systems*, World Scientific, v. 20, n. 05, p. 355–364, 2010. Citado na página 22.
- KWAPISZ, J. R.; WEISS, G. M.; MOORE, S. A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, ACM, v. 12, n. 2, p. 74–82, 2011. Citado 2 vezes nas páginas 17 e 21.
- MARTINS, E. R. d. F. L. M. *Estudo Comparativo de Algoritmos de Classificação para Fluxos Contínuos de Dados Aplicados ao Problema de Reconhecimento de Atividade Humana*. 2018. Disponível em: <http://www.propp.ufu.br/sites/propp.ufu.br/files/conteudo/acontece_na_pro_reitoria/vii_semana_de_inicio_cientifica_e_tecnologica_-_parte_5.pdf>. Acesso em: 12/01/2019. Citado 2 vezes nas páginas 12 e 34.
- NGUYEN, L. T.; ZENG, M.; TAGUE, P.; ZHANG, J. Recognizing new activities with limited training data. In: ACM. *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. [S.l.], 2015. p. 67–74. Citado 2 vezes nas páginas 22 e 23.

- RAVI, N.; DANDEKAR, N.; MYSORE, P.; LITTMAN, M. L. Activity recognition from accelerometer data. v. 5, n. 2005, p. 1541–1546, 2005. Citado 3 vezes nas páginas 11, 12 e 21.
- REPORTER, E. H. *Mobiquity Study: 70 Percent of People Track their Health and Fitness Daily with Mobile Apps*. 2014. Acesso em: 28/09/2018. Disponível em: <<https://electronichealthreporter.com/mobiquity-study-70-percent-of-people-track-their-health-and-fitness-daily-with-mobile-apps/>>. Citado na página 15.
- RIBEIRO, L. *O que é UML e Diagramas de Caso de Uso: Introdução Prática à UML*. 2012. Disponível em: <<https://www.devmedia.com.br/o-que-e-uml-e-diagramas-de-caso-de-uso-introducao-pratica-a-uml/23408>>. Acesso em: 18/07/2019. Citado na página 27.
- RUNTASTIC. *Runtastic*. 2018. Disponível em: <<https://www.runtastic.com/pt-br/>>. Acesso em: 01/10/2018. Citado na página 16.
- SAUDE, O. M. da. *Atividade Física*. 2014. Disponível em: <http://actbr.org.br/uploads/arquivo/957_FactSheetAtividadeFisicaOMS2014_port_REV1.pdf>. Acesso em: 02/07/2019. Citado na página 11.
- SILVA, A. V. da. *Monitoramento de atividade física com acelerômetro em cães da raça Border Collie de diferentes ambientes*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, Março 2015. Citado na página 11.
- SILVA, E.; BOTELHO, L.; SANTOS, I.; SANCHEZ, G. Computação ubíqua-definição e exemplos. *Revista de Empreendedorismo, Inovação e Tecnologia*, v. 2, n. 1, p. 23–32, 2015. Citado 3 vezes nas páginas 6, 14 e 15.
- STARNER, T. E. Wearable computers: No longer science fiction. *IEEE pervasive computing*, IEEE, v. 1, n. 1, p. 86–88, 2002. Citado na página 14.
- SUN, L.; ZHANG, D.; LI, B.; GUO, B.; LI, S. Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations. In: SPRINGER. *International conference on ubiquitous intelligence and computing*. [S.l.], 2010. p. 548–562. Citado 4 vezes nas páginas 11, 12, 16 e 22.
- TAN, P.-N.; STEINBACH, M.; KARPATNE, A.; KUMAR, V. *Introduction to Data Mining, (Second Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2013. ISBN 0133128903. Citado 2 vezes nas páginas 18 e 19.
- TECHNOLOGIES, S. T. *Power your passion for sports*. 2018. Disponível em: <<https://www.sports-tracker.com/>>. Acesso em: 13/10/2018. Citado na página 24.
- TWOMEY, N.; DIETHE, T.; FAFOUTIS, X.; ELSTS, A.; MCCONVILLE, R.; FLACH, P. A.; CRADDOCK, I. A comprehensive study of activity recognition using accelerometers. *Informatics*, v. 5, n. 2, p. 27, 2018. Citado 2 vezes nas páginas 12 e 23.
- WANG, J.; CHEN, Y.; HAO, S.; PENG, X.; HU, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, Elsevier, v. 119, p. 3–11, 2019. Citado na página 17.

- WEISS, G. M.; LOCKHART, J. W.; PULICKAL, T. T.; MCHUGH, P. T.; RONAN, I. H.; TIMKO, J. L. Actitracker: a smartphone-based activity recognition system for improving health and well-being. In: IEEE. *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on*. [S.l.], 2016. p. 682–688. Citado na página 12.
- WIKI, W. *ARFF Format*. 2018. Disponível em: <<https://waikato.github.io/weka-wiki/arff/>>. Acesso em: 29/11/2019. Citado na página 18.
- ZHAO, R.; WANG, J. Visualizing the research on pervasive and ubiquitous computing. *Scientometrics*, Springer, v. 86, n. 3, p. 593–612, 2011. Citado na página 14.