

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

UM SISTEMA DE LOCALIZAÇÃO *INDOOR*
USANDO *FINGERPRINTING* E DETECÇÃO DE
NOVIDADES PARA AVALIAÇÃO DE
CONFIANÇA

HELMER AUGUSTO DE SOUZA MOURÃO

UM SISTEMA DE LOCALIZAÇÃO *INDOOR*
USANDO *FINGERPRINTING* E DETECÇÃO DE
NOVIDADES PARA AVALIAÇÃO DE
CONFIANÇA

Tese apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas, Campus Universitário Senador Arthur Virgílio Filho, como requisito parcial para a obtenção do grau de Doutor em Informática.

ORIENTADOR: HORÁCIO ANTONIO BRAGA FERNANDES DE OLIVEIRA

Manaus - AM

Outubro de 2022

Ficha Catalográfica

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

M929s Mourão, Helmer Augusto de Souza
Um sistema de localização indoor usando fingerprinting e
detecção de novidades para avaliação de confiança / Helmer
Augusto de Souza Mourão . 2022
72 f.: il. color; 31 cm.

Orientador: Horácio Antonio Braga Fernandes de Oliveira
Tese (Doutorado em Informática) - Universidade Federal do
Amazonas.

1. Localização interna. 2. Sistema de posicionamento. 3.
Impressão digital. 4. Detecção de novidade. I. Oliveira, Horácio
Antonio Braga Fernandes de. 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

**"Um Sistema de Localização Indoor usando
Fingerprinting e Detecção de Novidades para Avaliação
de Confiança"**

Helmer Augusto de Souza Mourão

Tese de Doutorado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof. Horácio Antônio Braga Fernandes de Oliveira - PRESIDENTE

Prof. Felipe Leite Lobo- MEMBRO EXTERNO

Prof. Moysés Mendes Lima - MEMBRO EXTERNO

Prof. Eduardo James Pereira Souto - MEMBRO INTERNO

Prof. Raimundo da Silva Barreto - MEMBRO INTERNO

Manaus, 20 de outubro de 2022

Resumo

Os sistemas de localização para ambientes fechados são usados para localizar dispositivos móveis dentro de edifícios onde as soluções tradicionais, como o *Global Navigation Satellite Systems* (GNSS), não funcionam bem devido à falta de visibilidade direta dos satélites. O *fingerprinting* é uma das soluções mais conhecidas e precisas para localização em ambientes fechados, ele é dividido em duas fases distintas uma fase de treinamento (*Offline*) e uma de localização (*Online*). Uma das informações mais utilizadas é o *Received Signal Strength Indicator* (RSSI), por ser de fácil obtenção, porém valores de RSSI são conhecidos por serem instáveis e ruidosos devido aos obstáculos e à dinamicidade dos cenários, causando imprecisões nas estimativas de posição. Devido a esta característica do RSSI, várias metodologias têm sido propostas para tentar mitigar os efeitos do ruído na fase de treinamento, porém, as informações do RSSI também apresentam ruído na fase de localização. Esse ruído geralmente faz com que o sistema indique um local que não tem certeza se está correto, embora seja o mais provável com base em seus cálculos. Para minimizar este problema, este trabalho apresenta alguns métodos de verificação do nível de confiança das classificações usando as probabilidades da classificação aliado a algoritmos de Detecção de Novidade (*Novelty Detection*). Assim, neste trabalho, propomos LocFiND (*Localization using Fingerprinting and Novelty Detection*), uma solução baseada em *fingerprinting* que utiliza a detecção de novidade para avaliar a confiança das posições estimadas e, assim, tentar mitigar o ruído causado pelo RSSI na localização. Estimativas não confiáveis são descartadas e não encaminhadas à aplicação. Nossa solução foi avaliada usando informações coletadas em ambiente real de uma área escolar usando dispositivos baseados em *Bluetooth*. Nossa avaliação de desempenho mostra uma melhoria considerável na precisão e estabilidade da localização, descartando apenas algumas estimativas não confiáveis.

Abstract

Indoor localization systems are used to locate mobile devices inside buildings where traditional solutions such as Global Navigation Satellite Systems (GNSS) do not work well due to the lack of direct visibility from satellites. Fingerprinting is one of the most known and accurate solutions for indoor localization, it is divided into two distinct phases: a training phase (Offline) and a localization phase (Online). One of the most used information is the Received Signal Strength Indicator (RSSI), as it is easy to obtain, but RSSI values are known to be unstable and noisy due to obstacles and the dynamics of the scenarios, causing inaccuracies in the estimates of position. Due to this RSSI characteristic, several methodologies have been proposed to try to mitigate the noise effects in the training phase, however, the RSSI information also presents noise in the localization phase. This noise often causes the system to point to a location that it is not sure is correct, even though it is most likely based on its calculations. To minimize this problem, this work presents some methods to verify the confidence level of classifications using classification probabilities combined with novelty detection algorithms. Thus, in this work, we propose LocFiND (*Localization using Fingerprinting and Novelty Detection*), a solution based on fingerprinting that uses novelty detection to assess the confidence of the estimated positions and, thus, try to mitigate the noise caused by RSSI on localization. Unreliable estimations are discarded and not forwarded to the application. We implemented our solution in a real-world, large-scale school area using Bluetooth-based devices. Our performance evaluation shows considerable improvement in the localization accuracy and stability while discarding only a few, unreliable estimations.

Lista de Figuras

2.1	Relação entre sinais de localização, medições, métodos.	6
2.2	Localização por Tempo de Chegada.	15
2.3	Localização por Ângulo de Chegada.	17
2.4	Localização por Proximidade.	20
4.1	Arquitetura do LocFiND: a fase de treinamento gera o banco de dados de impressão digital, que é utilizado na fase online para estimar a posição dos dispositivos móveis.	40
4.2	Avaliação de confiança usando <i>Probabilidade de Predição</i> : classificações com baixas probabilidades são descartadas para melhorar a precisão.	42
4.3	Detecção de novidade usando One-Class SVM.	43
4.4	Avaliação de confiança usando a técnica <i>Detecção de Novidade</i> : a precisão é melhorada com o descarte de classificações não confiáveis.	44
4.5	Avaliação de confiança usando a <i>Abordagem Híbrida</i> : um meio-termo entre a precisão e o número de estimativas de posição.	45
5.1	Mapa do nosso testbed: 11 salas, 3 corredores, 15 nós âncora e 150 pontos de referência. Para cada ponto de referência, 100 amostras foram coletadas em um total de 15.000 amostras para cada dispositivo móvel. Foram usados 11 dispositivos móveis diferentes, 8 para treinamento e 3 para teste.	48
5.2	Dispositivos do tipo <i>smartwatch</i> baseados em Bluetooth que foram usados como dispositivos móveis.	48
5.3	Os pontos de acesso baseados em Bluetooth que foram usados como nós âncora.	49
5.4	Precisão da localização de ponto usando <i>probabilidade de previsão</i>	50
5.5	Erro médio da localização usando <i>probabilidade de previsão</i>	50
5.6	Precisão da localização de sala usando <i>probabilidade de previsão</i>	50
5.7	Função de Distribuição Acumulada do erro de localização usando <i>probabilidade de previsão</i> com limites $\alpha = 25\%$, 45% e 75% : mostra a porcentagem de estimativas (eixo y) com um erro menor do que o eixo x.	51

5.8	Precisão da localização de ponto usando <i>Detecção de Novidade</i>	53
5.9	Erro médio da localização usando <i>Detecção de Novidade</i>	53
5.10	Precisão da localização de sala usando <i>Detecção de Novidade</i>	53
5.11	Função de Distribuição Acumulada do erro de localização usando <i>Detecção de Novidade</i> com limites $\alpha = 25\%$, 45% e 75% : uma porcentagem maior das estimativas de posição resultou em erros menores.	54
5.12	Precisão da localização de ponto usando <i>Abordagem Híbrida</i>	55
5.13	Erro médio da localização usando <i>Abordagem Híbrida</i>	55
5.14	Precisão da localização de sala usando <i>Abordagem Híbrida</i>	56
5.15	Função de Distribuição Acumulada do erro de localização usando nossas três abordagens usando como menor probabilidade e corte 25%	56
5.16	Função de Distribuição Acumulada do erro de localização usando nossas três abordagens usando como menor probabilidade e corte 45%	57
5.17	Função de Distribuição Acumulada do erro de localização usando nossas três abordagens usando como menor probabilidade e corte 70%	57

Lista de Tabelas

2.1	Algumas das principais tecnologias usadas em localização para ambientes fechados.	12
2.2	Algumas das principais medidas usadas em localização para ambientes fechados.	18
2.3	Alguns dos principais métodos usadas em localização para ambientes fechados.	23

Sumário

Resumo	v
Abstract	vii
Lista de Figuras	viii
Lista de Tabelas	x
1 Introdução	1
1.1 Motivação	2
1.2 Objetivos	3
1.3 Organização do Documento	3
2 Fundamentação	5
2.1 Sinais de Comunicação Sem Fio Usados para Localização	6
2.1.1 WiFi	7
2.1.2 RFID	7
2.1.3 Infra-vermelho (IR)	8
2.1.4 Ultra Wideband (UWB)	9
2.1.5 Bluetooth LE (BLE)	9
2.1.6 Zigbee	10
2.1.7 Long Range (LoRa)	11
2.1.8 Considerações	11
2.2 Medidas	13
2.2.1 Intensidade do Sinal Recebido (RSSI)	13
2.2.2 Informações de Status do Canal (CSI)	14
2.2.3 Tempo de Chegada (ToA)	14
2.2.4 Diferença Entre os Tempos de Chegada (TDoA)	15
2.2.5 Ângulo de Chegada (AoA)	16
2.2.6 Considerações	17

2.3	Métodos Usados para Localizar	19
2.3.1	Proximidade	19
2.3.2	Triangulação	19
2.3.3	Técnicas <i>Fingerprinting</i> para Localização	20
2.3.4	Fase Online	22
2.3.5	Considerações	22
2.4	Aprendizagem de Máquina	23
2.4.1	Redes Neurais Artificiais (RNA)	24
2.4.2	k-Nearest Neighbor (kNN)	24
2.4.3	Support Vector Machine (SVM)	25
2.4.4	Considerações	25
2.5	Detecção de Novidade	26
2.5.1	Máquinas de Vetores de Suporte (SVM)	26
2.5.2	<i>Clustering</i>	27
2.5.3	Vizinho Mais Próximo	27
2.5.4	Estatísticas	27
3	Trabalhos Relacionados	29
3.1	Sistemas Baseados em WiFi	29
3.2	Sistemas Baseados em UWB	31
3.3	Sistemas Baseados em Infravermelho	32
3.4	Sistemas Baseados em Zigbee	33
3.5	LoRa (Long Range)	34
3.6	Sistemas Baseados em Bluetooth Low Energy (BLE)	35
3.7	Detecção de Ruído	36
4	<i>LocFiND-Localization Using Fingerprinting and Novelty Detection</i>	39
4.1	Usando Probabilidade de Previsão do kNN	41
4.2	Usando Detecção de Novidade	42
4.3	Abordagem Híbrida: Combinando Probabilidade de Previsão com Detecção de Novidade	44
5	Avaliação de Performance	47
5.1	Metodologia	47
5.2	Avaliação de Confiança Usando Probabilidade de Previsão	49
5.3	Avaliação de Confiança Usando Detecção de Novidade	52
5.4	Avaliação da Confiança Usando a Abordagem Híbrida	54
5.5	Aplicabilidade, Pontos Fortes e Fracos do Sistema Proposto	56

6	Conclusões e Trabalhos Futuros	59
6.1	Contribuições	59
6.1.1	Publicações	60
6.2	Próximos passos	60
	Referências Bibliográficas	61

Introdução

Com a crescente demanda de setores da indústria, governo, serviços, e mesmo dos consumidores finais, tem havido um grande desenvolvimento, produção e comercialização de dispositivos IoT (*Internet of Things*), o que junto ao aumento da presença de dispositivos como *smartphones*, sensores inteligentes e redes Wi-Fi em praticamente qualquer tipo de ambiente fechado, permite uma ampla gama de aplicações, entre elas a localização em ambientes fechados [Zafari et al., 2019].

A localização em ambientes fechados com alta precisão é um obstáculo tecnológico significativo para estender esses serviços IoT à áreas fechadas (salas, quartos, escritórios) onde muitos usuários passam a maior parte do tempo. Essa tarefa é, portanto, um problema em aberto na computação ubíqua, nos aplicativos sensíveis ao contexto e, especificamente nos serviços de Internet das Coisas (IoT) baseados em localização [Arbula & Ljubic, 2020].

O desafio é que, diferente de um *smartphone* que pode usar seu GPS (*Global Positioning System*) para obter informações de localização em ambientes externos, a localização por GPS pode ser problemática em ambientes fechados devido à sua fraca recepção de sinais de satélites [Wang et al., 2016].

Para resolver esse problema, nas últimas décadas uma série de soluções foram propostas usando diferentes configurações. Foram propostas soluções usando por exemplo tecnologias como Wi-Fi, Bluetooth, visão computacional, geomagnetismo, sensores inerciais, banda ultralarga (UWB), identificação por radiofrequência (RFID), ultrassom ou som, luz, e *Pedestrian Dead Reckoning* (PDR). No entanto, cada uma dessas tecnologias sofre de limitações de precisão, cobertura, custo, complexidade e aplicabilidade [Jang & Kim, 2019].

As abordagens de localização que usam a técnica chamada de *fingerprinting* normalmente usam RSSI ou CSI como parâmetros de correspondência de padrões para

determinar as posições dos dispositivos. Em comparação com as abordagens geométricas, as abordagens de *fingerprinting* são relativamente simples, facilmente incorporadas em dispositivos inteligentes de IoT e capazes de atingir uma precisão aceitável com o suporte da infraestrutura sem fio existente (Wi-Fi, celular, etc.) [Singh et al., 2021].

A localização *fingerprinting* funciona em duas fases: uma fase de treinamento offline e uma fase de localização online. Durante a fase de treinamento, as medições de radiofrequência (RF) (também conhecidas como assinaturas ou *fingerprinting*) em locais conhecidos são coletadas em um banco de dados. O banco de dados de impressão digital também é referido como o mapa de rádio [Yiu et al., 2017].

1.1 Motivação

No entanto, a tecnologia de localização para ambientes fechados requer um alto grau de precisão para atender às necessidades dos usuários [Zhu et al., 2020]. Para obter uma localização de alta precisão em ambientes fechados, os pesquisadores propuseram uma série de tecnologias e sistemas de posicionamento combinando várias tecnologias e métodos como redes sem fio, tecnologia de sensores, processamento de sinal e muitas outras tecnologias combinadas [de Oliveira, 2021]. Cada uma dessas técnicas tem suas próprias características que podem trazer benefícios ou limitações.

Na localização em ambientes fechados surgem diferentes fatores que podem dificultar a localização, mudanças no ambiente, atenuação do sinal e distorção devido à uma variedade de ruídos. Para aplicações que necessitam de um alto nível de precisão, os ruídos são o principal problema para a localização [Alkhawaja et al., 2019].

Devido à natureza dinâmica e imprevisível do canal de rádio e ao dispendioso esforço de calibração do mapa de rádio, os projetistas de sistemas de localização geralmente enfrentam limites e desafios práticos no desenvolvimento de técnicas de *fingerprinting* confiáveis e escaláveis que atendam à precisão exigida de aplicações baseadas em serviços de localização [Dashti et al., 2015].

Essa complexidade do ambiente fechado faz com que o RSSI sofra com efeitos como multipath e efeitos de sombra, que tornam o sinal RSSI instável. Isso pode fazer com que as informações da *fingerprinting* sejam inconsistentes com o RSSI coletado no estágio online. Em alguns casos, RSSI em locais diferentes são relativamente próximos uns dos outros, o que pode resultar em um grande erro nos resultados de posicionamento [Yan et al., 2018].

Devido a essa instabilidade natural do RSSI as soluções podem cometer erros na fase online se receberem uma informação de RSSI que tenha sofrido com algum

desses fenômenos descritos. Para tentar reduzir o erro causado por essa variação seria interessante verificar se a informação recebida se parece mais com as coletadas na fase *offline*.

Com esse intuito nós propomos um método usando mapas de RSSI e kNN similar a outros na literatura mas com o diferencial de uma etapa de avaliação, se aquela classificação é confiável, para descartar algumas classificações na fase *online* com o intuito de reduzir possíveis erros gerados por alguma variação gerada pelo ambiente no RSSI.

1.2 Objetivos

O objetivo dessa pesquisa é desenvolver e testar um método capaz de localizar pessoas dentro de ambientes fechados usando RSSI *fingerprinting* e kNN com o uso da técnica *Novelty Detection* para verificação do nível de confiança do resultado e descarte das estimativas menos confiáveis para melhoria da qualidade da localização.

Com esse intuito, são objetivos específicos desta pesquisa:

1. Demonstrar a eficácia do uso da probabilidade da classificação do kNN como nível de confiança de classificação;
2. Demonstrar a eficácia do uso da detecção de novidade como um segundo filtro para determinar o nível de confiança adequado;
3. Demonstrar a eficácia do uso de 2 níveis de confiança (probabilidades da classificação do kNN α e β com $\alpha > \beta$), o primeiro usando somente o limiar α e outro β combinado com a detecção de novidade;

1.3 Organização do Documento

Este documento está organizado em 6 capítulos onde, além deste capítulo introdutório 1, é possível identificar mais 5 capítulos. A primeira parte cobre os capítulos 2 e 3, que resume todas as pesquisas realizadas sobre as técnicas e tecnologias envolvidas na localização em ambientes fechados e o capítulo 4 descreve nossas técnicas e tecnologias envolvidas na nossa solução (LocFiND). Esses capítulos servem de base para a segunda parte da proposta, pois ajudam a entender as questões que envolvem essa área. A segunda parte deste documento descrita no capítulo 5 apresenta experimentos e métodos de análise que usamos para validar nossa solução e o capítulo 6 de conclusões e trabalhos futuros.

Fundamentação

As informações de posição são geralmente fornecidas por sistemas globais de navegação por satélite (GNSS - *Global Navigation Satellite Systems*), como GPS (*Global Positioning System*) ou o sistema europeu de navegação por satélite (sistema Galileo). No entanto, a precisão do posicionamento é afetada pelo ambiente, especialmente em cenários internos ou áreas urbanas densas onde a localização usando GNSS pode ser imprecisa ou mesmo impossível devido à interrupção da conexão com os satélites necessários [Yassin et al., 2017].

Com o intuito de resolver esse problema, nas últimas décadas houve o desenvolvimento de sistemas de localização sem fio em ambientes fechados. Várias tecnologias foram propostas para permitir um sistema de posicionamento, com isso diferentes sinais têm sido utilizados, incluindo WiFi, RFID, UWB, laser, luz visível, sinais acústicos e magnéticos, etc [Wu et al., 2018].

No que diz respeito a sistemas que provêm localização em ambientes fechados, existe uma relação entre precisão nas medições, dificuldade de implementação, custo para implementar, e ainda, escalabilidade para utilização em ambientes reais [Mainetti et al., 2014].

Mesmo com muitas abordagens diferentes e muitas pesquisas nessa área, o problema da localização em ambientes fechados ainda permanece sem solução amplamente aceita [Lymberopoulos et al., 2015].

Existem muitas técnicas que podem ser empregadas quando se procura estimar localização com muitas delas utilizando algum tipo de sinal de radiofrequência. Cada uma delas tem suas vantagens e desvantagens, adequando-se melhor a cada tipo de contexto.

Para o funcionamento de um sistema de localização precisamos definir qual ou quais as informações que serão utilizadas para se fazer a estimativa da localização, qual

vai ser o método de tratar esses dados, como eles serão agrupados e tratados, ou seja, precisamos de uma informação de entrada e de métodos ou algoritmos para tratá-los. A relação entre sinais de localização, medições, métodos é ilustrado na Figura 2.1.

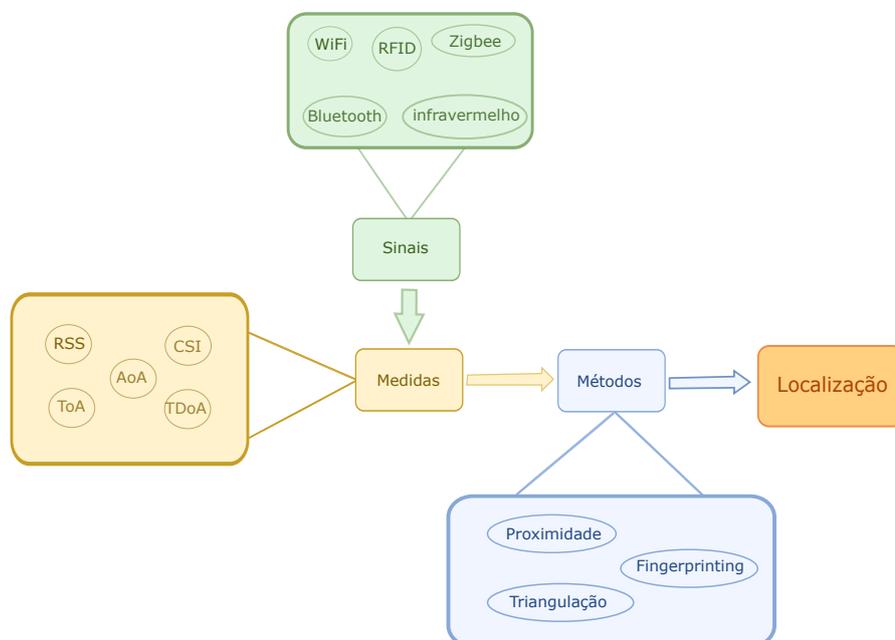


Figura 2.1: Relação entre sinais de localização, medições, métodos.

2.1 Sinais de Comunicação Sem Fio Usados para Localização

Ao selecionar uma tecnologia sem fio, fatores como alcance de transmissão, cobertura de rádio, taxa de bits, bem como a vida útil da bateria e os requisitos de energia devem sempre ser considerados para uma determinada aplicação [Sadowski & Spachos, 2018].

Os métodos de localização em ambientes fechados estimam a localização de uma entidade (por exemplo, uma pessoa ou objeto) usando tecnologias, como WiFi, UWB, Zigbee, Bluetooth, RFID, Celular, infravermelho (IR), sensores inerciais. No entanto, cada uma dessas tecnologias sofre limitações de precisão, cobertura, custo, complexidade e aplicabilidade [Gu et al., 2019].

A atenuação de sinais de comunicação sem fio em ambientes fechados é muito imprevisível por causa de fatores ambientais complexos. Além disso, a dinâmica ambiental (como movimentos humanos, abertura e fechamento de portas e rearranjo de móveis) também pode alterar a distribuição da intensidade do sinal em uma área. Para

piorar as coisas, alguns pesquisadores observaram experimentalmente que dois locais distantes em um espaço interno aberto podem ter impressões digitais de sinal semelhantes, o que de acordo com o esquema de posicionamento Wi-Fi, pode induzir a estimativa de localização incorreta [Wu et al., 2018].

2.1.1 WiFi

WiFi (*Wireless Fidelity*) é o nome dado a redes e dispositivos que implementam a especificação IEEE 802.11 para redes sem fio. Uma rede WiFi estruturada é composta por dispositivos que se comunicam por radiofrequência com um ou mais Pontos de Acesso (PA) [Tanenbaum, 2002].

A maioria dos smartphones atuais, laptops e outros dispositivos portáteis é habilitado para o uso do WiFi, o que o torna um candidato ideal para localização em ambientes fechados e uma das tecnologias mais amplamente estudadas na literatura de localização em ambientes fechados [Zafari et al., 2019].

Este aspecto se deve ao fato de que os sistemas de posicionamento que usam WiFi não requerem nenhum *software* adicional ou manipulação de *hardware*, mas são capazes de realizar a localização com base na infraestrutura existente. O RSSI é a informação mais utilizada devido à sua fácil extração em redes 802.11 e sua capacidade de rodar em qualquer *hardware* WiFi. Por outro lado, medidas que usam angulação ou necessitam de temporizadores são menos comuns em sistemas de posicionamento baseados em WiFi, pois as medições angulares e de atraso de tempo são mais complexas [Yassin et al., 2017].

2.1.2 RFID

A tecnologia de identificação por radiofrequência (RFID) foi rapidamente popularizada nas áreas de transporte inteligente, gestão logística, automação industrial e afins, e tem grande potencial de desenvolvimento devido à sua capacidade de coleta de dados rápida e eficiente [Cheng et al., 2020].

Uma rede RFID (*Radiofrequency Identification*) é composta essencialmente por três elementos principais [Bouet & dos Santos, 2008]: etiqueta que geralmente contém um número de identificação e está acoplada aos objetos ou pessoas que se pretende identificar, um leitor que detecta as etiquetas realizando determinadas operações sobre as mesmas e os servidores responsáveis pelo armazenamento e visualização de informação [Hahanov et al., 2007]. Tipicamente uma etiqueta RFID pode ser dividida em dois grupos distintos: etiquetas RFID passivas e ativas [Want, 2006].

A tecnologia RFID baseia em dois componentes principais um leitor (*reader*) e uma etiqueta que se comunicam através de sinais rádio para efetuar troca de informação, seu funcionamento diferente de outras tecnologias não requer qualquer tipo de contato sendo uma opção simples a captura de dados [Osório, 2011]. As etiquetas podem ser usadas em um objeto ou pessoa para fins de identificação e localização [Bouet & dos Santos, 2008].

Diferente das tecnologias WiFi onde a comunicação é iniciada ou finalizada deliberadamente pelos utilizadores, nos sistemas RFID a comunicação se estabelece automaticamente, iniciadas pelos leitores ou pelas etiquetas (no caso de etiquetas ativas) se estiverem ao alcance, essa característica levanta algumas questões ao nível da privacidade [Osório, 2011].

No entanto, o método de alimentação das etiquetas varia. Uma etiqueta ativa tem uma bateria interna ou está ligada a alguma outra fonte de energia e tem também seu circuito de comunicação de rádio frequência [Want, 2006]. Uma etiqueta ativa também pode ter funcionalidades adicionais, tais como a memória, e um sensor, ou um módulo de criptografia. Por outro lado, uma etiqueta passiva não tem nenhuma fonte de energia interna. Geralmente ela repassa o sinal recebido do leitor. Etiquetas passivas têm um tamanho menor e são mais baratas do que as etiquetas ativas, mas tem funcionalidades muito limitadas. Recentemente foram desenvolvidas etiquetas chamadas semi passivas. Nestas etiquetas a comunicação com os leitores é como nas etiquetas passivas, mas incorporaram uma bateria interna que alimenta constantemente o seu circuito interno [Bouet & dos Santos, 2008].

A localização baseada na tecnologia RFID pode ser categorizada em dois tipos: localização de leitores e localização de etiquetas. No caso de localização baseada em leitor, o leitor RFID geralmente é anexado à pessoa ou objeto rastreado enquanto as etiquetas são instaladas no ambiente. Tais tags fixas podem ser ativas ou passivas. Esta técnica de localização tem a vantagem de fornecer a localização absoluta do objeto em movimento e também a informação de deslocamento associada ao movimento. Por outro lado, a localização baseada em tags é adequada para muitas aplicações, como localizar mercadorias em armazéns ou rastrear bagagem em aeroportos. Ele pode fornecer a mesma precisão de localização obtida com a abordagem baseada em leitor [Hatem et al., 2020].

2.1.3 Infra-vermelho (IR)

A tecnologia de comunicação infravermelha (IR) é amplamente adotada, barata e prontamente disponível. Os sinais IR são usados em muitas aplicações diferentes, desde

controles remotos de consumidor até transferência de dados (IrDA) [Arbula & Ljubic, 2020].

Eles são relativamente direcionais, econômicos e fáceis de montar, mas têm uma desvantagem importante, não atravessam objetos sólidos. Isso pode ser visto como uma qualidade já que um sistema infravermelho instalado em um ambiente fechado não interfere em um sistema semelhante instalado nas salas ou nos prédios adjacentes [Tanenbaum, 2002]. Outra vantagem do uso de emissores e detectores de infravermelho é que eles são pequenos e baratos, porém conseguem operar a um alcance de até 6 metros [Kaspar, 2005].

Suas desvantagens baseia-se nas limitações da tecnologia que tem um alcance curto [Gu et al., 2019], além de necessitar de linha de visada para funcionar e ainda poder sofrer interferência de fontes de luz muito intensas [Casas et al., 2007].

Sistemas de localização que usam essa tecnologia costumam funcionar como sistemas baseados em proximidade, devido ao seu curto alcance, desse modo quando um dispositivo móvel tem comunicação com um ponto de referência da rede, sua localização pode ser estimada como na proximidade do ponto de referência.

2.1.4 Ultra Wideband (UWB)

O UWB é definido como a onda de rádio cuja largura de banda fracionária é maior em 20% ou pelo menos 500 MHz com base na *Federal Communication Commission* (FCC) em 2002. Este impulso ultra curto (nanossegundos) com uma alta largura de banda consegue diminuir os efeitos de multipercurso e também resultam em um melhor tempo de resolução, o que melhora muito a precisão do posicionamento [Xiao et al., 2016].

A tecnologia UWB é comumente usada por pesquisadores e pela indústria em vários campos, como posicionamento interno, a fim de apresentar melhorias em termos de obtenção de resolução e precisão de medição de alto alcance, baixa probabilidade de interceptação, imunidade multipercurso e a capacidade de combinar posicionamento e comunicação de dados em um sistema [Yassin et al., 2017]. A dificuldade para o uso do UWB é que ele exige uma infraestrutura específica e rede proprietária, e o uso de métodos de cálculo de distância baseados em tempo que dependem de sincronização estrita entre transmissores e receptores [Zafari et al., 2019].

2.1.5 Bluetooth LE (BLE)

Sistemas de localização que usam o sinal de Bluetooth normalmente são similares aos que usam Wi-Fi, muitas vezes baseados no RSSI como informação.

Soluções baseadas em Bluetooth, especialmente o recentemente lançado padrão Bluetooth *Low Energy* (BLE), têm sido adotadas como uma tecnologia promissora como informação de entrada para localização em ambientes fechados [Wu et al., 2018]. Isso se deve a características como sinal mais estável, baixo consumo de energia e alta precisão de posicionamento [Zhu et al., 2020].

Projetar um sistema de localização para ambientes fechados baseado em BLE de baixo custo com alta precisão de localização é um requisito fundamental para aplicações de IoT, mas é igualmente desafiador e difícil. Visto que o RSSI é usado para localizar o alvo, a exatidão e precisão da localização dependem da estabilidade do RSSI BLE, que na verdade, depende da potência de transmissão usada pelos dispositivos BLE [Qureshi et al., 2019]. Como o padrão BLE permite que os dispositivos BLE operem em vários níveis de potência de transmissão, é intuitivo que em um nível de potência de transmissão alto, o RSSI BLE seja mais estável em comparação com o RSSI BLE em um nível de potência de transmissão baixo [Castillo-Cara et al., 2017, Sie & Kuo, 2017].

Semelhante à localização baseada em WiFi, o Bluetooth coleta o sinal RSSI e combina aprendizado de máquina e outros algoritmos para localização [Zhu et al., 2020]. Apple e Google propuseram, respectivamente, dispositivos iBeacons e Eddystone [Bin Aftab, 2017], que têm as características de tamanho pequeno e baixo custo, permitindo a implantação em larga escala de dispositivos Bluetooth em ambientes internos.

2.1.6 Zigbee

O padrão 802.15.4 estabelece a camada Física (PHY) e a camada de acesso ao meio (MAC). Já o padrão ZigBee define a camada de rede. Desta forma, ZigBee é desenvolvido sobre a pilha do 802.15.4 e usa a camada física desta pilha [Rodrigues, 2011].

O IEEE 802.15.4 é um padrão que especifica a camada física e de controle de acesso para redes sem fio pessoais de baixa taxa de transmissão, mantida pela Zigbee Alliance com vista a satisfazer a necessidade de criação de redes de baixo custo e baixo consumo [IEEE, 2005][Yang & Yang, 2009].

O ZigBee é uma tecnologia semelhante ao Bluetooth que opera a cerca de um quarto da taxa máxima de dados de 1 Mbps do Bluetooth. A baixa taxa de dados o torna inadequado para aplicativos com transmissão de dados de alta velocidade, mas a solução permite uma vida útil da bateria de vários anos e conexão a um grande número de nós [Loganathan et al., 2019].

É um padrão para redes de área pessoal (PANs), que visa a baixa taxa de dados, curto alcance de comunicação e baixo custo. Adicionalmente, as aplicações devem ter

baixa taxa de transferência de dados [Rodrigues, 2011]. Ele pode prover uma rede *mesh* que oferece alta confiabilidade, e também torna mais fácil a implantação [Larranaga et al., 2010].

Permite uma boa comunicação *peer-to-peer*, com a rede suportando três tipos de topologias: estrela, árvore e malha cada uma com até 65535 nós, com uma estrutura auto-organizável e auto-reparável, que permite o *relay* de dados através de *multi-hop* entre quaisquer dois nós [Osório, 2011].

Por ser uma tecnologia de radiofrequência, a localização pode ser feita usando o RSSI como informação para estimativa de distâncias similar aos sistemas usados em WiFi e Bluetooth.

2.1.7 Long Range (LoRa)

A tecnologia LoRa foi desenvolvida para Low Power Wide Area Networks (LPWANs) para suportar aplicações M2M (Machine-to-Machine) e IoT (Internet das Coisas) [Lam et al., 2019].

LoRa utiliza arquitetura em estrela de longo alcance em vez da arquitetura de rede mesh mais comum. Na topologia em estrela, os nós individuais não precisam encaminhar os dados de outros nós, o que preserva a energia e aumenta a vida útil da bateria [Kim et al., 2021].

Com custo e consumo de energia relativamente baixos, o sinal do LoRa com alcance em nível de cidade é resiliente ao efeito ou ruído de multicaminhos [Kim et al., 2021]. Portanto, os sinais LoRa têm boa capacidade de penetração e estabilidade e é mais estável que WiFi e BLE [Islam et al., 2019].

No entanto, em redes LoRa, quando uma estação base recebe um sinal e seu valor RSSI é pequeno, não sabemos se a estação base está distante do transmissor ou é altamente afetada por ruído não-Gaussiano, isso é um desafio para localização em redes LoRa [Lam et al., 2019].

2.1.8 Considerações

Para localização em ambientes fechados cada tipo de tecnologia tem suas vantagens e desvantagens que podem ser apropriadas para resolver o problema em vários níveis dependendo dos requisitos de cada sistema. Uma breve comparação entre algumas das principais tecnologias avaliadas neste trabalho podem ser vistas na Tabela 2.1.

Para nossa aplicação, que é a localização de alunos em salas de aula, precisamos de uma tecnologia que não tenha um custo muito elevado, uma precisão razoável e

Tecnologia	Vantagens	Desvantagens
WiFi	Fácil de implementar, expandir e medir, baixo custo	Dificuldades na obtenção de base para <i>fingerprinting</i> , muito afetado por mudanças no ambiente
RFID	alta precisão, baixo custo	Baixo alcance de comunicação, implantação de equipamentos tediosos.
Infravermelho	alta precisão	Requer equipamentos especiais, alto custo, não penetra obstáculos, alto consumo de energia, muito afetado por mudanças no ambiente
UWB	alta precisão, alta penetração, altas taxas de transmissão	Requer equipamentos especiais, alto custo
Bluetooth (BLE)	Baixo consumo de energia, baixo custo, alta taxa de transferência, tecnológica bem popular	alta propensão a ruído, baixa precisão
Zigbee	Baixo consumo de energia, baixo custo	muito afetado por mudanças no ambiente, disponível em poucos equipamentos
LoRa	Baixo consumo de energia, baixo custo	disponível em poucos equipamentos, ruído não-Gaussiano

Tabela 2.1: Algumas das principais tecnologias usadas em localização para ambientes fechados.

baixo consumo de energia.

O uso de Infravermelho e UWB apesar de terem uma alta precisão tem um custo alto para obtenção de equipamentos com essas tecnologias. O Zigbee tem a vantagem do baixo custo e baixo consumo de energia, porém é uma tecnologia que não é muito difundida ainda sendo utilizada em poucos equipamentos. Soluções baseadas em RFID tem o problema da escalabilidade, seriam necessárias muitas TAGs para ter uma boa cobertura. Para o uso de LoRa é necessário uma tecnologia específica que ainda não é muito difundida e ainda tem muitos desafios tecnológicos para serem resolvidos.

Sobram o WiFi e Bluetooth (BLE) que têm em comum os pontos negativos com problemas com ruído e precisão não tão boa. O Bluetooth (LE) tem vantagem do baixo consumo de energia e ambos são tecnologias bem difundidas e de grande disponibilidade então, optamos pelo Bluetooth (LE).

2.2 Medidas

Em geral, a medição envolve a transmissão e recepção de sinais entre os componentes de *hardware* do sistema. Um sistema de localização que usa redes sem fio consiste em pelo menos dois componentes de *hardware* separados: um transmissor de sinal e uma unidade de medição. Este último geralmente carrega a maior parte da "inteligência" do sistema [Liu et al., 2007]. Existem, várias formas de medir um sinal de comunicação sem fio algumas são descritas a seguir.

2.2.1 Intensidade do Sinal Recebido (RSSI)

As abordagens baseadas no uso da intensidade do sinal recebido (*RSSI-Received Signal Strength Indicator*) são as mais simples e amplamente utilizadas para localização em ambientes fechados. O RSSI pode ser usado para estimar a distância entre um transmissor e um receptor usando a lógica de que quanto maior o valor RSSI menor a distância entre transmissor receptor [Zafari et al., 2019].

Usando o RSSI e um modelo simples de propagação de perda [Kumar et al., 2009], a distância d entre um transmissor e um receptor pode ser estimada a partir como (2.1)

$$RSSI = -10n \log_{10}(d) + A \quad (2.1)$$

onde n é o expoente de perda (que varia de 2 em ambientes abertos para 4 em ambientes fechados) e A é o valor RSSI a uma distância de referência do receptor.

O uso de força de sinal tem a vantagem de não necessitar de *hardware* extra, pois essa informação é disponibilizada pelo padrão WiFi, porém ela não tem um comportamento muito previsível, tendo vários fatores que influenciam sua variação que vão desde o *layout* do ambiente até a sua temperatura como observado em [Bahl & Padmanabhan, 2000].

Como o RSSI naturalmente tem uma variação, para ele ser melhor utilizado como informação principal, são necessárias técnicas para se tratar e estimar a localização. Muitas dessas técnicas usam abordagens probabilísticas como em [Bahl & Padmanabhan, 2000] e [Smailagic & Kogan, 2002] para geolocalização.

2.2.2 Informações de Status do Canal (CSI)

O CSI (*Channel Status Information*) são informações que são estimadas pelo canal representando as propriedades de um link de comunicação [Wu et al., 2013].

Com o uso de sistemas de multiplexação por divisão de frequência ortogonal (OFDM) e sistemas de múltiplas entradas e saídas múltiplas (MIMO) no protocolo 802.11a/n, os sinais de informações de estado do canal (CSI) podem ser extraídos de equipamentos WiFi comerciais [Dang et al., 2019]. Em contraste com os sinais RSSI, que fornecem apenas informações de amplitude, os sinais CSI podem fornecer tanto a fase da subportadora quanto às informações de amplitude, bem como descrições melhores das mudanças de sinal do transmissor para o receptor do que as fornecidas pelos sinais RSSI [Yang et al., 2013a]. Mais especificamente, CSI descreve como um sinal se propaga do (s) transmissor (es) para o (s) receptor (es) e revela o efeito combinado de, para por exemplo, espalhamento, desvanecimento e queda de energia com a distância. Em resumo, a precisão do CSI influencia muito o desempenho geral do sistema OFDM [Wu et al., 2013].

Com isso ele fornece uma fonte de informações muito mais rica do que o RSSI, com isso as interfaces de rede que usam OFDM, com estimativas de canal que são menos suscetíveis a interferência do que usando somente o espectro de propagação (devido à correlação mais baixa) e são calibrados. Ambos os fatores levam a melhores medições do que como feito anteriormente [Halperin et al., 2010].

2.2.3 Tempo de Chegada (ToA)

A técnica ToA se baseia na ideia de que a distância entre o alvo móvel para a unidade de medição é diretamente proporcional ao tempo de propagação [Liu et al., 2007].

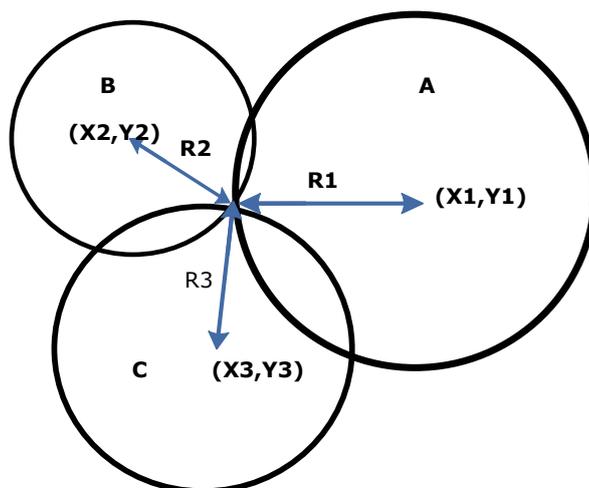


Figura 2.2: Localização por Tempo de Chegada.

A fim de permitir o posicionamento 2-D, as medições devem ser feitas com respeito a sinais de pelo menos três pontos de referência, como mostrado na Figura 2.2.

Para os sistemas baseados em ToA, o tempo de propagação de sentido único é medido, e a distância entre a unidade de medição e o transmissor de sinal é calculada. Em geral, os resultados diretos ToA têm dois problemas. Em primeiro lugar, todos os transmissores e os receptores do sistema têm que ser precisamente sincronizados, o que necessita de *hardware* específico [Fink et al., 2010]. Em segundo lugar, todos os contadores precisam ser rotulados para cada sinal de transmissão a fim de medir as unidades e discernir a distância que o sinal viajou [Liu et al., 2007]. ToA é limitado pelo requisito de pelo menos 3 estações base (ou ANs) para gerar uma localização 2-D [Yassin et al., 2017].

2.2.4 Diferença Entre os Tempos de Chegada (TDoA)

A técnica TDoA funciona medindo o tempo exato de chegada de um sinal transmitido em três ou mais locais separados. Como as ondas de rádio viajam a uma velocidade conhecida (à velocidade da luz), por meio do cálculo da diferença dos tempos de chegada de pares de receptores, é possível calcular as hipérboles em que o dispositivo de transmissão está situado [Assad, 2007].

Esta técnica pode funcionar com qualquer dispositivo de transmissão e só requer modificações nos receptores [Malekitabar et al., 2005].

Para cada medição TDoA, o transmissor deve situar-se em uma hiperboloide com uma diferença de alcance constante entre as duas unidades de medição. A equação da

hiperboloide é dada pela equação (2.2)

$$R_{i,j} = (x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 - (x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2 \quad (2.2)$$

onde (x_i, y_i, z_i) e (x_j, y_j, z_j) representam os receptores fixos i e j , e (x, y, z) , representam as coordenadas do alvo [Drane et al., 1998]. Exceto as soluções exatas para a equação TDoA hiperbólica mostrada em (2.2) por meio de regressão não linear, a solução mais fácil é para a linearização das equações através do uso de uma série de Taylor e criar um algoritmo iterativo [Torrieri, 1984].

É ainda necessária a sincronização temporal entre os vários pontos de acesso, o que torna este método dependente de *hardware* e como tal uma solução mais dispendiosa [Fink et al., 2010].

Outra desvantagem dessa técnica se deve a em ambientes internos termos muitos pontos de acesso e dispositivos sem fios existentes, os sinais são refletidos nos obstáculos existentes como paredes, mobílias e até mesmo pessoas [Osório, 2011]. TDoA tem como uma limitação o requisito de pelo menos 3 estações base (ou ANs) para gerar uma localização 2-D [Yassin et al., 2017].

2.2.5 Ângulo de Chegada (AoA)

Em AoA, a localização do alvo desejada pode ser encontrada pela intersecção de vários pares de linhas em ângulo, cada um formado pelo raio circular a partir de uma estação de base ou uma estação móvel para o alvo [Liu et al., 2007].

Como mostrado na Figura 2.3, os métodos AoA podem usar pelo menos dois pontos de referência conhecidos (A, B) , e dois ângulos medidos θ_1, θ_2 para localizar em 2-D o ponto P . Localização AoA pode ser feita tanto com antenas direcionais ou com *arrays* de antenas [Liu et al., 2007].

As vantagens de AoA são que uma estimativa da posição pode ser determinada com apenas três unidades de medição de posicionamento 3-D ou duas unidades de medição de 2-D, e não necessita de sincronização entre as unidades de medição além de não ser necessária nenhum tipo de sincronização para que ele funcione.

Por outro lado a necessidade de *hardware* específico [Rodrigues, 2011], que geralmente tem custo elevado [Röhrig & Müller, 2009] e a queda na qualidade da localização quando o alvo está muito distante dos receptores [Liu et al., 2007] são desvantagens dessa abordagem. Algumas literaturas também chamam AoA como direção de chegada (DoA).

O desempenho das técnicas de AoA é altamente dependente do alcance forne-

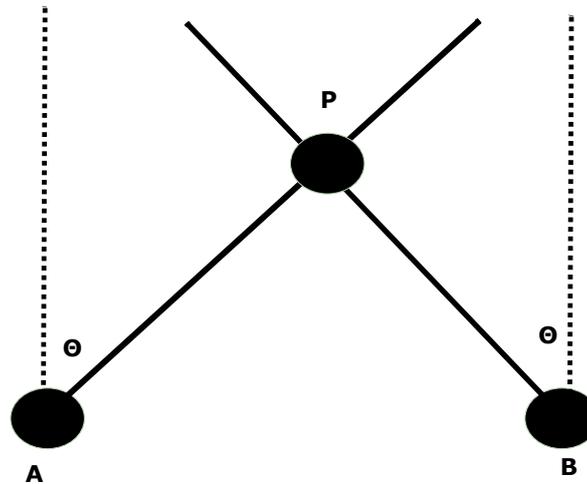


Figura 2.3: Localização por Ângulo de Chegada.

cido com erros significativos de estimativa de posição a partir de erros relativamente pequenos nas medições de AoA. Eles são restritos pela frequência do transmissor e pela quantidade de antenas. Assim, são utilizados apenas para localização em aplicações com exigência de baixa precisão ou em combinação com outras medições. Além disso, os sistemas AoA são sensíveis ao multicaminho angular, um efeito importante no ambiente fechado [Yassin et al., 2017].

2.2.6 Considerações

Cada tipo de medida pode ser usada para localização em ambientes fechados tem suas próprias particularidades, algumas necessitam e/ou funcionam melhor com certos tipos de tecnologia e outras requerem configurações extras que podem dificultar sua implementação. Uma breve comparação entre algumas das principais medidas avaliadas neste trabalho podem ser vistas na Tabela 2.2.

Apesar do uso de AoA ter alta precisão principalmente em menores distâncias são necessárias antenas direcionais o que dificulta e encarece a implantação. ToA e TDoA também tem boa precisão porém necessitam de trabalho extra em sincronização de relógios dos equipamentos o que precisa de ajustes muito finos do que depende em muito sua performance.

O CSI aparece com uma excelente solução em muitos aspectos melhor que o RSSI porém a dificuldade de modificações necessárias para ser habilitada em algumas interfaces.

Medidas	Vantagens	Desvantagens
RSS	Fácil de implementar, custo benefício, disponível na maioria das tecnologias sem fio	muito afetado por mudanças no ambiente, muito afetado por multipath, baixa precisão
CSI	Mais robusto que RSS e multipath comparado com RSS	Não está disponível em muitas interfaces de rede
ToA	Alta precisão	Requer sincronização entre transmissores e receptores pode requerer timestamps e múltiplas antenas nos transmissores e receptores, necessita de linha de visada para uma boa performance
TDoA	Não requer fingerprinting, não requer sincronização de relógios entre o device e o RN	requer sincronização de relógios entre RNs, pode requerer time stamps, requer muita banda de comunicação
AoA	Alta precisão	requer antenas direcionais e <i>hardware</i> mais complexo, requer algoritmos mais complexos, performance cai quanto maior a distância entre transmissor e receptor

Tabela 2.2: Algumas das principais medidas usadas em localização para ambientes fechados.

2.3 Métodos Usados para Localizar

Para localização em ambientes fechados existem métodos determinísticos que podem estimar a localização mais rapidamente usando técnicas de medição clássicas como Tri- lateração/Triangulação e proximidade e métodos probabilísticos baseados em teoria de probabilidades e/ou Aprendizado de máquina. Métodos determinísticos, usam cálculos mais simples do que métodos probabilísticos, os métodos probabilísticos estimam a localização normalmente com mais precisão. Devido ao fato de que os métodos proba- bilísticos requerem mais informações do que os métodos determinísticos, eles realizam uma detecção de localização mais precisa em comparação com os métodos determinís- ticos [Zekavat & Buehrer, 2012]. No entanto, eles são ineficazes para dispositivos mais limitados, como sensores, por terem baixa capacidade de computação [Turgut et al., 2016].

Alguns dos métodos que podem ser usados para localização em ambientes fechados são descritos a seguir.

2.3.1 Proximidade

As abordagens de proximidade determinam a localização de um objeto detectando se o objeto está próximo de um local conhecido ou de uma área [Gu et al., 2019].

Os algoritmos de proximidade fornecem informações de localização relativa sim- bólica. Normalmente, ele depende de uma densa grade de antenas, cada uma com uma posição bem conhecida. Quando um alvo móvel é detectado por uma única an- tena, considera-se que está localizado com ela. Quando mais de uma antena detecta o alvo móvel, ela é considerada colocada com aquela que recebe o sinal mais forte. Este método é relativamente simples de implementar [Liu et al., 2007]. Normalmente são usados em sistemas de localização usando infravermelho e RFID [Ngamakeur et al., 2020].

Quando um único sensor detecta uma pessoa, a pessoa é considerada co-localizada com ele, conforme mostrado na Figura 2.4. Por exemplo, a pessoa P1 ativa o nodo sensor A2. Assim, a localização de pessoa P1 é considerada igual à localização do sensor A2. Este método é em geral fácil de implementar usando diferentes tipos de tecnologias.

2.3.2 Triangulação

Com base nas propriedades geométricas dos triângulos [Gu et al., 2019], são adequados para uso especialmente na comunicação em linha de visada. As técnicas de medição

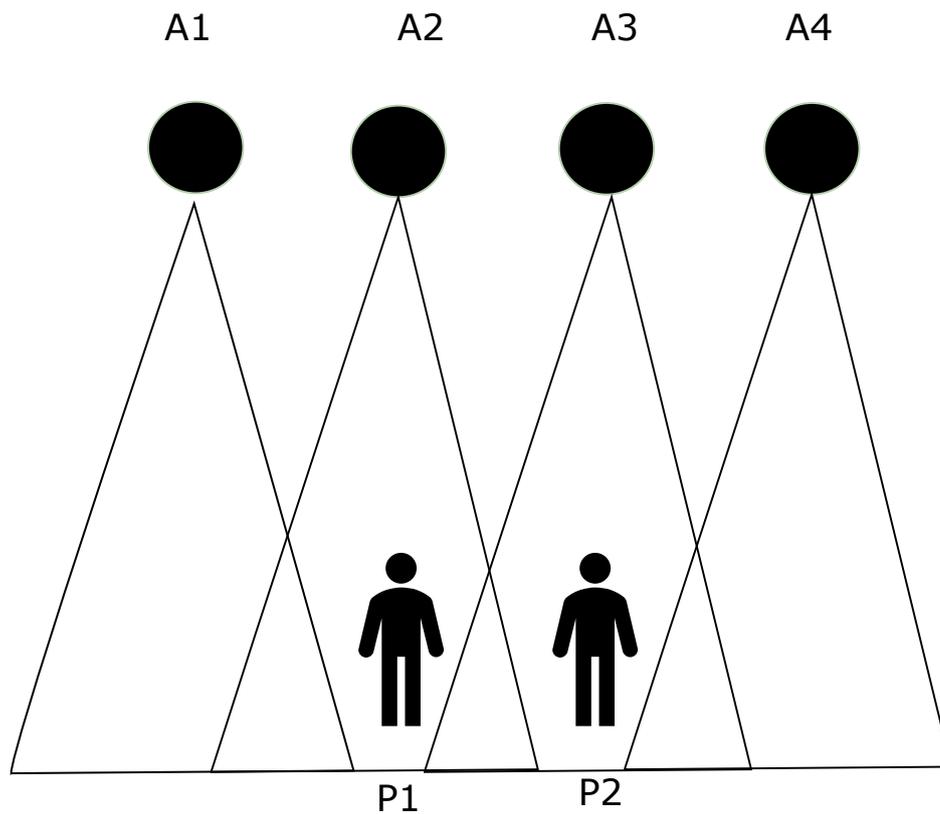


Figura 2.4: Localização por Proximidade.

do ângulo de chegada (AoA) ou do ângulo de partida (AoD) normalmente usam os algoritmos de triangulação [Turgut et al., 2016]. Cada método de triangulação tem vantagens e desvantagens. ToA é a técnica mais precisa, que pode filtrar os efeitos de caminhos múltiplos nas situações internas. RSSI e ToA precisam saber a posição de pelo menos três elementos de referência, para estimar a posição de um objeto. AoA requer apenas dois elementos de medição de posição para realizar estimativa de localização [Gu et al., 2009]. No entanto, quando o objeto alvo a ser localizado está longe, o método AoA pode conter alguns erros, o que resultará em menor precisão [Chen et al., 2006].

2.3.3 Técnicas *Fingerprinting* para Localização

Assim como a impressão digital de uma pessoa é única em um indivíduo, a sua impressão digital pode ser usada para reconhecer ele individualmente. Da mesma forma, o princípio fundamental da técnica de impressão digital (*fingerprinting*) para localização é usar um padrão de uma informação particular sobre o seu entorno para reconhecer uma localização correspondente [Xu & GmbH, 2016].

Muitos sistemas de localização são concebidos com base nas técnicas *fingerprintin-*

ting, visto que ela apresenta bons resultados além de ser relativamente fácil e barata de implementar em comparação com outras técnicas [Abdat et al., 2010].

A localização usando *Fingerprinting* pode, teoricamente, usar qualquer fenômeno físico que difere entre locais, até mesmo luz ou temperatura. É claro que é benéfico usar fontes que sejam temporalmente mais ou menos estáveis. Assim, a maioria dos sistemas de posicionamento para ambientes fechados usa sinais de rádio como GSM, Bluetooth ou WiFi [Bolliger, 2011].

Um sistema de posicionamento baseado em *Fingerprinting* tem duas fases: a fase de treinamento (*off-line*) e fase de determinação da posição (*on-line*) em que é feita a análise da posição mais compatível para estimativa de localização [Atia et al., 2013, Au et al., 2013].

2.3.3.1 Fase Offline

A fase de treinamento ou *offline* é focada na coleta e processamento de dados preliminares. Um bom método de coleta deve selecionar cuidadosamente os pontos de treinamento: nem muito poucos, o que reduz a precisão da localização, nem demais, o que exige uma quantidade maior de dados trabalho de coleta [Gao, 2015].

Um grande número de sistemas de localização para ambientes fechados foi desenvolvido com base nos valores RSSI [Chen et al., 2015, Cui et al., 2017]. O ambiente de interesse (onde deve ser estimada a localização de algum elemento) é dividida com a ajuda de uma planta baixa ou mapa do local em partes ou células menores, que serão utilizados como pontos de referência [Jiang et al., 2016]. Os valores RSSI dos *Access Points* (APs) são coletados dentro de cada célula e armazenados em um mapa de rádio, de modo que cada *fingerprinting* mantenha as informações sobre a localização em uma célula e uma lista de valores RSSI [Minaev et al., 2017].

Já que a maioria dos dados brutos sem treinamento são caóticos e redundantes alguns algoritmos de localização processam os dados brutos e salvam apenas um modelo de *fingerprinting* para a fase de teste (*Online*) [Gao, 2015].

A partir da base de dados de *fingerprinting*, métodos estatísticos e de Aprendizado de Máquina supervisionado são então utilizados para criação de modelos que, a partir dos dados de entrada, seja possível extrapolar e estimar a localização do elemento móvel. Alguns modelos usados são: Naive Bayes, K-vizinhos mais próximos (do inglês K-Nearest Neighbors - KNN), redes neurais, entre outros [de Oliveira, 2021].

2.3.4 Fase Online

Na fase de teste online, quando o usuário móvel se desloca para um local desconhecido, as impressões digitais correspondentes à posição atual são enviadas ao servidor para localização. Como o banco de dados reserva as impressões digitais já conhecidas e suas posições correspondentes, o algoritmo de posição pode estimar a posição atual por meio da busca de impressões digitais correspondentes no banco de dados [Gao, 2015]. É mais provável que o local com as impressões digitais mais correspondentes seja próximo à posição atual [Minaev et al., 2017].

O ponto fraco do método tradicional é que na fase online os dispositivos podem não estar calibrados, então os dispositivos são capazes de apresentar valores diferentes de RSS para a mesma ponto de acesso [Laoudias et al., 2013].

Na fase *on-line* para determinar a posição de um dispositivo sem fio, o sistema tenta corresponder o sinal de um dispositivo com as entradas da tabela. A correspondência mais próxima da entrada é a provável posição do dispositivo. Normalmente essas correspondência é calculada por meio de técnicas de reconhecimento de padrões: k-vizinhos mais próximos (kNN), redes neurais (RNA), métodos probabilísticos, e máquinas de vetores de suporte (SVM) [Abdat et al., 2010].

2.3.5 Considerações

Similar ao que acontece com as medidas os métodos também tem suas próprias particularidades, algumas necessitam e/ou funcionam melhor com certos tipos de tecnologia e/ou requerem configurações extras que podem dificultar sua implementação ou ainda podem limitar o escopo e implantação de um sistema. Uma breve comparação entre alguns dos principais métodos avaliadas neste trabalho podem ser vistas na Tabela 2.3.

No geral, os métodos de proximidade são muito simples e fáceis de implantar, mas sua precisão de localização depende do número de dispositivos de ancoragem (por exemplo, terminais POS, leitores RFID) que detectam as etiquetas. A cobertura e aplicabilidade das abordagens de proximidade são pobres, uma vez que esses dispositivos de ancoragem são geralmente instalados em certas áreas, tornando difícil dimensionar sistemas baseados em proximidade [Gu et al., 2009].

Os métodos de triangulação são muito dependentes de qual tecnologia sem fio utilizada. UWB ou sinais de luz geralmente podem atingir altas precisões. Já usada com outras tecnologias sem fio como WiFi e Bluetooth pode propagar erros gerados por mudanças no ambiente e por *multipath*.

A impressão digital é um dos métodos de localização interna comumente usados devido à sua capacidade de fazer uso da infraestrutura existente (por exemplo, APs

Métodos	Vantagens	Desvantagens
Proximidade	Simple e fáceis de implantar	Baixa precisão, localização simbólica, dificuldade de escalar para grandes áreas
Triangulação	Algoritmos Simples, pode ter alta acurácia dependendo da tecnologia usada	Muito dependente de características da tecnologia sem fio utilizada
Fingerprinting	Fácil de usar	Base de impressões digitais precisa ser obtida novamente e/ou atualizada se houverem mudança no ambiente

Tabela 2.3: Alguns dos principais métodos usadas em localização para ambientes fechados.

WiFi) ou assinaturas internas (por exemplo, magnetismo, recursos visuais de objetos). O principal desafio é a construção de um banco de dados de impressão digital, que pode ter um custo proibitivo para ambientes de grande escala se for feito manualmente.

No nosso caso o *fingerprinting* apesar do custo adicional para obtenção da base é o método que mais se ajusta a nossas escolhas anteriores.

2.4 Aprendizagem de Máquina

No passado recente, testemunhamos a adoção de técnicas de aprendizado de máquina em localização em ambientes internos e sua eficácia na extração de conhecimento, descobrindo, aprendendo e melhorando a precisão da localização. Essas abordagens são mais eficazes do que as abordagens matemáticas tradicionais para problemas não lineares complexos que são muito complicados com a escrita de regras e/ou equações [Roy & Chowdhury, 2021]. Algumas das principais técnicas que são usadas em localização em ambientes fechados são: redes neurais artificiais, kNN e SVM.

2.4.1 Redes Neurais Artificiais (RNA)

Redes Neurais Artificiais (RNA) são usadas em uma série de cenários de classificação e previsão. Para localização, a RNA deve ser treinada usando os valores RSS e as coordenadas correspondentes que são obtidas durante a fase *off-line* [Altini et al., 2010].

Como um dos algoritmos mais conhecidos em aprendizado de máquina, a RNA tem se desenvolvido rapidamente nos últimos anos. Como sua estrutura de rede de múltiplas camadas pode se ajustar a diferentes mudanças complexas de um modelo não linear, ela também é adequada para problemas de localização. Contanto que o modelo de treinamento funcione bem, ele pode localizar com precisão e pode resistir às influências do ruído e do efeito de caminhos múltiplos [Zhu et al., 2020].

A Rede Neural detecta inerentemente relacionamentos não lineares complexos entre os recursos dependentes e independentes. Normalmente, a camada oculta tem a capacidade de identificar a inter-relação entre as variáveis de entrada. O modelo baseado em Rede Neural é computacionalmente eficiente e um pequeno conjunto de parâmetros precisa ser ajustado para manter um baixo custo de comunicação. Além disso, o modelo baseado em rede neural é facilmente retreinado quando um conjunto de dados interno está desatualizado ou as novas amostras precisam ser coletadas para refletir as mudanças no ambiente interno. Uma das preocupações importantes é que esse modelo pode ser muito sujeito a *overfitting* [Roy & Chowdhury, 2021].

2.4.2 k-Nearest Neighbor (kNN)

O k-Nearest-Neighbours (kNN) [Aha et al., 1991] é um método de classificação não paramétrico, que é simples, mas eficaz em muitos casos. É um classificador baseado em distância que classifica as instâncias com base em sua similaridade [Bozkurt et al., 2015].

Funciona encontrando as k instâncias de treinamento que estão mais próximas da instância a ser classificada, então é usada a classificação dessas k instâncias a mais próxima para definir a classe da nova amostra [Bramer, 2020]. A melhor escolha de número de vizinhos depende dos dados. Geralmente, valores maiores de k reduzem o efeito do ruído na classificação, mas tornam os limites entre as classes menos distintas [D'Aloia et al., 2020].

O algoritmo kNN é fácil de implementar e requer apenas dois parâmetros, ou seja, o valor de k e a função de distância (por exemplo, Euclidiana ou Minkowski ou Manhattan, etc.). Conforme o conjunto de dados cresce, mais tempo é necessário para calcular a distância entre um novo ponto de dados e cada ponto de dados existente. Portanto,

o desempenho do kNN diminui rapidamente com o crescimento do conjunto de dados. Além disso, o desequilíbrio de dados cria problemas em kNN [Roy & Chowdhury, 2021].

2.4.3 Support Vector Machine (SVM)

O SVM é um método de aprendizado de máquina baseado em teoria estatística, que apresenta certas vantagens no processamento de dados de alta dimensão e problemas não lineares, principalmente no tratamento do problema de regressão entre impressões digitais e posições [Zhu et al., 2020].

As SVMs são usadas em sistemas de localização treinando os vetores de suporte usando o mapa de rádio que consistem em pontos em uma grade. As SVMs analisam a relação entre as impressões digitais treinadas e seus pontos na grade, considerando cada ponto da grade como uma classe. As impressões digitais RSSI testadas são tomadas como uma entrada para o SVM que prevê a classe à qual a amostra testada pertence [Salamah et al., 2016].

Sua capacidade de se adaptar facilmente a ambientes multi classe, nos quais os dados não são linearmente separáveis, é um o critério básico pelo qual foram selecionados para resolver muitos problemas de localização [Chriki et al., 2017].

O SVM escala relativamente bem com dados dimensionais elevados (ou seja, os conjuntos de dados com um grande número de features). Os modelos SVM são bastante estáveis, portanto, as pequenas alterações nos dados devido à natureza dinâmica do ambiente interno não afetam o (s) hiperplano (s) do SVM. Além disso, essas vantagens, o SVM também tem alguns problemas. Selecionar uma função de kernel apropriada é uma tarefa muito complicada e complexa. Um grande número de vetores de suporte são gerados usando um kernel de alta dimensão que por sua vez reduz o tempo de treinamento. Com o aumento do tamanho do conjunto de dados de treinamento, o requisito de memória aumenta para armazenar todos os vetores de suporte [Roy & Chowdhury, 2021].

2.4.4 Considerações

Em nossos testes iniciais para definição do algoritmo de classificação constatamos que com nossa base de impressões digitais os 3 obtiveram resultados muito similares. Com isso apesar das desvantagens escolhemos usar o kNN devido ele ter um resultado igual ou superior aos demais para nosso problema mesmo tendo uma complexidade menor e um custo computacional e tempo de resposta menor para o tamanho da nossa base de teste.

2.5 Detecção de Novidade

De fato, novidade, anomalia e detecção de outliers são termos relacionados a encontrar padrões que são diferentes dos padrões normais, usuais. Enquanto os termos anomalia e outliers dão a ideia de um padrão indesejado, o termo novidade indica um conceito emergente ou novo que precisa ser incorporado ao padrão normal [Faria et al., 2016].

A detecção de novidade refere-se à identificação de padrões novos ou anormais embutidos em uma grande quantidade de dados normais. A novidade (anomalia, *outlier*, exceção) acontece quando uma amostra não está de acordo com o comportamento normal esperado [Miljković, 2010].

Na detecção de novidade a descrição da normalidade é aprendida construindo um modelo com vários exemplos representando instâncias positivas (ou seja, dados indicativos de comportamento normal do sistema). Padrões anteriormente não vistos são testados comparando-os com o modelo de normalidade, muitas vezes resultando em alguma forma como um padrão que ainda não foi visto (novidade) [Pimentel et al., 2014]. Detectar amostras novas e nunca vistas é uma habilidade importante para qualquer sistema de classificação. Ao perceber que não é possível treinar um sistema de aprendizado de máquina com todos os tipos de dados que o sistema provavelmente encontrará, torna-se importante ser capaz de diferenciar entre informações de dados conhecidos e desconhecidos durante a fase *online* [Markou & Singh, 2003].

Técnicas de detecção de novidades podem ser implementadas usando diferentes técnicas, como as baseadas em máquinas de vetores de suporte (SVM), baseadas em *clustering*, baseadas em vizinho mais próximo (kNN) e as estatísticas [Sayyed & Deolekar, 2017].

2.5.1 Máquinas de Vetores de Suporte (SVM)

Uma abordagem que usa Máquinas de Vetores de Suporte (SVM) assume que pontos normais pertencem a regiões de dados de alta densidade e novos padrões a regiões de dados de baixa densidade [Miljković, 2010]. É construído um esquema de limite esférico que agrupa e inclui a maioria das amostras de dados normais. Qualquer amostra nova que caia fora dos limites do grupo é detectada como novidade. Em trabalhos recentes sobre detecção de novidades baseada em SVM, uma variação chamada OC-SVM (SVM de uma classe) é usada uma vez que tem um kernel que permite separar facilmente dados normais e novos em espaços com muitas dimensões [Ding et al., 2014].

2.5.2 *Clustering*

Na detecção de novidade baseada em cluster, existem várias maneiras de identificar a novidade, como clusters que estão localizados longe de outros clusters, ou menores ou mais esparsos que outros clusters, ou pontos que não pertencem a nenhum grupo são considerados dados anormais [Amarbayasgalan et al., 2018]. Mas o desempenho do método baseado em cluster é muito dependente de qual algoritmo é usado, assim como dados de alta dimensão em cluster com distância são desafios especiais para algoritmos de mineração de dados [Zimek et al., 2012].

Um método de detecção de novidade baseado em *Clustering* usando o algoritmo de *Clustering* fuzzy de Gustafson-Kessel e usado para detectar perdas não técnicas em um sistema elétrico [Viegas et al., 2018].

2.5.3 Vizinho Mais Próximo

Esta técnica assume que os dados normais estão próximos de suas vizinhanças, enquanto as novidades potenciais estão longe de seus vizinhos. É um método muito simples e eficaz, mas a desvantagem é que ele precisa armazenar todos os pontos de dados de treinamento que são usados para calcular a distância entre um novo ponto de dados não visto e todos os pontos de dados fornecidos [Ding et al., 2014].

A abordagem de k-vizinhos mais próximos (k-NN) é baseada na suposição de que pontos de dados normais têm vizinhos próximos no conjunto de treinamento “normal”, enquanto novos pontos estão localizados longe desses pontos [Hautamaki et al., 2004]. Um ponto é declarado como outlier se estiver localizado longe de seus vizinhos. A distância euclidiana é uma escolha popular para atributos contínuos univariados e multivariados, mas outras medidas, como a distância de Mahalanobis, podem ser usadas [Pimentel et al., 2014].

Trabalhos recentes sobre a detecção de novidades com base nos vizinhos mais próximos giram em torno de dois temas: conceber estratégias de detecção mais eficazes e reduzir o tamanho dos dados armazenados necessários, mantendo a precisão da detecção [Ding et al., 2014].

2.5.4 Estatísticas

As abordagens estatísticas são principalmente baseadas em dados de modelagem com base em suas propriedades estatísticas e usando essas informações para estimar se uma amostra de teste vem da mesma distribuição ou não [Markou & Singh, 2003].

A Gaussian Mixture Modeling (GMM) modela a distribuição geral estimando a densidade usando menos kernels do que o número de padrões no conjunto de treinamento [Markou & Singh, 2003]. Os parâmetros são estimados pelo critério de maximum likelihood (ML) usando o algoritmo Expectation Maximization (EM) [Miljković, 2010].

A cadeia de Markov (Hidden Markov Model - HMM) é uma técnica estatística usada para modelar dados sequenciais (adequado para padrões). O HMM é composto por vários estados juntamente com a probabilidade de se mover entre pares de estados (probabilidades de transição) [Norris, 1998]. HMM são treinadas em dados normais (usando o algoritmo de Baum-Welch). Um novo padrão sequencial é detectado se a probabilidade de observar tal sequência calculada usando pontuações abaixo de um limite (pontuação de Viterbi) [Miljković, 2010].

Trabalhos Relacionados

Os métodos de localização mais básicos são baseados na estimativa da distância do dispositivo para outros nós (conhecidos como âncoras) com posições conhecidas. Diferentes tipos de tecnologia, métricas e técnicas já foram estudadas para realizar a tarefa de estimar a posição de um objeto ou pessoa em ambientes fechados. Usando algumas tecnologias citadas neste trabalho, como Wi-Fi, UWB, Infravermelho, Zigbee, LoRa, e Bluetooth Low Energy (tecnologia escolhida para implementação deste trabalho) essa revisão apresentará trabalhos que trazem estas tecnologias no estado da arte para localização em ambientes internos.

3.1 Sistemas Baseados em WiFi

Os métodos usados em muitos trabalhos necessitam ou de tecnologias específicas que demandam hardware extra ou específico. O *Received Signal Strength Indicator* (RSSI) está disponível na maioria das implementações de redes sem fio, ele caracteriza a atenuação de sinais de rádio durante a propagação e foi adotado em uma grande quantidade de sistemas de localização internos [Yang et al., 2013b]. Essa métrica tem as vantagens de baixo custo, ampla cobertura e sem a necessidade de qualquer adição ou modificação de *hardware* [Kaemarungsi & Krishnamurthy, 2012].

Em RADAR [Bahl & Padmanabhan, 2000, Bahl et al., 2000], um dos primeiros sistemas de localização usando RSS e impressão digital, as leituras de RSS são feitas em diferentes pontos em o lugar onde a localização será realizada. Cada leitura contém os valores RSS recebidos dos roteadores WiFi. Essas informações são registradas em um banco de dados de impressão digital. Após este treinamento, quando um usuário solicita sua localização, os valores RSS da solicitação são comparados com aqueles

observados na fase de treinamento usando o classificador KNN, que retornará a posição mais provável.

No Sistema Horus [Youssef & Agrawala, 2004, Youssef et al., 2003], uma técnica de agrupamento conjunto é proposta para a estimativa de localização usando um método probabilístico. Neste sistema, cada coordenada possível em que o candidato pode estar é considerada uma classe ou categoria, escolhe-se o local onde sua probabilidade é maior. Aumentar o número de amostras em cada local de amostragem melhora a precisão, pois melhora a estimativa da média e desvio padrão da distribuição Gaussiana. Sistemas probabilísticos semelhantes também foram propostos em outros trabalhos [Castro et al., 2001, Pinto et al., 2021, Roos et al., 2002]. Um método baseado em treinamento para determinar a localização usando classificadores baseados em redes neurais é proposto em [Battiti et al., 2002]. Em uma abordagem mais ampla, os autores em [Saha et al., 2003] compararam os vários classificadores mencionados: redes neurais, kNN e métodos probabilísticos. Seus resultados mostram que todas as técnicas tiveram um bom desempenho em diferentes cenários.

Em [Sharma, 2006] é apresentada uma abordagem de trilateração baseada no centro de massa ponderada para localizar um dispositivo sem fio com base na intensidade do sinal recebido dos pontos de acesso em locais conhecidos. Esta abordagem consiste principalmente em duas fases: (1) O cálculo da distância dos valores RSSI de vários pontos de acesso conforme recebido pelo dispositivo móvel, (2) Determinação da localização mais provável do dispositivo sem fio usando coordenadas de vários pontos de acesso fixos ou conhecidos e distâncias calculadas do dispositivo a partir desses pontos de acesso. O filtro de Kalman também é usado em ambas as fases, a fim de remover o componente de ruído de medição e aumentar a precisão da estimativa. O algoritmo proposto fornece uma solução para rastreamento de localização de dispositivos móveis em ambiente interno, onde a configuração dos pontos de acesso, como potência de transmissão, etc., não é fixa e os movimentos no ambiente que afetam a atenuação do sinal são tão imprevisíveis que qualquer modelagem matemática do sinal de RF a propagação é inviável.

Em [Hamza, 2010] diversos experimentos foram realizados confirmando que as técnicas de localização para ambientes internos baseadas em treinamento (*fingerprinting*) têm desempenho melhor do que as tradicionais baseadas em trilateração/multilateração. O principal motivo apontado foi o fenômeno de *multipath*, em que o comportamento do sinal é afetado pelos diversos obstáculos encontrados em tais cenários. Entretanto, uma observação levantada neste trabalho é que as técnicas de localização baseada em treinamento sofrem degradações na precisão da localização em ambientes dinâmicos, onde as propriedades do local podem mudar com o tempo

(rearrançamento de armários, mesas, divisórias, etc).

Em [Wen & Liang, 2015] usa redes WiFi 802.11ac que permite o uso da tecnologia OFDM (*orthogonal frequency-division multiplexing modulation*) que permitiu o uso de CSI (*Channel State Information*) com isso podem ser obtidas informações combinadas de AoA e ToA, uma abordagem de pairing automático é proposta para identificar o caminho da linha de visada (LOS). Foram feitas simulações e um experimento em campo que obteve um RMSE (root mean squared error) de 0,4357 m.

Em [Yang et al., 2015], é utilizado o algoritmo probabilístico KNN (pKNN) que usa a probabilidade de RSS no rádio-mapa como uma ponderação para calcular a distância euclidiana. Esta solução também filtra os valores RSS cuja probabilidade é inferior a 3 %. Um experimento foi feito usando uma rede sem fio com 4 APs e um dispositivo móvel. Para um valor de $K=4$ o algoritmo p-KNN pode atingir um erro menor que 2 metros em mais de 80% das vezes. Também pode melhorar a probabilidade cumulativa em até 9% na faixa de 0,5 m a 2 m em comparação com o algoritmo KNN.

3.2 Sistemas Baseados em UWB

Como a tecnologia *Ultrawide bandwidth* (UWB) permite ajustes finos no atraso de envio, várias técnicas de localização baseadas em UWB utilizam ToA (*Time Of Arrival*) para medir o alcance entre um receptor e um transmissor. Nas técnicas baseadas em ToA a distância do alvo móvel para a medição unidade é diretamente proporcional ao tempo de propagação do sinal [Liu et al., 2007], o que o torna a tecnologia UWB ideal para essa medição.

Em [Dardari et al., 2008] usando UWB e ToA são feitos modelos para a estimativa do intervalo de erro e o excesso de atraso com base em dados medidos de alcance real dos dispositivos. Esses modelos são usados em vários algoritmos de multilateração para determinar a posição do alvo. Usando medições em um cenário interno real, é investigado como a precisão da localização é afetada pelo número de beacons e pela disponibilidade de informações a priori sobre o ambiente e a geometria da rede. Também é avaliado como vários alvos cooperam medindo intervalos não apenas dos beacons, mas também uns dos outros. Um algoritmo de multilateração iterativo que incorpora informações coletadas através da cooperação é então proposto com o objetivo de melhorar a precisão da localização. Usando resultados numéricos, é demonstrado o impacto da cooperação na precisão da localização.

Em [Conti et al., 2008] é investigado usando UWB e ToA como a precisão da localização é afetada pelo número de beacons e pela disponibilidade de informações a

priori sobre o ambiente e a geometria da rede. Também é examinado o caso em que vários alvos estão presentes no mesmo ambiente e cooperam medindo faixas não apenas dos beacons, mas também uns dos outros. Um algoritmo iterativo de multilateração que incorpora informações coletadas por meio da cooperação é então proposto com o objetivo de melhorar a precisão da localização.

Uma variação do ToA chamada de diferença de tempo de chegada (TDoA) explora a diferença nos tempos de propagação dos sinais de diferentes transmissores, medidos no receptor [Zafari et al., 2019]. O UWB é particularmente adequado para sistemas de localização para ambientes fechados que usam TDoA como em [Yuqiang Zhang & Junhui Zhao, 2005, Zhang et al., 2006].

Outra técnica usada é o *Angle of Arrival* (AoA), também conhecido como direção de chegada (DoA), refere-se ao ângulo de incidência em que os sinais de rádio de um transmissor chegam ao receptor. Dadas as posições de dois AoAs e de seus receptores, as coordenadas bidimensionais do transmissor podem ser determinadas por triangulação. Para usar AoA é necessário hardware extra como antenas direcionais ou um conjunto de antenas, o que torna o receptor mais complexo e caro [Hou et al., 2018].

3.3 Sistemas Baseados em Infravermelho

A comunicação usando infravermelho é relativamente direcional, econômica e fácil de montar, mas têm uma desvantagem importante, não atravessa objetos sólidos. Isso pode ser visto como uma qualidade já que um sistema infravermelho instalado em um ambiente fechado não interfere em um sistema semelhante instalado nas salas ou nos prédios adjacentes [Tanenbaum, 2002].

Em [Mukhopadhyay et al., 2018] é proposta uma técnica que utiliza a saída analógica de sensores infravermelhos piroelétricos (ou passivos) (PIR) para localização em ambientes fechados. Foram propostos dois modelos (hiperbólico e um modelo linear por partes) para estimativa de distância com base na caracterização da saída do sensor analógico em termos do valor pico a pico e sua relação com a distância do sensor. Uma vez que as distâncias do alvo para os sensores PIR (ou âncoras) são estimadas, técnicas baseadas em multilateração ou regressão de vetor de suporte (SVR) são usadas para calcular as coordenadas de localização. Usando quatro sensores PIR, é demonstrada a localização de uma pessoa em uma área de $7\text{ m} \times 7,5\text{ m}$ com a técnica baseada em regressão superando as outras técnicas em termos de precisão e alcançando um erro de localização com RMS de 0,65 m. Também foi comparado a complexidade computacional e os requisitos de memória ou armazenamento das técnicas propostas, o que

é importante para implementação distribuída em dispositivos com recursos limitados, como nós sensores.

Em [Arbula & Ljubic, 2020] é apresentado um novo método que utiliza sinal infravermelho (IR) de baixo alcance na linha de visão (LOS) para fornecer estimativa de ângulo de chegada (AoA) de alta precisão. A proposta foi testada no contexto desafiador da navegação de um carrinho de supermercado. Os resultados da avaliação variam de precisão centimétrica alcançada no contexto estático-1D até 1 m de erro médio de localização obtido para um carrinho móvel se movendo a 140 cm/s em uma configuração 2D. Esses resultados mostram que, para o contexto de supermercados, a precisão de localização adequada pode ser alcançada, juntamente com o tempo real suporte à navegação, usando tecnologia IR prontamente disponível com componentes de hardware baratos.

3.4 Sistemas Baseados em Zigbee

O ZigBee é uma tecnologia com baixo consumo de energia ($1mW$ ou menos), mas que ainda fornece alcance de até 150 metros em ambientes externos [Ramya et al., 2011].

Em [Yang et al., 2020] é proposto um algoritmo de trilateração para localização indoor com base no RSSI. Em primeiro lugar, todos os dados de medição brutos são pré-processados por um filtro Gaussiano para reduzir a influência do ruído de medição. Em segundo lugar, a potência de transmissão e o expoente de perda de caminho são estimados por um novo método de ajuste de curva de mínimos quadrados (LSCF) na localização baseada em RSSI. Em terceiro lugar, um novo algoritmo de trilateração é proposto com base na teoria de valores extremos, que constrói uma função de erro não linear dependendo das distâncias e da posição dos nós âncora. A validação foi feita usando tanto simulação como com o uso de módulos de comunicação sem fio CC2530 que usam a tecnologia Zigbee.

Em [Kimoto et al., 2019] é apresentado o MuCHLoc, um sistema de localização usando sensores e impressão digital que melhora a precisão da localização utilizando a diversidade de canais. Usando a extração da diversidade de canais do RSS de pontos de acesso Wi-Fi (APs) medidos em múltiplos canais ZigBee por meio da localização usando impressão digital. Com isso aumentando a estabilidade do RSS aumentando as dimensões das impressões digitais. Foram conduzidos experimentos coletando RSS de APs Wi-Fi em um ambiente prático durante a troca de canais ZigBee e avaliados a precisão da localização. As avaliações revelaram que o MuCHLoc melhora a precisão da localização em aproximadamente 15% em comparação com a localização usando um

único canal. Também mostramos que MuCHLoc é eficaz em um ambiente de rádio dinâmico onde o canal de propagação de rádio é instável devido ao movimento de objetos, incluindo humanos.

3.5 LoRa (Long Range)

A tecnologia de rádio LoRa tem sido muito utilizada devido às suas características técnicas versáteis e robustas, juntamente com sua capacidade de alcançar longos alcances de comunicação combinados com baixo consumo de energia e baixo custo [Gkotsiopoulos et al., 2021].

No trabalho de [Lam et al., 2019] que sugere que é um dos pioneiros em usar tecnologia LoRa para localização com algoritmos de localização baseados em RSSI e redução de efeito do ruído gaussiano e não gaussiano durante a localização. A análise de desempenho dos seus modelos de simulação e experimentos reais mostram que os algoritmos de localização propostos funcionam adequadamente em ambientes externos e internos de grande escala. Em ambientes externos, seu desempenho é comparável ao GPS.

Em [Kim et al., 2021] são realizados experimentos para demonstrar a capacidade do LoRa para localização interna em um ambiente de apartamento. Os resultados experimentais mostram que a localização interna baseada em LoRa tem uma precisão melhor que 1,6 m no cenário de linha de visão e 3,2 m no cenário extremo sem linha de visão com uma precisão melhor que 25 cm em todos os casos, sem usar qualquer filtragem de dados nas estimativas de localização.

Em [Micheletti & Godoy, 2022] é apresentado um sistema de localização 3D aprimorado baseado em comunicação LoRa e altímetro para localização em ambientes fechados. O método utiliza a técnica tradicional de trilateração com o indicador de intensidade do sinal (RSSI) com melhoramentos experimentais dos parâmetros para melhorar a precisão da localização. Testes experimentais realizados em dois cenários indoor mostraram uma melhora significativa de até 80%, com intervalo de confiança (95%) de 0,121m, na precisão da localização 3D, além de resultados melhores que trabalhos relacionados na literatura.

3.6 Sistemas Baseados em Bluetooth Low Energy (BLE)

Com o surgimento do Bluetooth *Low Energy* (BLE), foram propostas tecnologias para localizar um objeto usando *beacons* com essa tecnologia. Apple e Google propuseram, respectivamente, dispositivos iBeacons e Eddystone [Bin Aftab, 2017].

Em [Martin et al., 2014] apresentam um conjunto de ferramentas de localização desenvolvidas em torno do protocolo iBeacon, fornecendo uma visão aprofundada da viabilidade do Bluetooth *Low Energy* como uma tecnologia de posicionamento para ambientes fechados. Neste sistema um dispositivo móvel ouve os *broadcasts* dos *beacons* Bluetooth LE em um serviço de segundo plano, estimando e retransmitindo estimativas de distância do *beacon* agregadas e informações de identificação do *beacon* correspondentes para um servidor central em intervalos fixos. Esse servidor central irá enviar a informação da localização baseadas em um modelo usando os *beacons* e as distâncias desse device para os *beacons*. O sistema obtém um erro médio de estimativa de posição de 0,53 metros.

Em [Motohashi et al., 2017] foi desenvolvido um aplicativo usando iBeacon para capturar o movimento físico em ambientes fechados do usuário e coletar seu *feedback* temporal. Com isso foi desenvolvido um método de previsão comparando os dados comportamentais espaciais com as respostas explícitas do usuário para definir pontos de interesse. Com isso, a localização e a duração dos pontos que um usuário visitou são usadas pelo modelo para determinar os pontos de interesse. Os resultados em relação às respostas explícitas do usuário alcançaram uma precisão de cerca de 90%).

Em [Campaña et al., 2017] é usado um sistema de navegação interna baseado em impressão digital usando 30 *beacons* Eddystone Bluetooth *Low Energy* (BLE) no Centro de Tecnologias da Informação da Universidade ESPOL. Foi usada uma arquitetura centralizada com um servidor responsável por salvar e processar as impressões digitais, para que possam ser usadas em um algoritmo de aprendizado de máquina para prever um local. Este artigo apresenta um estudo comparativo de três técnicas de aprendizado de máquina usadas para melhorar a precisão da impressão digital BLE. Os resultados mostraram que Random Forest é 30% mais preciso do que *Naive Bayes*, é capaz de estimar corretamente o localização da sala de vários usuários com precisão de 91%.

Em [Sun et al., 2021] é proposto um método para localizar um usuário em movimento usando sensores BLE em ambientes fechados e complexos. Foi desenvolvido um modelo de localização baseado em Rede Neural Convolutiva (CNN) baseado na imagem 2D composta pelo RSSI recebido dos eixos x e y. Os experimentos foram con-

duzidos em um ambiente de construção onde treze dispositivos beacon foram instalados em diferentes locais para registrar as coordenadas. Para comparação foram adotados algoritmos de aprendizado de máquina e aprendizado profundo para prever a localização de um usuário em um ambiente fechado. Os resultados experimentais indicam que o método baseado em CNN otimizado proposto mostra alta precisão (97,92% com erro de 2,8%) para rastrear a localização de um usuário em movimento em um edifício complexo sem calibração complexa em comparação com outros métodos.

Em [Nguyen & Thuy Le, 2021] usa BLE com impressão digital baseada em RSSI usando apenas vetores RSSI sem nenhum conhecimento prévio. Foi usado Regressão de Processo Gaussiana (GPR) combinado com o algoritmo kNN modificado para melhorar o desempenho. Na fase *online* do sistema, o GPR dá uma previsão para a localização usando o vetor RSSI. Esta previsão ajuda o algoritmo kNN, portanto, limita a região de busca, levando a reduzir o custo computacional. Também é apresentada uma análise sobre a distribuição dos k pontos mais próximos que avalia a confiança da localização estimada para a extração de uma lista de pontos confiáveis. Além disso, também é apresentado um processo de extrapolação usando esta lista de confiança para ter uma trajetória otimizada. A análise de precisão e tempo da proposta foi realizada usando a base de dados BBIL (BLE Beacon Indoor Localization Dataset) que contém sinais RSSI muito ruidosos devido ao movimento rápido dos objetos. Os resultados experimentais mostram que o sistema apresenta um desempenho muito melhor do que os algoritmos tradicionais kNN ou WkNN. O RMSE da trajetória ótima é de 1,78 m na sala com uma dimensão de 10 m \times 25 m, o que é bom em comparação com outros métodos onde a localização inicial é conhecida.

3.7 Detecção de Ruído

Devido a fenômenos conhecidos como multicaminhos, efeitos de sombra e algumas restrições de *hardware*, variações inesperadas nas medições RSS não são incomuns. Se a medição estiver contaminada com algum tipo de ruído, seja na fase *offline* ou *online*, a estimativa de posição resultante pode ter seus erros de localização aumentados Chen & Juang [2012].

Em [Ansari et al., 2007] são usados filtros Bayesianos não lineares para obter resultados de rastreamento precisos de dados de localização com ruído. Isso permite o rastreamento de alta precisão de movimento não linear, mesmo na presença de ruído de medição não gaussiano, o que é uma melhoria significativa em geral em comparação com muitos sistemas de localização de rede de sensores de última geração. É usado também

uma combinação de sinais de rádio e ultrassom para obter estimativas de distâncias necessárias para o processo de filtragem, embora em princípio qualquer técnica de estimativa de distância possa ser usada. Também são descritos métodos aprimorados de detecção de outlier e pós-calibração para aprimorar a qualidade das estimativas de distância obtidas em comparação com sistemas anteriores.

Em [He & Chan, 2016] para reduzir os efeitos do ruído é proposto um sistema de localização chamado de Tilejunction, com base na média e variância do sinal medido ele mapeia o RSSI do alvo em cada AP para uma estrutura denominada de "Tile" (calculado com algoritmos de geometria computacional, e é representado por um conjunto de restrições lineares) onde o alvo provavelmente está dentro. Usando uma nova métrica de comparação de sinais aleatórios, foi formulado um problema de programação linear (PL) para localizar o alvo na junção dos 'Tiles'. Para melhorar ainda mais a eficiência computacional, o Tilejunction emprega uma medida de teoria da informação para usar apenas os APs cujos sinais mostram diferenciação suficiente naquela região. Ele também particiona a região em vários *clusters* para reduzir substancialmente o espaço de pesquisa na otimização de LP. As simulações e medições experimentais mostram que ele supera outras abordagens recentes de última geração (por exemplo, RADAR, divergência KL, etc.) com erro de localização significativamente menor (geralmente em mais de 30%).

Em [Chen & Wang, 2018] uma nova rede neural de regressão generalizada de largura de banda múltipla (GRNN) com a abordagem de posicionamento interno de filtro de *outlier* (GROF) é proposta. O método GROF é baseado no GRNN, para o qual é adotado um novo tipo de arquitetura de kernel de largura de banda múltipla para alcançar um desempenho de regressão mais flexível do que o GRNN tradicional. Além disso, um esquema de filtragem de *outlier* que adota o método do vizinho mais próximo (kNN) é incorporado ao módulo de localização de modo a melhorar a robustez da localização contra mudanças ambientais.

Em [Bhatti et al., 2020] foi desenvolvida uma técnica de detecção de ruído denominada iF_Ensemble para ambientes de localização indoor usando Wi-Fi, analisando RSSs usando a combinação de métodos de aprendizado de máquina supervisionado, não supervisionado e ensemble. Nesta pesquisa, a floresta de isolamento (iForest) é usada como um método de aprendizagem não supervisionado. O método de aprendizado supervisionado inclui classificadores SVM, KNN e floresta aleatória (RF).

LocFiND-Localization Using Fingerprinting and Novelty Detection

Neste capítulo, apresentamos nosso método proposto chamado LocFiND (*Localization Using Fingerprinting e Novelty Detection*). Propomos 3 variações do algoritmo, uma simples sem usar a novidade e 2 mais complexas que usam a novidade para refinar os resultados.

O principal objetivo do trabalho proposto é fornecer a localização de indivíduos em ambientes fechados para seu monitoramento. Nesse tipo de ambiente é esperado que a solução tradicional (GPS) acabe por não funcionar muito bem [Basiri et al., 2017]. Nessas aplicações é comum que o indivíduo a ser localizado carregue algum tipo de dispositivo conhecido como nó móvel que possa ser usado para estimar sua localização.

Outra característica que deve ser levada em consideração deve ser que a solução não seja muito invasiva, cada indivíduo carrega um dispositivo de pequeno porte com bateria que serve de nó móvel. Para o monitoramento, os nós móveis enviam as informações para dispositivos âncora de modo que não haja interferência na rede sem fio de dados do local.

Como em outras soluções, LocFiND pode ser dividido em duas fases distintas, conforme ilustrado na Figura 4.1. Na primeira fase de treinamento, são coletadas amostras para treinar o algoritmo. O banco de dados de impressão digital obtido segue o formato descrito abaixo:

$$A_i = \{RSS_{i,1}, RSS_{i,2}, \dots, RSS_{i,n}, CLASS_i\} \quad (4.1)$$

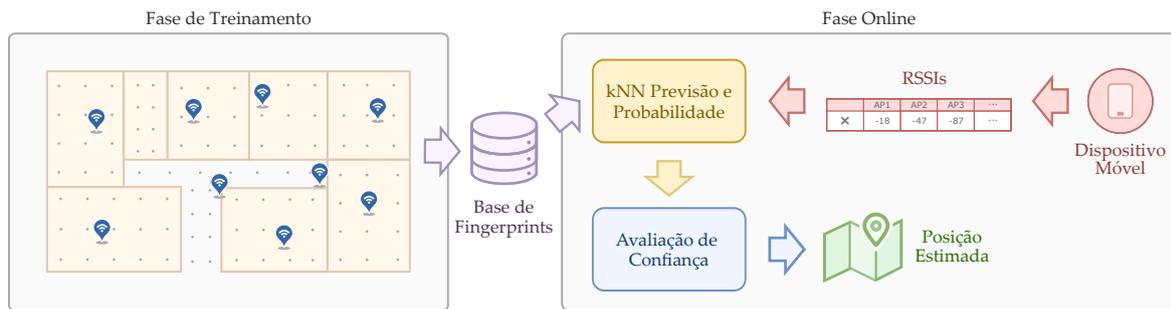


Figura 4.1: Arquitetura do LocFiND: a fase de treinamento gera o banco de dados de impressão digital, que é utilizado na fase online para estimar a posição dos dispositivos móveis.

onde A_i é o conjunto de pacotes coletados i , $RSS_{i,n}$ é o RSSI do pacote i recebido no nó âncora n , e $CLASS_i$ é a classe (rótulo) que pode ser mapeado para o local em que o pacote foi enviado. Todos esses dados serão usados pelo algoritmo de treinamento supervisionado para gerar um modelo.

Para o rótulo das amostras foram utilizadas 2 granularidades. Para a classificação de sala $CLASS_i = \{41, 57, 40, 53, 39, 47, 51, 62, 45, 43, 61, 46, 16, 63\}$, onde cada número representa o ID de uma sala. Para a classificação de pontos específicos dentro das salas, os rótulos têm o formato R_X_Y que representa R como o ID da sala, X como a linha e Y como a coluna dentro da sala (com células de aproximadamente 2×2 m).

Na segunda fase, online, são executados os procedimentos para estimar a localização dos dispositivos móveis. Uma arquitetura centralizada foi escolhida principalmente devido às limitações dos nós móveis (por exemplo, economia de energia e *hardware* limitado). Nesta fase, um nó móvel envia um pacote de dados que será recebido pelos nós âncora. Esses nós âncora medirão seu próprio RSSI para este pacote e encaminharão as informações para um servidor central. O servidor poderá então estimar o local em que o pacote foi enviado e, portanto, a localização do dispositivo móvel.

Nesta fase, o algoritmo kNN, utilizando o banco de dados de impressões digitais da fase de treinamento, retornará não apenas a classe, mas também as probabilidades da amostra pertencer a cada uma das diferentes classes. Essas probabilidades serão então avaliadas por nossas abordagens propostas para avaliação de confiança para a classificação final.

Conforme mencionado, os valores de RSSI apresentam uma grande variação mesmo quando o nó é estático. Isso se deve a muitas características e obstáculos que podem alterar o sinal momentaneamente causando erros tanto na fase de treinamento quanto na fase *online*. Essas variações inesperadas nas informações são tratadas como

ruído e geralmente são detectadas na fase de treinamento.

Muitos trabalhos na literatura tratam do ruído RSSI na fase de treinamento [Swangmuang & Krishnamurthy, 2008], mas a informação do sinal na fase *online* sofre do mesmo problema ou pior, já que na maioria dos casos, o treinamento é feito em um prédio vazio enquanto a fase online ocorre em um cenário mais populoso e dinâmico. Para piorar a situação, em alguns cenários como o nosso, não podemos calcular a média de várias amostras na fase *online*, pois os pacotes são enviados por dispositivos móveis com restrição de energia. Assim, precisamos de soluções que também levem em consideração o impacto dessa variação do RSSI na fase *online*, sem exigir o envio de pacotes de dados adicionais.

Uma etapa intuitiva seria descartar e ignorar dados não confiáveis no momento da localização. Porém, após algum tempo sem informações de localização, essas informações ausentes começarão a ser perceptíveis. Portanto, não podemos simplesmente começar a ignorar qualquer informação que possa não ser confiável. Resta saber como definir, na fase online, se um conjunto de RSSI's recebidos pode ser considerado confiável para uma correta classificação. Nos referimos a esta etapa como *Avaliação de confiança* em nossa arquitetura LocFiND.

Como mencionado antes, propomos três técnicas diferentes no LocFiND para incluir esta avaliação de confiança para as posições estimadas:

1. *Probabilidade da previsão*: usa a probabilidade de previsão estimada para decidir se aceita a classe estimada.
2. *Deteção de Novidade*: usa a técnica de deteção de novidades para identificar a confiabilidade da amostra classificada.
3. *Abordagem Híbrida*: usa deteção de probabilidade e novidade para melhorar as soluções isoladas.

Essas três técnicas são explicadas e detalhadas nas próximas seções.

4.1 Usando Probabilidade de Previsão do kNN

Na primeira técnica, chamada *Probabilidade de Predição*, avaliamos a confiança sobre a posição estimada usando a probabilidade de classificação, disponível na maioria das implementações kNN. Por exemplo, na implementação do scikit-learn, a função *predict_proba* retorna as probabilidades que uma amostra possui de pertencer a cada uma das diferentes classes (posições treinadas, rótulos). A classe mais provável é o resultado da classificação. Para a estimativa de probabilidade, os preditores geralmente

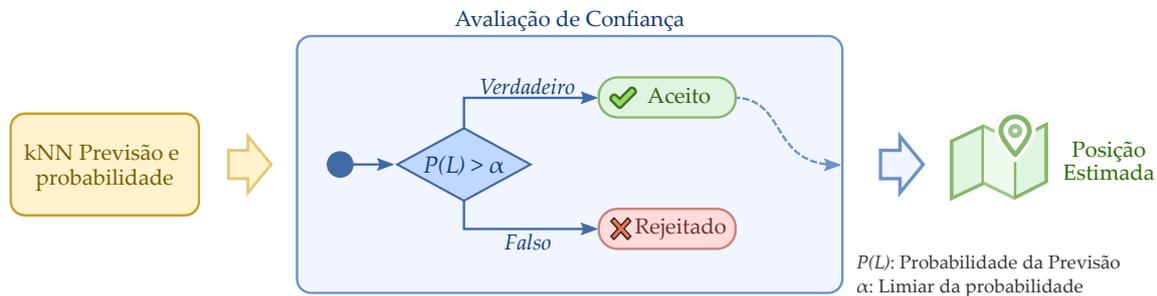


Figura 4.2: Avaliação de confiança usando *Probabilidade de Predição*: classificações com baixas probabilidades são descartadas para melhorar a precisão.

implementam métodos que quantificam a confiança das previsões. No kNN, as probabilidades de cada classe são a média ponderada normalizada dos indicadores para as k-classes mais próximas, ponderada pela distância inversa [Buitinck et al., 2013].

A abordagem *probabilidade de previsão* usa o modelo gerado pelo kNN usando o banco de dados de impressão digital da fase de treinamento. O modelo avalia cada nova amostra online recebida e retorna as probabilidades de pertencer a cada uma das classes disponíveis. Nessa abordagem, escolhemos a classe com maior probabilidade e, se essa probabilidade for maior que um limite α (uma variável que pode ser ajustada), a classificação é aceita e a nova posição estimada é enviada para a aplicação. Caso contrário, o resultado é rejeitado. A Figura 4.2 mostra o diagrama completo para esta avaliação de confiança.

4.2 Usando Detecção de Novidade

Na solução anterior, como mostrado na Seção 4.1, mesmo ao descartar as classificações que têm uma probabilidade de predição baixa, ainda temos algumas classificações erradas. Nesta seção, definimos uma nova abordagem que ajuda a decidir quais classificações devem ser descartadas. Nesta solução, a que nos referimos simplesmente como *Detecção de Novidade*, verificamos (ou seja, confirmamos) se a amostra classificada realmente pertence à classe que foi estimada pelo classificador (o kNN escolhe a classe que mais se aproxima). Para isso, utilizamos o conceito de novidade, conceito utilizado na área de detecção de *outliers*. A Detecção de Novidade pode ser definida como a tarefa de reconhecer se uma amostra testada difere, em alguns aspectos, dos dados coletados durante a fase de treinamento [Pimentel et al., 2014].

Detectar amostras novas e nunca vistas é uma habilidade importante para qual-

quer sistema de classificação. Sabendo que não é possível treinar um sistema de aprendizado de máquina com todos os tipos de dados que o sistema provavelmente encontrará, torna-se importante ser capaz de diferenciar entre informações de dados conhecidos e desconhecidos durante a fase online [Markou & Singh, 2003].

A detecção de novidade refere-se à identificação de novos padrões ou padrões anormais embutidos em uma grande quantidade de dados normais. A novidade (anomalia, *outlier*, exceção) acontece quando uma amostra não está de acordo com o comportamento normal esperado [Miljković, 2010]. Ao contrário das abordagens de detecção de *outliers*, que são feitas no treinamento, a detecção de novidade define se uma nova amostra pertence ou não a uma classe na fase *online*. Uma arquitetura de detecção de novidades é ilustrada na Figura 4.3.

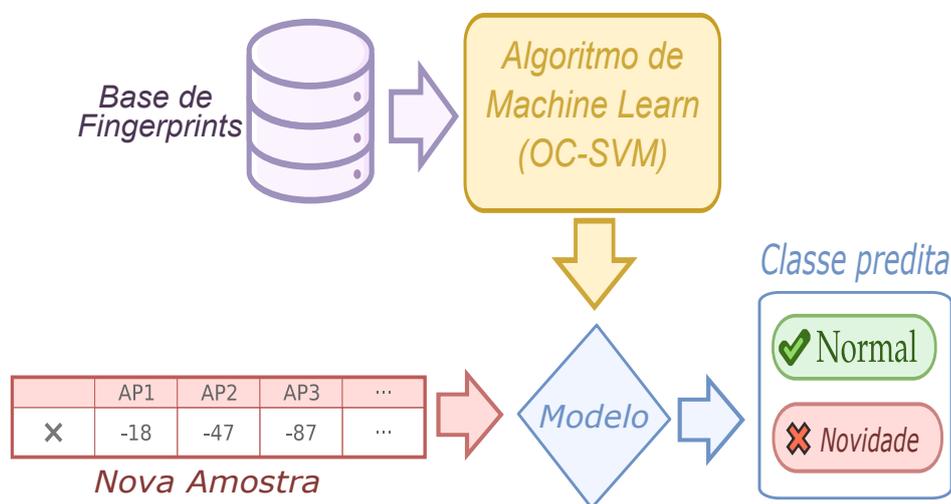


Figura 4.3: Detecção de novidade usando One-Class SVM.

Técnicas de detecção de novidades podem ser implementadas usando diferentes técnicas, como as baseadas em máquinas de vetor de suporte, baseadas em *clustering*, baseadas em vizinho mais próximo e as estatísticas [Sayyed & Deolekar, 2017].

Uma técnica de detecção de novidades não supervisionada comumente usada é baseada em SVM. Ele constrói um esquema de limite esférico que agrupa e inclui a maioria das amostras de dados normais. Qualquer amostra nova e não vista que caia fora dos limites do grupo é detectada como novidade. Em trabalhos recentes sobre detecção de novidades baseada em SVM, uma variação chamada OC-SVM (*One-Class SVM*) é usada uma vez que tem um truque de kernel que permite separar facilmente dados normais e novos em espaços dimensionais grandes [Ding et al., 2014], que é o caso do nosso sistema de localização sistema com 15 dimensões (ou seja, nós âncora).

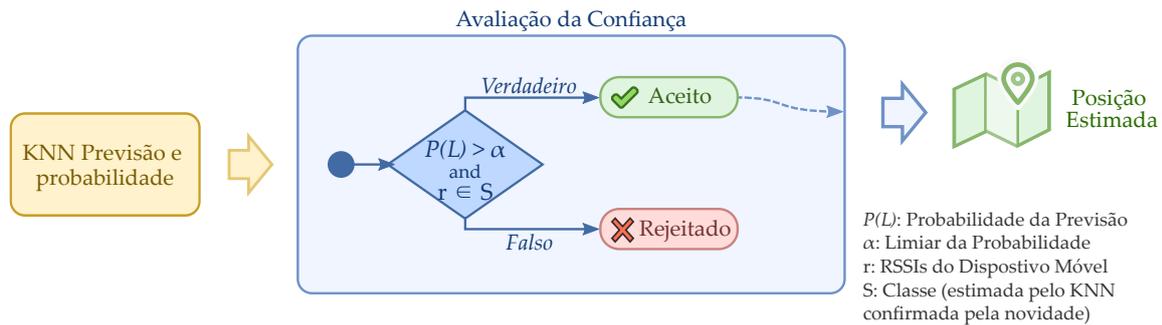


Figura 4.4: Avaliação de confiança usando a técnica *Detecção de Novidade*: a precisão é melhorada com o descarte de classificações não confiáveis.

Nesta variação do LocFiND, primeiro usamos a probabilidade de previsão, como na seção anterior 4.1, para ter uma primeira avaliação do próprio kNN em relação à confiabilidade da classificação. Então, para ter uma segunda validação de resultado, usamos o conceito de detecção de novidade para confirmar que uma determinada amostra realmente pertence àquela classe.

Assim, a mesma probabilidade de previsão explicada anteriormente com o limite α é usada. No entanto, as classificações são posteriormente validadas usando a detecção de novidades implementada usando o One-Class SVM. Além de passar no teste de limite, uma nova validação é feita pelo algoritmo de novidade que testa se aquela amostra realmente pertence à classe escolhida (aquela com a maior probabilidade retornada pelo kNN). Somente quando ambos os testes são aprovados, uma classificação é aceita. Este processo de decisão completo para a avaliação da confiança é ilustrado na Figura 4.4.

4.3 Abordagem Híbrida: Combinando Probabilidade de Previsão com Detecção de Novidade

Para entender a *abordagem híbrida*, precisamos primeiro entender o comportamento do limite de probabilidade α . Quanto mais baixo for o limite de probabilidade, maiores serão os erros de classificação, resultando em um erro de localização médio mais alto. No entanto, neste caso, também temos um número maior de estimativas de posição, o que é bom para a aplicação. Se aumentarmos esse limite, melhoramos a precisão, mas também reduzimos o número de classificações aceitas, deixando a aplicação sem estimativas de posição.

Na *abordagem híbrida*, inicialmente usamos o mesmo limite de probabilidade α que nas seções anteriores. No entanto, em vez de simplesmente rejeitar as classificações que não passam neste filtro, usamos a detecção de novidades para dar uma segunda chance para algumas dessas classificações que seriam descartadas de outra forma. Para isso, usamos um segundo limite de probabilidade inferior β e usamos a detecção de novidades para confirmar as classificações acima desse segundo limite.

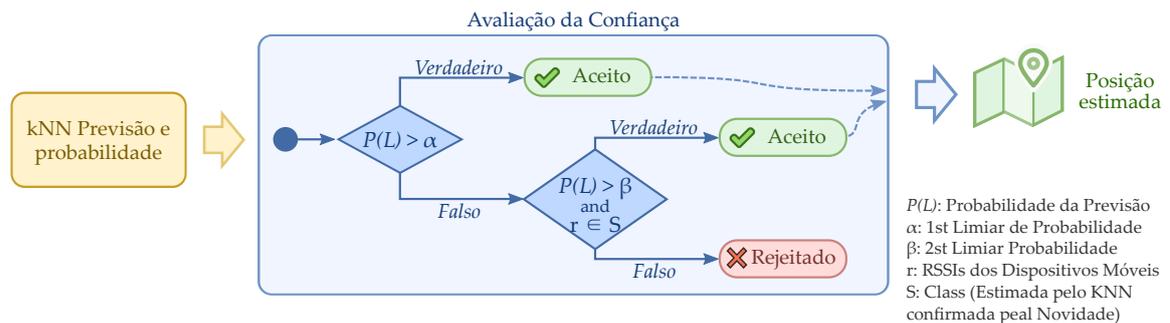


Figura 4.5: Avaliação de confiança usando a *Abordagem Híbrida*: um meio-termo entre a precisão e o número de estimativas de posição.

O resultado desse processo de decisão, que é ilustrado na Figura 4.5, é que as classificações com altas probabilidades de previsão (limite α) não são verificadas pela detecção de novidade. Classificações com probabilidades de previsão mais baixas, mas ainda mais altas do que o limite β , serão confirmadas pela detecção de novidade. Por fim, as classificações com probabilidades de previsão inferiores a β ou que não passaram na verificação de novidade são descartadas.

Avaliação de Performance

Nesta seção, avaliamos o desempenho de nossa solução LocFiND proposta e suas abordagens para a *avaliação de confiança* das posições estimadas. Discutiremos os resultados de nossa solução comparados de várias maneiras, para ter uma melhor visão de seu funcionamento e um melhor entendimento de seu comportamento.

5.1 Metodologia

Avaliamos o desempenho de nossa solução com uma base coletada em um ambiente de teste do mundo real, mostrado na Figura 5.1. Como nosso foco é em aplicações baseadas em dispositivos móveis pequenos, de baixo processamento, de baixo consumo de energia e operados por bateria, usamos nós e âncoras com a tecnologia Bluetooth Low Energy (BLE). A Figura 5.2 mostra o *hardware* usado para os dispositivos móveis, enquanto a Figura 5.3 mostra o *hardware* usado como nós âncora para a coleta de dados de localização.

Como nossa principal aplicação é localizar alunos, nosso ambiente de teste foi implementado em uma área escolar de $645 m^2$ consistindo em 15 espaços (11 salas e 3 corredores). O mapa completo da área está representado na Figura 5.1. Para cobrir toda a área, 15 nós de ancoragem foram instalados no teto das salas em locais onde era de certa forma conveniente conectá-los à rede elétrica. Esses nós âncora são representados em azul na Figura 5.1.

Para executar a fase de treinamento do LocFiND e gerar o banco de dados de impressão digital, foram distribuídos uniformemente 150 pontos de referência (RPs) em toda a área. Esses RPs são descritos como pontos cinza na Figura 5.1. Para cada RP, fomos a esse local e começamos a enviar pacotes de dados usando 11 dispositivos móveis diferentes. Coletamos um total de 100 amostras por RP para um total de 15.000

amostras para cada dispositivo móvel em nosso banco de dados de impressão digital. Finalmente, em nossos resultados mostrados nas seções a seguir, usamos as amostras de 8 dispositivos móveis como treinamento e as amostras de 3 outros dispositivos móveis como teste. Assim, as amostras usadas para teste são de dispositivos nunca vistos no treinamento.



Figura 5.1: Mapa do nosso testbed: 11 salas, 3 corredores, 15 nós âncora e 150 pontos de referência. Para cada ponto de referência, 100 amostras foram coletadas em um total de 15.000 amostras para cada dispositivo móvel. Foram usados 11 dispositivos móveis diferentes, 8 para treinamento e 3 para teste.



Figura 5.2: Dispositivos do tipo *smartwatch* baseados em Bluetooth que foram usados como dispositivos móveis.

Para escolher os parâmetros de kNN e OC-SVM (usados para detecção de novidades) foi usado o método *GridSearchCV* do *scikit-learn*, que é uma seleção de parâmetros



Figura 5.3: Os pontos de acesso baseados em Bluetooth que foram usados como nós âncora.

usando validação cruzada na base de treinamento, evitando assim *overfitting*. Usando *GridSearchCV* com 5 *folds*, definimos os parâmetros principais em kNN o valor de $k = 9$ e OC-SVM o valor $\gamma = 0,0025$. Por fim, também usamos diferentes dispositivos para treinamento e teste, de modo que os valores em teste sejam de dispositivos nunca vistos. Isso aumenta as chances de a solução proposta funcionar com novos dispositivos em um aplicativo do mundo real.

Como base de comparação para todos os resultados, uma solução tradicional baseada em impressão digital que usa o banco de dados RSSI e kNN (com $K = 9$) resultou em uma precisão de 96% para a classificação da sala e uma precisão de 36% para classificação no ponto de uma sala e um erro médio de 1,9 m. Claro, como neste caso não há avaliação de confiança, 100% das estimativas de posição foram enviadas para a aplicação. Nosso objetivo para LocFiND é melhorar as classificações e o erro médio, entregando o máximo possível de estimativas de posição para a aplicação.

5.2 Avaliação de Confiança Usando Probabilidade de Previsão

Conforme mencionado na Seção 4.1, a avaliação de confiança usando *probabilidade de previsão* usa o limite α para decidir se aceita ou descarta uma classe/posição estimada com base na probabilidade da amostra pertencer a essa classe. Somente amostras com probabilidade de predição superior a α são aceitas.

A Figura 5.4 mostra a relação entre o valor do limite α e a precisão final das classificações aceitas (em vermelho), bem como a porcentagem de estimativas aceitas

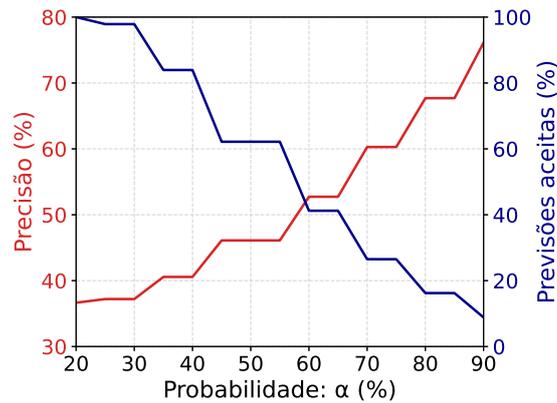


Figura 5.4: Precisão da localização de ponto usando *probabilidade de previsão*.

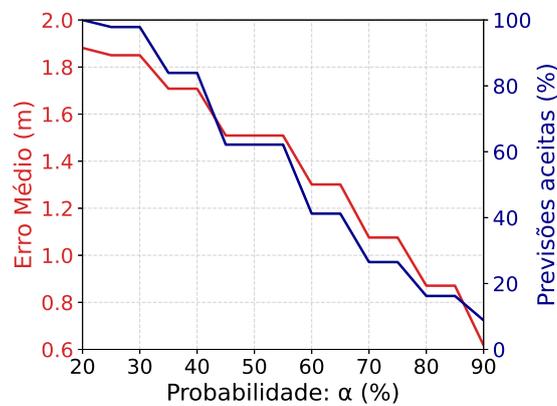


Figura 5.5: Erro médio da localização usando *probabilidade de previsão*.

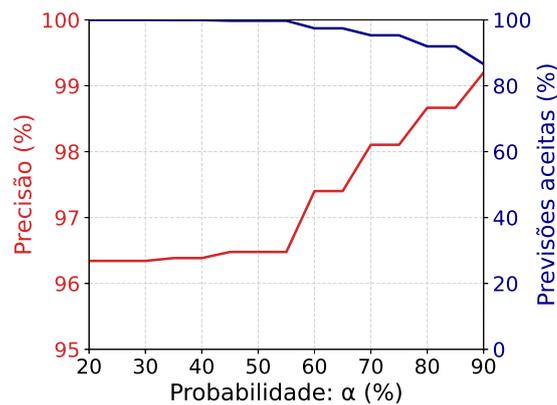


Figura 5.6: Precisão da localização de sala usando *probabilidade de previsão*.

(em azul). Da mesma forma, a Figura 5.5 mostra o comportamento do erro médio de localização (em vermelho) e, novamente, a porcentagem de estimativas aceitas (em azul) à medida que aumentamos o limite α . Finalmente, a Figura 5.6 mostra o comportamento da precisão da estimativa da sala, ou seja, estamos usando o nome da

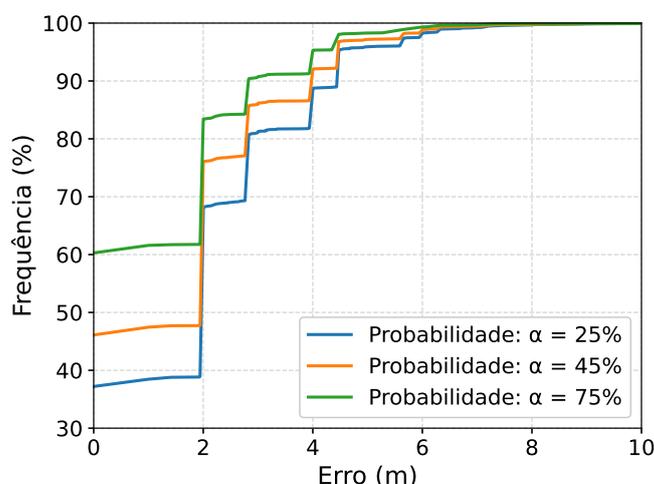


Figura 5.7: Função de Distribuição Acumulada do erro de localização usando *probabilidade de previsão* com limites $\alpha = 25\%$, 45% e 75% : mostra a porcentagem de estimativas (eixo y) com um erro menor do que o eixo x.

sala como classes. Esses resultados mostram várias características de nossa solução LocFiND proposta que discutiremos nos próximos parágrafos.

Como esperado, conforme aumentamos o limite α , o número de estimativas aceitas diminui (curva azul nas Figuras 5.4, 5.5, e 5.6). No entanto, uma vez que estamos aceitando previsões de qualidade superior, a precisão aumenta (Figura 5.4) e, assim, o erro médio diminui (Figura 5.5) e a precisão da sala também aumenta (Figura 5.6).

Por exemplo, em comparação com um kNN tradicional, no caso das Figuras 5.4 e 5.5, um limite $\alpha = 40\%$ aceita 85% das previsões enquanto aumenta a precisão de 36% para 41% e diminui o erro médio de localização de 1,9 m para 1,7 m. No caso da precisão da sala, para $\alpha = 40\%$, o número de previsões aceitas permaneceu quase o mesmo em 99,6%, enquanto aumentou apenas ligeiramente a precisão para 96,4%. Isso ocorre principalmente porque é mais fácil estimar salas e o classificador é capaz de fazer isso com probabilidades mais altas.

Por outro lado, quando usamos um limite $\alpha = 80\%$, apenas 20% das estimativas de posição foram encaminhadas para a aplicação, conforme ilustrado nas Figuras 5.4 e 5.5. No entanto, essas estimativas eram muito confiáveis, tendo quase 70% de precisão e um erro de localização inferior a 0,9 m. Um resultado melhor veio da precisão de sala, em que 92% das previsões foram aceitas com uma precisão de quase 99%.

Para entender melhor como os erros de localização são distribuídos, a Figura 5.7 mostra a função de distribuição cumulativa para os limites $\alpha = 25\%$, 45% e 75% . Este gráfico mostra a porcentagem de estimativas (eixo y) com um erro menor que o eixo

x. Nesta curva, quanto menor o valor do eixo x e maior o valor do eixo y, melhor, pois isso significa que a maioria das estimativas têm erros menores. Como podemos ver, por exemplo, ao usar um limite $\alpha = 25\%$, apenas 68% das estimativas resultaram em um erro menor ou igual a 2 m. Por outro lado, para um limite $\alpha = 75\%$, quase 85% das estimativas resultaram em erros menores ou iguais a 2 m.

Assim, como podemos ver, a escolha de qual limite usar realmente depende do cenário e do tipo de aplicação. O principal desafio é decidir qual é o melhor caso para nosso problema atual: um número maior de estimativas de posição com uma precisão menor ou menos estimativas com uma precisão maior. No entanto, podemos tornar esta decisão mais fácil se pudermos melhorar esta solução usando melhores mecanismos de avaliação de confiança. Assim, em nossa solução LocFiND, usamos a Detecção de Novidades para melhorar essa avaliação de confiança.

5.3 Avaliação de Confiança Usando Detecção de Novidade

Conforme mencionado na Seção 4.2, para melhorar ainda mais a precisão do sistema de localização, introduzimos o teste *detecção de novidade*. Este teste avalia se uma amostra pertence a uma determinada classe (classe/posição predita pelo kNN). Esta abordagem ainda usa o limite α , mas usa a detecção de novidades para melhorar ainda mais a avaliação de confiança.

Semelhante à seção anterior, as Figuras 5.8, 5.9 e 5.10 mostram a relação entre o valor do limite α (eixo x), a porcentagem de estimativas aceitas (em azul) e a precisão final do classificações aceitas (curva vermelha em 5.8), o erro médio de localização (curva vermelha em 5.9) e a precisão da localização de sala (curva vermelha em 5.10).

Como podemos ver, a detecção de novidade foi de fato capaz de melhorar tanto a precisão da localização quanto o erro médio. No entanto, foi à custa de um número menor de previsões aceitas. Por exemplo, em comparação com a probabilidade de previsão em 5.2, como podemos ver na Figura 5.8, para um limite de $\alpha = 40\%$, a precisão da classificação está perto de 45%, contra 40% na abordagem anterior. Quanto ao erro médio de localização, mostrado na Figura 5.9, ele diminuiu de 1,8 m para 1,7 m. No entanto, em ambos os casos, as previsões aceitas foram de 85% para 68%. No caso da precisão da sala, aumentou de 96,4% para 97,7%, enquanto as previsões aceitas diminuíram de 99,6% para 87%.

Por outro lado, para um limite $\alpha = 80\%$ os resultados não mudam significativamente, pois, na maioria dos casos, as estimativas com altas probabilidades de predição

também passaram no teste de detecção de novidades.

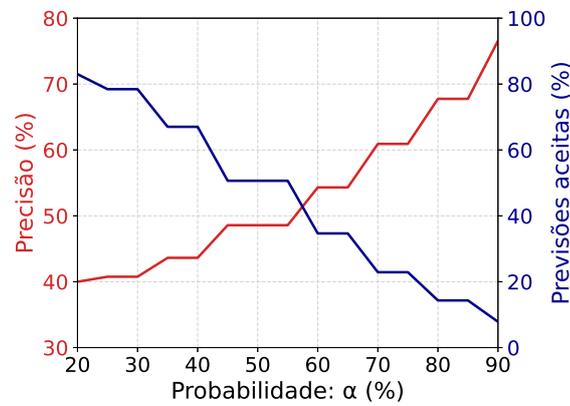


Figura 5.8: Precisão da localização de ponto usando *Deteção de Novidade*.

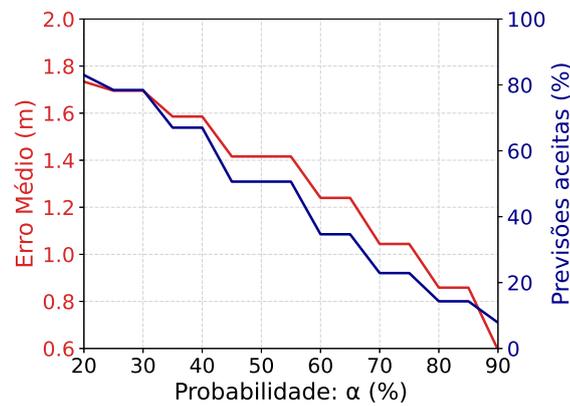


Figura 5.9: Erro médio da localização usando *Deteção de Novidade*.

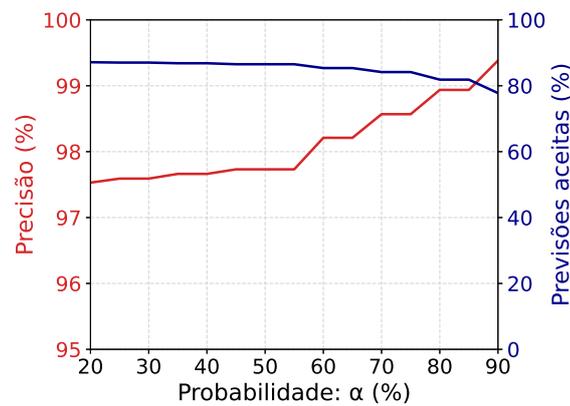


Figura 5.10: Precisão da localização de sala usando *Deteção de Novidade*.

Finalmente, a Figura 5.11 mostra a função de distribuição cumulativa para os limites $\alpha = 25\%$, 45% e 75% . Quando comparados aos resultados anteriores, podemos

notar algumas melhorias nos erros de localização. Por exemplo, ao usar um limite $\alpha = 25\%$ e considerando as estimativas com erros menores ou iguais a 2 m, o número de estimativas aumentou de 68% para 72%.

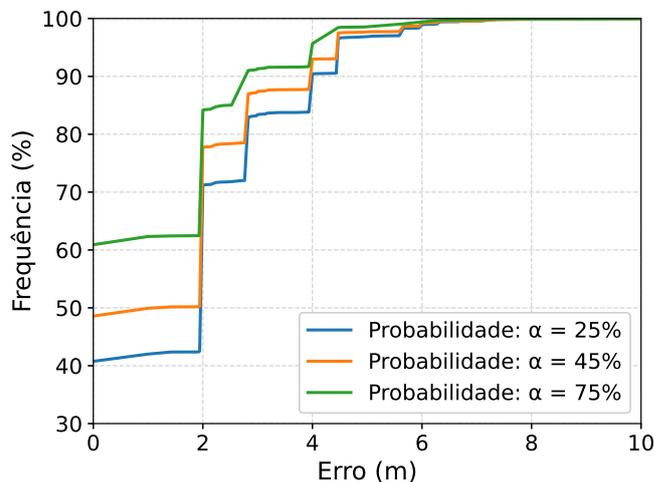


Figura 5.11: Função de Distribuição Acumulada do erro de localização usando *Detecção de Novidade* com limites $\alpha = 25\%$, 45% e 75%: uma porcentagem maior das estimativas de posição resultou em erros menores.

Como podemos ver, embora a detecção de novidades seja capaz de melhorar a precisão, foi à custa de previsões aceitas. Na próxima seção, mostraremos os resultados de nossa abordagem híbrida para avaliação de confiança, que tenta minimizar este problema.

5.4 Avaliação da Confiança Usando a Abordagem Híbrida

Nesta seção, mostramos e discutimos os resultados de uma terceira forma de avaliar a confiança, a *abordagem híbrida* para avaliar a confiança das posições estimadas. Conforme mencionado na Seção 4.3, ele combina as duas abordagens anteriores para melhorar a precisão enquanto reduz o número de estimativas descartadas.

Observou-se nas seções anteriores que quanto menor o limite de probabilidade, maior o erro de localização. Assim, nesta solução híbrida, usamos a probabilidade de predição para valores mais altos de α como um primeiro filtro inicial. Em seguida, usamos um limite de probabilidade mais baixo β para obter estimativas possivelmente erradas e aplicamos o teste de detecção de novidades mais rigoroso para confirmar suas previsões. Assim, espera-se mitigar a perda de previsões nas maiores probabilidades

e, dessa forma, ter algumas estimativas perdidas apenas nas predições com menores probabilidades.

Como nas seções anteriores, as Figuras 5.12, 5.13 e 5.14 mostram nossos resultados ao usar a abordagem híbrida para a avaliação de confiança. No caso da Figura 5.12, podemos notar que a precisão da localização é muito próxima à obtida pela abordagem de detecção de novidades (como mostrado na Seção 5.3). Por outro lado, o número de predições aceitas é muito semelhante àquelas obtidas pela abordagem de predição de probabilidade (como mostrado na Seção 5.2). Por exemplo, para o limite $\alpha = 40$, a precisão é 42%, perto da segunda abordagem e, em termos de previsões aceitas, aumentou de 64% para 70%, semelhante ao Primeira abordagem. Esses resultados mostram que essa abordagem híbrida pode chegar a um meio-termo, melhorando a localização e diminuindo apenas ligeiramente o número de estimativas descartadas.

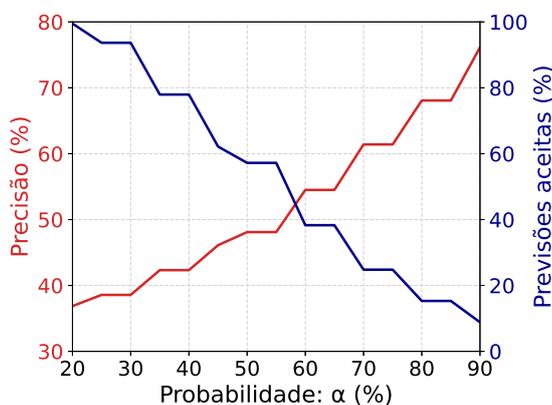


Figura 5.12: Precisão da localização de ponto usando *Abordagem Híbrida*.

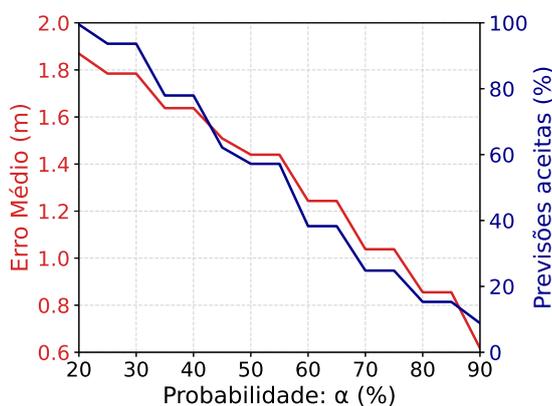


Figura 5.13: Erro médio da localização usando *Abordagem Híbrida*.

O mesmo pode ser observado para o erro médio de localização, mostrado na Figura 5.13. Finalmente, no caso da precisão da sala, mostrada na Figura 5.14, a

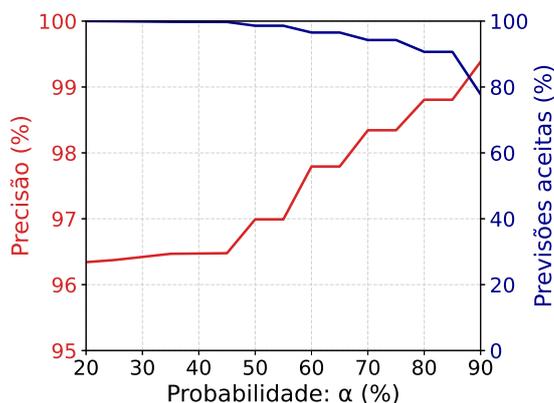


Figura 5.14: Precisão da localização de sala usando *Abordagem Híbrida*.

abordagem híbrida também foi capaz de melhorar essa precisão ao ter quase o mesmo número de amostras aceitas quando comparada à probabilidade de predição.

As Figuras 5.15, 5.16 e 5.17 mostram a função de distribuição cumulativa do erro de localização usando nossas três abordagens para a avaliação da confiança em três configurações de limite diferentes. Como podemos ver, a abordagem híbrida possui uma curva de erro quase igual ou melhor que a técnica mais restritiva, mas, como visto nos resultados anteriores, possui um número maior de estimativas aceitas.

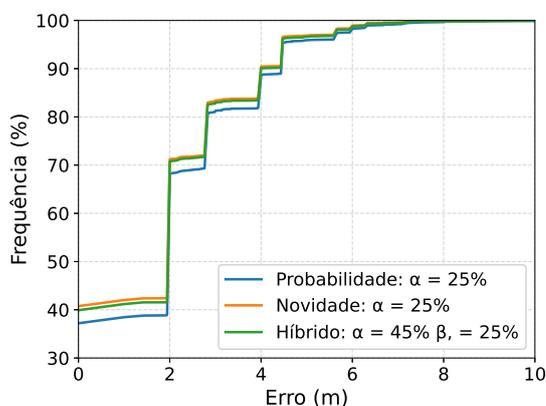


Figura 5.15: Função de Distribuição Acumulada do erro de localização usando nossas três abordagens usando como menor probabilidade e corte 25%.

5.5 Aplicabilidade, Pontos Fortes e Fracos do Sistema Proposto

Como mencionado antes, nossa principal motivação para implementar um sistema de localização para ambientes fechados é localizar os alunos dentro de uma escola. No

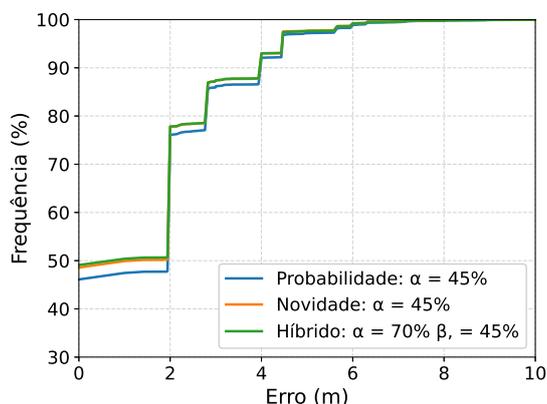


Figura 5.16: Função de Distribuição Acumulada do erro de localização usando nossas três abordagens usando como menor probabilidade e corte 45%.

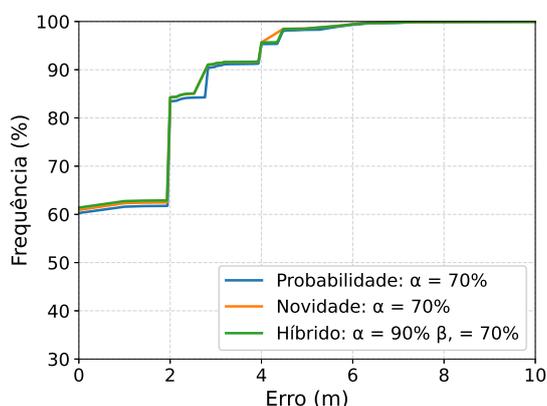


Figura 5.17: Função de Distribuição Acumulada do erro de localização usando nossas três abordagens usando como menor probabilidade e corte 70%.

entanto, na primeira versão do nosso sistema, as estimativas de posição instável resultaram em alunos pulando de sala em sala, embora estivessem na verdade estáticos. Para reduzir esses saltos e melhorar a precisão e estabilidade do nosso sistema, começamos a procurar soluções para avaliar e melhorar a confiança das previsões.

Assim, propusemos e aplicamos a solução LocFiND apresentada ao nosso sistema de localização e foi capaz de melhorar significativamente os resultados e estabilidade, mantendo a sensação de um sistema em tempo real. Nossa solução tem a vantagem de escolher os resultados mais confiáveis, com base na avaliação da probabilidade e da novidade. Isso torna o sistema mais resiliente, pois mesmo que tenhamos algum problema de comunicação, perda de pacotes ou interferência até um limite, o sistema pode continuar funcionando apenas descartando as informações problemáticas. Para outros cenários, dependendo dos requisitos da aplicação, nossa solução proposta pode ser facilmente ajustada para ter uma porcentagem aceitável de estimativas descartadas

enquanto melhora consideravelmente a precisão e estabilidade do sistema.

No entanto, o descarte de dados causa alguns problemas para avaliar os resultados, uma vez que a quantidade final de resultados é menor do que se usássemos todos os dados. Neste trabalho, a avaliação foi baseada na precisão e erro médio. A precisão pode ser definida como $a = h/n$, e o erro médio como $ae = e/n$, onde h é o número de resultados de classificação corretos e e é a soma dos erros associados ao resultados considerados válidos pelos algoritmos e n é o número de saídas válidas do algoritmo. Porém, como nosso algoritmo descarta alguns dados, o valor de n para eles torna-se diferente, tornando a comparação entre eles não totalmente justa. Assim, na maioria dos resultados, incluímos a porcentagem de previsões aceitas para deixar claro que, embora o erro tenha diminuído, o número de previsões aceitas também diminuiu.

Finalmente, sabe-se que os valores RSSI informados pelos dispositivos são um tanto dependentes do *hardware* e podem variar de um fabricante para outro. Embora isso possa afetar o desempenho dos sistemas de localização, no caso da nossa aplicação, na qual implementamos a solução proposta na área escolar, isso não foi um problema, pois todo o hardware era semelhante e do mesmo fabricante.

Conclusões e Trabalhos Futuros

Neste trabalho, propomos e avaliamos o desempenho de uma nova abordagem para localização em ambientes fechados, o LocFiND (*Localization using Fingerprinting e Novelty Detection*). Em nossa solução, as informações recebidas na fase online só serão aceitas após passar por uma avaliação de confiança. Em seguida, propomos três abordagens diferentes para essa avaliação de confiança. Nossa abordagem híbrida usa a probabilidade de previsão e a detecção de novidades para melhorar a precisão da localização enquanto tenta descartar o mínimo de informações possível. Na seção 6.1, nos descrevemos as principais contribuições do nosso trabalho, Seção 6.2 indica nossos próximos passos na pesquisa.

6.1 Contribuições

A principal contribuição de nossa solução é a possibilidade de remover resultados não confiáveis (que podem ter sido afetados por informações ruidosas) na fase *online*, enquanto a maioria dos trabalhos na literatura avalia o impacto do ruído na fase de treinamento ou coleta de dados. A principal preocupação de nossa abordagem é como descartar o menor número de estimativas possíveis e, ao mesmo tempo, aumentar a precisão e a estabilidade. Portanto, nossa solução oferece uma importante troca entre a precisão e o número de estimativas que podem ser descartadas sem comprometer o funcionamento da aplicação.

Nossos resultados mostram que podemos ter configurações menos restritas, para um limite $\alpha = 40\%$, que melhora a precisão da posição de 36% para 44% e o erro médio de localização de 1,9 m para 1,6 m em comparação com um kNN tradicional, descartando apenas 30% das estimativas. Por outro lado, podemos ter configurações mais restritas, para um limite $\alpha = 80\%$, que pode melhorar a precisão para quase 70%

e o erro médio de localização para 0,9 m, mas aceitando apenas 18% das estimativas. Em termos de precisão da sala, que era o foco principal de nosso aplicativo, LocFiND foi capaz de melhorar a precisão de 96% para quase 99,5%, descartando apenas 20% das estimativas para um limite $\alpha = 90\%$.

6.1.1 Publicações

Este trabalho resultou na publicação em *journal*:

- *Indoor Localization System Using Fingerprinting and Novelty Detection for Evaluation of Confidence* em *Future Internet* 2022, 14, 51. <https://doi.org/10.3390/fi14020051>

6.2 Próximos passos

Em trabalhos futuros, pretendemos propor e analisar outros métodos de avaliação de confiança. Por exemplo, o valor residual, usado principalmente no método dos mínimos quadrados, pode ser usado para calcular uma "distância" entre a nova amostra e os vizinhos mais próximos no banco de dados de impressão digital. Outra técnica que pode ser avaliada é o uso de filtros usados para detecção de ruídos como por exemplo filtro de kalman e suas variações. Nosso objetivo é avaliar como esta e outras medidas de discrepâncias podem ser usadas para melhorar nossa solução.

Outro ponto que podemos avançar é o problema gerado pelo descarte de informações não confiáveis na fase de localização. Pretendemos definir, dado o tempo de resposta exigido pelo sistema, quantas informações podemos descartar sem prejudicar o seu funcionamento. Podemos definir quantidades mínimas de dados distribuídos ao longo do tempo como um requisito do aplicativo, que define o limite máximo de descarte de dados.

Referências Bibliográficas

- Abdat, M.; Wan, T.-C. & Supramaniam, S. (2010). Survey on indoor wireless positioning techniques: Towards adaptive systems. Em *Distributed Framework and Applications (DFmA), 2010 International Conference on*, pp. 1--5.
- Aha, D. W.; Kibler, D. & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6:37--66. 10.1023/A:1022689900470.
- Alkhawaja, F.; Jaradat, M. & Romdhane, L. (2019). Techniques of indoor positioning systems (ips): A survey. Em *2019 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1--8.
- Altini, M.; Brunelli, D.; Farella, E. & Benini, L. (2010). Bluetooth indoor localization with multiple neural networks. Em *IEEE 5th International Symposium on Wireless Pervasive Computing 2010*, pp. 295--300.
- Amarbayasgalan, T.; Jargalsaikhan, B. & Ryu, K. H. (2018). Unsupervised novelty detection using deep autoencoders with density based clustering. *Applied Sciences*, 8(9).
- Ansari, J.; Riihijarvi, J. & Mahonen, P. (2007). Combining particle filtering with cricket system for indoor localization and tracking services. Em *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1--5.
- Arbula, D. & Ljubic, S. (2020). Indoor localization based on infrared angle of arrival sensor network. *Sensors*, 20(21).
- Assad, M. A. (2007). A real-time laboratory testbed for evaluating localization performance of wifi rfid technologies. Master's thesis, Worcester Polytechnic Institute.
- Atia, M. M.; Noureldin, A. & Korenberg, M. J. (2013). Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *IEEE Transactions on Mobile Computing*, 12(9):1774--1787.

- Au, A. W. S.; Feng, C.; Valaee, S.; Reyes, S.; Sorour, S.; Markowitz, S. N.; Gold, D.; Gordon, K. & Eizenman, M. (2013). Indoor tracking and navigation using received signal strength and compressive sensing on a mobile device. *IEEE Transactions on Mobile Computing*, 12(10):2050–2062.
- Bahl, P. & Padmanabhan, V. (2000). Radar: an in-building rf-based user location and tracking system. Em *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pp. 775--784 vol.2.
- Bahl, P.; Padmanabhan, V. N. & Balachandran, A. (2000). Enhancements to the radar user location and tracking system. Technical report.
- Basiri, A.; Lohan, E. S.; Moore, T.; Winstanley, A.; Peltola, P.; Hill, C.; Amirian, P. & Figueiredo e Silva, P. (2017). Indoor location based services challenges, requirements and usability of current solutions. *Computer Science Review*, 24:1–12.
- Battiti, R.; Nhat, T. L. & Villani, A. (2002). Location-aware computing: A neural network model for determining location in wireless lans. Technical report.
- Bhatti, M. A.; Riaz, R.; Rizvi, S. S.; Shokat, S.; Riaz, F. & Kwon, S. J. (2020). Outlier detection in indoor localization and internet of things (iot) using machine learning. *Journal of Communications and Networks*, 22(3):236–243.
- Bin Aftab, M. U. (2017). *Building Bluetooth Low Energy Systems*. Packt Publishing Ltd.
- Bolliger, P. (2011). *Robust Indoor Positioning through Adaptive Collaborative Labeling of Location Fingerprints*. PhD thesis, ETH Zurich, Zurich, Switzerland.
- Bouet, M. & dos Santos, A. (2008). Rfid tags: Positioning principles and localization techniques. Em *Wireless Days, 2008. WD '08. 1st IFIP*, pp. 1--5.
- Bozkurt, S.; Elibol, G.; Gunal, S. & Yayan, U. (2015). A comparative study on machine learning algorithms for indoor positioning. Em *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–8.
- Bramer, M. (2020). *Principles of Data Mining*. Springer, London.
- Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; Layton, R.; VanderPlas, J.; Joly, A.; Holt, B. & Varoquaux, G. (2013). API design for machine learning software:

- experiences from the scikit-learn project. Em *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108--122.
- Campaña, F.; Pinargote, A.; Domínguez, F. & Peláez, E. (2017). Towards an indoor navigation system using bluetooth low energy beacons. Em *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, pp. 1--6.
- Casas, R.; Cuartielles, D.; Marco, A.; Gracia, H. & Falco, J. (2007). Hidden issues in deploying an indoor location system. *Pervasive Computing, IEEE*, 6(2):62--69.
- Castillo-Cara, M.; Lovón-Melgarejo, J.; Bravo-Rocca, G.; Orozco-Barbosa, L. & García-Varea, I. (2017). An empirical study of the transmission power setting for bluetooth-based indoor localization mechanisms. *Sensors*, 17(6).
- Castro, P.; Chiu, P.; Kremenek, T. & Muntz, R. R. (2001). A probabilistic room location service for wireless networked environments. Em *Proceedings of the 3rd international conference on Ubiquitous Computing, UbiComp '01*, pp. 18--34, London, UK, UK. Springer-Verlag.
- Chen, J.-C.; Wang, Y.-C.; Maa, C.-S. & Chen, J.-T. (2006). Network-side mobile position location using factor graphs. *IEEE Transactions on Wireless Communications*, 5(10):2696--2704.
- Chen, Y.-C. & Juang, J.-C. (2012). Outlier-detection-based indoor localization system for wireless sensor networks. *International Journal of Navigation and Observation*, 2012.
- Chen, Z. & Wang, J. (2018). Grof: Indoor localization using a multiple-bandwidth general regression neural network and outlier filter. *Sensors*, 18(11).
- Chen, Z.; Zou, H.; Jiang, H.; Zhu, Q.; Soh, Y. C. & Xie, L. (2015). Fusion of wifi, smartphone sensors and landmarks using the kalman filter for indoor localization. *Sensors*, 15(1):715--732.
- Cheng, S.; Wang, S.; Guan, W.; Xu, H. & Li, P. (2020). 3dlra: An rfid 3d indoor localization method based on deep learning. *Sensors*, 20(9).
- Chriki, A.; Touati, H. & Snoussi, H. (2017). Svm-based indoor localization in wireless sensor networks. Em *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 1144--1149.

- Conti, A.; Dardari, D. & Win, M. Z. (2008). Experimental results on cooperative uwb based positioning systems. Em *2008 IEEE International Conference on Ultra-Wideband*, volume 1, pp. 191–195.
- Cui, W.; Zhang, L.; Li, B.; Guo, J.; Meng, W.; Wang, H. & Xie, L. (2017). Received signal strength based indoor positioning using a random vector functional link network. *IEEE Transactions on Industrial Informatics*, PP:1–1.
- D’Aloia, M.; Longo, A.; Guadagno, G.; Pulpito, M.; Fornarelli, P.; Laera, P. N.; Manni, D. & Rizzi, M. (2020). Iot indoor localization with ai technique. Em *2020 IEEE International Workshop on Metrology for Industry 4.0 IoT*, pp. 654–658.
- Dang, X.; Si, X.; Hao, Z. & Huang, Y. (2019). A novel passive indoor localization method by fusion csi amplitude and phase information. *Sensors*, 19(4).
- Dardari, D.; Andrea, C.; Jaime, L. & Win, M. (2008). The effect of cooperation on uwb-based positioning systems using experimental data. *EURASIP Journal on Advances in Signal Processing*, 2008.
- Dashti, M.; Yiu, S.; Yousefi, S.; Perez-Cruz, F. & Claussen, H. (2015). Rssi localization with gaussian processes and tracking. Em *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6.
- de Oliveira, A. F. (2021). 3d indoor localization with bluetooth low energy networks using machine learning techniques. Master’s thesis, Unesp - Universidade Estadual Paulista, São João da Boa Vista.
- Ding, X.; Li, Y.; Belatreche, A. & Maguire, L. P. (2014). An experimental evaluation of novelty detection methods. *Neurocomputing*, 135:313–327.
- Drane, C.; Macnaughtan, M. & Scott, C. (1998). Positioning gsm telephones. *Communications Magazine, IEEE*, 36(4):46–54, 59.
- Faria, E. R.; Gonçalves, I. J. C. R.; de Carvalho, A. C. P. L. F. & Gama, J. (2016). Novelty detection in data streams. *Artificial Intelligence Review*, 45(2):235–269.
- Fink, A.; Beikirch, H.; Voss, M. & Schroder, C. (2010). Rssi-based indoor positioning using diversity and inertial navigation. Em *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pp. 1–7.
- Gao, L. (2015). Channel state information fingerprinting based indoor localization: a deep learning approach. Master’s thesis, Auburn University, Department Electrical Engineering, Auburn, Alabama.

- Gkotsiopoulos, P.; Zorbas, D. & Douligeris, C. (2021). Performance determinants in lora networks: A literature review. *IEEE Communications Surveys & Tutorials*, 23(3):1721–1758.
- Gu, F.; Hu, X.; Ramezani, M.; Acharya, D.; Khoshelham, K.; Valaee, S. & Shang, J. (2019). Indoor localization improved by spatial context—a survey. *ACM Comput. Surv.*, 52(3).
- Gu, Y.; Lo, A. & Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys Tutorials*, 11(1):13–32.
- Hahanov, V.; Filippenko, I. & Lavrova, L. (2007). Contemporary rfid systems and identification problems. Em *CAD Systems in Microelectronics, 2007. CADSM '07. 9th International Conference - The Experience of Designing and Applications of*, pp. 343–344.
- Halperin, D.; Hu, W.; Sheth, A. & Wetherall, D. (2010). Predictable 802.11 packet delivery from wireless channel measurements. *SIGCOMM Comput. Commun. Rev.*, 40(4):159–170.
- Hamza, L. N. C. (2010). A dynamic system approach for radio location fingerprinting in wireless local area networks. *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, pp. 253–258.
- Hatem, E.; Abou-Chakra, S.; Colin, E.; Laheurte, J.-M. & El-Hassan, B. (2020). Performance, accuracy and generalization capability of rfid tags' constellation for indoor localization. *Sensors*, 20(15).
- Hautamaki, V.; Karkkainen, I. & Franti, P. (2004). Outlier detection using k-nearest neighbour graph. Em *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, pp. 430–433 Vol.3.
- He, S. & Chan, S.-H. G. (2016). Tilejunction: Mitigating signal noise for fingerprint-based indoor localization. *IEEE Transactions on Mobile Computing*, 15(6):1554–1568.
- Hou, Y.; Yang, X. & Abbasi, Q. H. (2018). Efficient aoa-based wireless indoor localization for hospital outpatients using mobile devices. *Sensors*, 18(11).
- IEEE (2005). Draft amendment to ieee standard for information technology-telecommunications and information exchange between systems-part 15.4: Wireless

- medium access control (mac) and physical layer (phy) specifications for low-rate wireless personal area networks (lr-wpans): Amendment to add alternate phy (amendment of ieee std 802.15.4). *IEEE Std P802.15.4a/D5*, pp. --.
- Islam, B.; Islam, M. T.; Kaur, J. & Nirjon, S. (2019). Lorain: Making a case for lora in indoor localization. Em *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 423–426.
- Jang, B. & Kim, H. (2019). Indoor positioning technologies without offline fingerprinting map: A survey. *IEEE Communications Surveys & Tutorials*, 21(1):508–525.
- Jiang, Q.; Ma, Y.; Liu, K. & Dou, Z. (2016). A probabilistic radio map construction scheme for crowdsourcing-based fingerprinting localization. *IEEE Sensors Journal*, 16(10):3764–3774.
- Kaemarungsi, K. & Krishnamurthy, P. (2012). Analysis of wlan’s received signal strength indication for indoor location fingerprinting. *Pervasive and Mobile Computing*, 8(2):292–316. Special Issue: Wide-Scale Vehicular Sensor Networks and Mobile Sensing.
- Kaspar, D. (2005). Application of directional antennas in rf-based indoor localization systems. Master’s thesis, Swiss Federal Institute of Technology Zurich.
- Kim, K.; Li, S.; Heydariaan, M.; Smaoui, N.; Gnawali, O.; Suh, W.; Suh, M. J. & Kim, J. I. (2021). Feasibility of lora for smart home indoor localization. *Applied Sciences*, 11(1).
- Kimoto, R.; Ishida, S.; Yamamoto, T.; Tagashira, S. & Fukuda, A. (2019). Muchloc: Indoor zigbee localization system utilizing inter-channel characteristics. *Sensors*, 19(7).
- Kumar, P.; Reddy, L. & Varma, S. (2009). Distance measurement and error estimation scheme for rssi based localization in wireless sensor networks. Em *2009 Fifth International Conference on Wireless Communication and Sensor Networks (WCSN)*, pp. 1–4.
- Lam, K.-H.; Cheung, C.-C. & Lee, W.-C. (2019). Rssi-based lora localization systems for large-scale indoor and outdoor environments. *IEEE Transactions on Vehicular Technology*, 68(12):11778–11791.

- Laoudias, C.; Piché, R. & Panayiotou, C. (2013). Device self-calibration in location systems using signal strength histograms. *Journal of Location Based Services*, 7:165–181.
- Larranaga, J.; Muguira, L.; Lopez-Garde, J.-M. & Vazquez, J.-I. (2010). An environment adaptive zigbee-based indoor positioning algorithm. Em *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pp. 1--8.
- Liu, H.; Darabi, H.; Banerjee, P. & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1067–1080.
- Loganathan, A.; Ahmad, N. S. & Goh, P. (2019). Self-adaptive filtering approach for improved indoor localization of a mobile node with zigbee-based rssi and odometry. *Sensors*, 19(21).
- Lymberopoulos, D.; Liu, J.; Yang, X.; Choudhury, R. R.; Handziski, V. & Sen, S. (2015). A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned. Em *Proceedings of the 14th International Conference on Information Processing in Sensor Networks, IPSN '15*, p. 178–189, New York, NY, USA. Association for Computing Machinery.
- Mainetti, L.; Patrono, L. & Sergi, I. (2014). A survey on indoor positioning systems. Em *2014 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 111–120.
- Malekitabar, A.; Aghababa, H.; Radfar, M. & Khalaj, B. (2005). Minimizing the error of time difference of arrival method in mobile networks. Em *Wireless and Optical Communications Networks, 2005. WOCN 2005. Second IFIP International Conference on*, pp. 328--332.
- Markou, M. & Singh, S. (2003). Novelty detection: a review—part 1: statistical approaches. *Signal Processing*, 83(12):2481 – 2497.
- Martin, P.; Ho, B.-J.; Grupen, N.; Muñoz, S. & Srivastava, M. (2014). An ibeacon primer for indoor localization: Demo abstract. Em *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, BuildSys '14*, p. 190–191, New York, NY, USA. Association for Computing Machinery.
- Micheletti, J. A. & Godoy, E. P. (2022). Improved indoor 3d localization using lora wireless communication. *IEEE Latin America Transactions*, 20(3):481–487.

- Miljković, D. (2010). Review of novelty detection methods. Em *The 33rd International Convention MIPRO*, pp. 593–598.
- Minaev, G.; Visa, A. & Piché, R. (2017). Comprehensive survey of similarity measures for ranked based location fingerprinting algorithm. Em *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–4.
- Motohashi, Y.; Sato-Shimokawara, E.; Chan, R. Y.-Y.; Zhou, N. & Yamaguchi, T. (2017). Using ibeacon to detect user behavior from indoor physical movement. Em *2017 IEEE/SICE International Symposium on System Integration (SII)*, pp. 799–804.
- Mukhopadhyay, B.; Sarangi, S.; Srirangarajan, S. & Kar, S. (2018). Indoor localization using analog output of pyroelectric infrared sensors. Em *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6.
- Ngamakeur, K.; Yongchareon, S.; Yu, J. & Rehman, S. U. (2020). A survey on device-free indoor localization and tracking in the multi-resident environment. *ACM Comput. Surv.*, 53(4).
- Nguyen, D. D. & Thuy Le, M. (2021). Enhanced indoor localization based ble using gaussian process regression and improved weighted knn. *IEEE Access*, 9:143795–143806.
- Norris, J. (1998). *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Osório, R. M. M. M. (2011). Sistema de identificação e localização baseado em dispositivos de comunicações sem fios. Master's thesis, Faculdade de Engenharia da Universidade do Porto.
- Pimentel, M. A.; Clifton, D. A.; Clifton, L. & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99:215 – 249.
- Pinto, B. H. O. U. V.; de Oliveira, H. A. B. F. & Souto, E. J. P. (2021). Factor optimization for the design of indoor positioning systems using a probability-based algorithm. *Journal of Sensor and Actuator Networks*, 10(1).
- Qureshi, U. M.; Umair, Z. & Hancke, G. P. (2019). Evaluating the implications of varying bluetooth low energy (ble) transmission power levels on wireless indoor localization accuracy and precision. *Sensors*, 19(15).

- Ramya, C. M.; Shanmugaraj, M. & Prabakaran, R. (2011). Study on zigbee technology. Em *2011 3rd International Conference on Electronics Computer Technology*, volume 6, pp. 297–301.
- Rodrigues, M. L. (2011). Localização em ambientes internos utilizando múltiplas tecnologias sem fio. Master's thesis, Universidade Federal de Minas Gerais. Departamento de Ciência da Computação.
- Röhrig, C. & Müller, M. (2009). Indoor location tracking in non-line-of-sight environments using a ieee 802.15.4a wireless network. Em *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems, IROS'09*, pp. 552--557, Piscataway, NJ, USA. IEEE Press.
- Roos, T.; Myllymäki, P.; Tirri, H.; Misikangas, P. & Sievänen, J. (2002). A probabilistic approach to wlan user location estimation. *International Journal of Wireless Information Networks*, 9(3):155--164.
- Roy, P. & Chowdhury, C. (2021). A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent & Robotic Systems*, 101(3):63.
- Sadowski, S. & Spachos, P. (2018). Rssi-based indoor localization with the internet of things. *IEEE Access*, 6:30149–30161.
- Saha, S.; Chaudhuri, K.; Sanghi, D. & Bhagwat, P. (2003). Location determination of a mobile device using ieee 802.11b access point signals. Em *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, volume 3, pp. 1987--1992 vol.3.
- Salamah, A. H.; Tamazin, M.; Sharkas, M. A. & Khedr, M. (2016). An enhanced wifi indoor localization system based on machine learning. Em *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8.
- Sayyed, S. & Deolekar, R. (2017). A survey on novelty detection using level set methods. Em *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 415–418.
- Sharma, N. K. (2006). A weighted center of mass based trilateration approach for locating wireless devices in indoor environment. Em *Proceedings of the 4th ACM International Workshop on Mobility Management and Wireless Access, MobiWac '06*, p. 112–115, New York, NY, USA. Association for Computing Machinery.

- Sie, M.-K. & Kuo, C.-H. (2017). Indoor location estimation using ble beacon with multiple transmission power levels. Em *2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, pp. 323–324.
- Singh, N.; Choe, S. & Punmiya, R. (2021). Machine learning based indoor localization using wi-fi rssi fingerprints: An overview. *IEEE Access*, 9:127150–127174.
- Smailagic, A. & Kogan, D. (2002). Location sensing and privacy in a context-aware computing environment. *Wireless Communications, IEEE*, 9(5):10–17.
- Sun, D.; Wei, E.; Ma, Z.; Wu, C. & Xu, S. (2021). Optimized cnns to indoor localization through ble sensors using improved pso. *Sensors*, 21(6).
- Swangmuang, N. & Krishnamurthy, P. (2008). Location fingerprint analyses toward efficient indoor positioning. Em *2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 100–109.
- Tanenbaum, A. (2002). *Computer Networks*. Prentice Hall Professional Technical Reference, 4th edição.
- Torrieri, D. (1984). Statistical theory of passive location systems. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-20(2):183–198.
- Turgut, Z.; Aydin, G. Z. G. & Sertbas, A. (2016). Indoor localization techniques for smart building environment. *Procedia Computer Science*, 83:1176–1181. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.
- Viegas, J. L.; Esteves, P. R. & Vieira, S. M. (2018). Clustering-based novelty detection for identification of non-technical losses. *International Journal of Electrical Power & Energy Systems*, 101:301–310.
- Wang, B.; Chen, Q.; Yang, L. T. & Chao, H. (2016). Indoor smartphone localization via fingerprint crowdsourcing: challenges and approaches. *IEEE Wireless Communications*, 23(3):82–89.
- Want, R. (2006). An introduction to rfid technology. *Pervasive Computing, IEEE*, 5(1):25–33.
- Wen, F. & Liang, C. (2015). Fine-grained indoor localization using single access point with multiple antennas. *IEEE Sensors Journal*, 15(3):1538–1544.

- Wu, C.; Yang, Z. & Liu, Y. (2018). *Wireless Indoor Localization: A Crowdsourcing Approach*.
- Wu, K.; Xiao, J.; Yi, Y.; Chen, D.; Luo, X. & Ni, L. M. (2013). Csi-based indoor localization. *IEEE Transactions on Parallel and Distributed Systems*, 24(7):1300–1309.
- Xiao, J.; Zhou, Z.; Yi, Y. & Ni, L. M. (2016). A survey on wireless indoor localization from the device perspective. *ACM Comput. Surv.*, 49(2).
- Xu, Y. & GmbH, K. U. P. (2016). *Autonomous Indoor Localization Using Unsupervised Wi-Fi Fingerprinting*. Kassel University Press.
- Yan, M.; Xu, F.; Bai, S. & Wan, Q. (2018). A noise reduction fingerprint feature for indoor localization. Em *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6.
- Yang, B.; Guo, L.; Guo, R.; Zhao, M. & Zhao, T. (2020). A novel trilateration algorithm for rssi-based indoor localization. *IEEE Sensors Journal*, 20(14):8164–8172.
- Yang, H. & Yang, S.-H. (2009). Connectionless indoor inventory tracking in zigbee rfid sensor network. Em *Industrial Electronics, 2009. IECON '09. 35th Annual Conference of IEEE*, pp. 2618--2623.
- Yang, L.; Chen, H.; Cui, Q.; Fu, X. & Zhang, Y. (2015). Probabilistic-knn: A novel algorithm for passive indoor-localization scenario. Em *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5.
- Yang, Z.; Zhou, Z. & Liu, Y. (2013a). From rssi to csi: Indoor localization via channel response. *ACM Comput. Surv.*, 46(2).
- Yang, Z.; Zhou, Z. & Liu, Y. (2013b). From rssi to csi: Indoor localization via channel response. *ACM Comput. Surv.*, 46(2).
- Yassin, A.; Nasser, Y.; Awad, M.; Al-Dubai, A.; Liu, R.; Yuen, C.; Raulefs, R. & Abou-tanios, E. (2017). Recent advances in indoor localization: A survey on theoretical approaches and applications. *IEEE Communications Surveys Tutorials*, 19(2):1327–1346.
- Yiu, S.; Dashti, M.; Claussen, H. & Perez-Cruz, F. (2017). Wireless rssi fingerprinting localization. *Signal Processing*, 131:235–244.

- Youssef, M. & Agrawala, A. (2004). Handling samples correlation in the horus system. Em *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pp. 1023--1031 vol.2.
- Youssef, M. A.; Agrawala, A. & Shankar, A. U. (2003). Wlan location determination via clustering and probability distributions. Em *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, PERCOM '03*, pp. 143--, Washington, DC, USA. IEEE Computer Society.
- Yuqiang Zhang & Junhui Zhao (2005). Indoor localization using time difference of arrival and time-hopping impulse radio. Em *IEEE International Symposium on Communications and Information Technology, 2005. ISCIT 2005.*, volume 2, pp. 964–967.
- Zafari, F.; Gkelias, A. & Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys Tutorials*, 21(3):2568–2599.
- Zekavat, R. & Buehrer, R. M. (2012). *RF Fingerprinting Location Techniques*, pp. 487–520.
- Zhang, C.; Kuhn, M.; Merkl, B.; Fathy, A. E. & Mahfouz, M. (2006). Accurate uwb indoor localization system utilizing time difference of arrival approach. Em *2006 IEEE Radio and Wireless Symposium*, pp. 515–518.
- Zhu, X.; Qu, W.; Qiu, T.; Zhao, L.; Atiquzzaman, M. & Wu, D. O. (2020). Indoor intelligent fingerprint-based localization: Principles, approaches and challenges. *IEEE Communications Surveys Tutorials*, 22(4):2634–2657.
- Zimek, A.; Schubert, E. & Kriegel, H.-P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(5):363–387.