



Universidade Federal de Pernambuco
Centro de Informática

Graduação em Engenharia da Computação

**Aplicação de Algoritmo de Localização
Indoor Baseado em Sinais de
Radiofrequência**

José Renê Santos Leite

Trabalho de Graduação

Recife
5 de julho de 2018

Universidade Federal de Pernambuco
Centro de Informática

José Renê Santos Leite

Aplicação de Algoritmo de Localização Indoor Baseado em Sinais de Radiofrequência

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: *Prof. Dr. Nivan Roberto Ferreira Júnior*

Recife
5 de julho de 2018

Dedico esse trabalho à minha mãe, Elisabeth. Que sua memória viva, através de ações, em minha história.

*Quero assistir ao sol nascer, Ver as águas dos rios correr, Ouvir os
pássaros cantar. Eu quero nascer, Quero viver.*
—CARTOLA (Preciso me encontrar, 1976)

Resumo

O desenvolvimento de sistemas de localização em tempo real se tornou um grande auxílio para os sistemas baseados em localização. Sistemas de recomendação, jogos e busca de rotas são exemplos de sistemas que podem ser baseados na localização. O problema da localização em tempo real em ambientes *outdoor* já é bem resolvido pelo sistema GPS, porém o GPS não é capaz de realizar predições de qualidade em ambientes *indoor*. A capacidade de rastrear pessoas e objetos em ambientes *indoor* tem aplicações médicas, militares, logísticas e de entretenimento. Um sistema de localização *indoor* é um sistema para localizar pessoas ou objetos no interior de um dado estabelecimento utilizando ondas de rádio, campos magnéticos, sinais acústicos ou qualquer outra informação sensorial que possa ser coletada por dispositivos móveis. O trabalho será focado no uso de ondas de radiofrequência para localização *indoor*, através de mapeamento de ambiente prévio seguido da localização através de inteligência artificial com o método knn, tendo como entrada os sinais de radiofrequência medidos pelo dispositivo a ser localizado e como saída as coordenadas da localização do dispositivo.

Palavras-chave: localização *indoor*, localização por radiofrequência

Abstract

Development of indoor localization systems provides aid to location-based systems. Recommendation systems and route searching are examples of location-based systems. The outdoor localization problem is already solved at a high level of accuracy by the GPS system although the GPS suffers from poor quality when used inside buildings. The ability to track a person or an object located in an indoor environment has health, military and entertaining applications. An indoor localization system is a system able to find a person or an object located on the inside of a building using radio frequency waves, magnetic fields, acoustic signals or another kind of sensor information available in mobile devices. The presented work restricts itself to radio-frequency analysis to locate at indoor environments separated in two phases. The first phase is the environment mapping which is an off-line phase to collect information about the place of choice. The second phase is the online phase where the approximate position of the device is estimated through the artificial intelligence method of knn.

Keywords: indoor localization, radio-frequency localization

Sumário

1	Introdução	1
1.1	Contexto	1
1.2	Objetivos	1
1.3	Estrutura do Trabalho	2
2	Conceitos Básicos	3
2.1	Contexto do Usuário	3
2.2	Localização Indoor e Localização Outdoor	3
2.3	Localização Outdoor	4
2.4	Sistema de Localização Indoor	4
2.5	Técnicas Baseadas em Radiofrequência	4
2.6	Fingerprinting	5
2.7	Métricas de Comparação de Fingerprints	5
2.7.1	Distância Euclideana	6
2.7.2	Distância Manhattan	6
2.7.3	Distância Chebyshev	6
2.8	Inteligência Artificial	7
2.8.1	Método k-NN	7
2.8.2	Estratégia de União de Vizinhos com Peso Constante	7
2.8.3	Estratégia de União de Vizinhos com Peso Inverso da Distância	7
2.8.4	Estratégia de União de Vizinhos com Peso Exponencial na Distância	7
3	Trabalhos Relacionados	9
3.1	RADAR	9
3.2	Horus	11
3.3	Variações do RADAR	13
4	Base de Treinamento e Base de Teste	15
4.1	Falta de Bases Comuns Para Localização Indoor	15
4.2	UJIIndoorLoc	15
5	Experimentos e Análise	17
5.1	Proposta	17
5.2	Experimentos com a métrica Manhattan	18
5.2.1	Utilizando a União de Vizinhos com Peso Constante	18
5.2.2	Utilizando a União de Vizinhos com Peso Inverso da Distância	20

5.2.3	Utilizando a União de Vizinhos com Peso Exponencial na Distância	22
5.3	Experimentos com a métrica euclideana	24
5.3.1	Utilizando a União de Vizinhos com Peso Constante	24
5.3.2	Utilizando a União de Vizinhos com Peso Inverso da Distância	26
5.3.3	Utilizando a União de Vizinhos com Peso Exponencial na Distância	28
5.4	Experimentos com a métrica Chebyshev	30
5.4.1	Utilizando a União de Vizinhos com Peso Constante	30
5.4.2	Utilizando a União de Vizinhos com Peso Inverso da Distância	32
5.4.3	Utilizando a União de Vizinhos com Peso Exponencial na Distância	34
5.5	Análise dos Resultados	36
6	Conclusão e Trabalhos Futuros	39
6.1	Conclusão	39
6.2	Trabalhos Futuros	40

Lista de Figuras

3.1	Exemplo de localizações das estações base para o método RADAR	9
3.2	Posição problema para algoritmo do RADAR	10
5.1	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso constante.	18
5.2	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso inverso da distância.	20
5.3	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso exponencial na distância.	22
5.4	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso constante.	24
5.5	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica euclideana usando a união de vizinhos com peso inverso da distância.	26
5.6	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica euclideana usando a união de vizinhos com peso inverso da distância.	28
5.7	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso constante.	30
5.8	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Chebyshev usando a união de vizinhos com peso inverso da distância.	32
5.9	Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Chebyshev usando a união de vizinhos com peso inverso da distância.	34

Lista de Tabelas

5.1	Erro em metros para Manhattan e estratégia de união de k vizinhos com peso constante, k variando de 1 a 35.	19
5.2	Erro em metros para métrica Manhattan e estratégia de união de k vizinhos com peso inverso da distância, k variando de 1 a 35.	21
5.3	Erro em metros para métrica Manhattan e estratégia de união de k vizinhos com peso exponencial na distância, k variando de 1 a 35.	23
5.4	Erro em metros para métrica euclideana e estratégia de união de k vizinhos com peso constante, k variando de 1 a 35.	25
5.5	Erro em metros para métrica euclideana e estratégia de união de k vizinhos com peso inverso da distância, k variando de 1 a 35.	27
5.6	Erro em metros para métrica euclideana e estratégia de união de k vizinhos com peso exponencial na distância, k variando de 1 a 35.	29
5.7	Erro em metros para métrica Chebyshev e estratégia de união de k vizinhos com peso constante, k variando de 1 a 35.	31
5.8	Erro em metros para métrica Chebyshev e estratégia de união de k vizinhos com peso inverso da distância, k variando de 1 a 35.	33
5.9	Erro em metros para métrica Chebyshev e estratégia de união de k vizinhos com peso exponencial na distância, k variando de 1 a 35.	35
5.10	Comparação dos melhores resultados no percentil 50 das várias métricas e estratégias, em metros.	36

CAPÍTULO 1

Introdução

1.1 Contexto

A presença de dispositivos móveis é crescente em nossa sociedade. Em 2016 as compras de smartphones superaram 1,4 bilhão de unidades [Gartner, 2017]. Além disso é comum para os usuários passar mais de duas horas diárias interagindo com o dispositivo [Hackernoon, 2017], que virou uma presença constante na rotina. Essa presença despertou o interesse para o desenvolvimento de aplicações sensíveis ao contexto. Alguns exemplos são a mídia entregue de acordo com o contexto do usuário [InLoco, 2018] e o rastreamento de objetos em armazéns [LIPS, 2018].

Como aparelhos móveis possuem diversos sensores e comunicação com redes sem fio, é possível monitorar e até reagir ao contexto. Esses serviços podem estar focados na reação à localização do usuário. Para isso um sistema de localização *indoor* passa a ser necessário.

1.2 Objetivos

O objetivo deste trabalho é o de realizar um estudo sobre contexto de usuário, avançando de forma cada vez mais específica para localização como um contexto de interesse e finalmente chegando no problema da localização *indoor*.

Dentro de localização *indoor* explicar sobre os métodos existentes de forma geral com as várias maneiras de realizar predições usando variadas fontes, como câmeras, sensores de infravermelho, sensores acústicos, sinais enviados por redes móveis de telefonia, sensores de Wi-Fi e especificando para o uso de métodos que se limitam à análise de radiofrequência, especificamente a frequência de 2.4GHz utilizada por dispositivos Wi-Fi.

Dos métodos que utilizam radiofrequência, mostrar o porquê desse método funcionar, mostrar quais são as informações necessárias para realizar uma predição e começar a apresentar a potência do sinal das redes visíveis como a chave para encontrar uma relação com posição.

O objetivo principal será apresentar o conceito de *fingerprinting* e chegar à conclusão de que o problema é uma regressão em que temos como entrada uma leitura de sinais Wi-Fi com suas potências respectivas e como saída as coordenadas do usuário que enviou as leituras. Isso será feito com a aplicação de um método canônico na área de localização *indoor* em uma base de dados que foi criada como uma tentativa de padronizar o uso de bases para estudos dessa área de pesquisa.

Da inteligência artificial, o método k-NN será explicado e utilizado como uma forma de melhorar o resultado das predições realizadas.

1.3 Estrutura do Trabalho

O trabalho é composto por 6 capítulos, incluindo este capítulo introdutório. O Capítulo 2 discorrerá sobre conceitos básicos de contexto de usuário e suas utilidades desde criação de aplicações de simples monitoramento até a criação de sistemas sensíveis ao contexto, aprofundando o conceito de localização *indoor* explicando a diferença entre outdoor e *indoor*. Também explicando os métodos outdoor conhecidos e apresentando a existência de técnicas baseadas em radiofrequência e explicando as métricas de comparação entre medidas de radiofrequência.

O Capítulo 3 apresenta alguns algoritmos de localização *indoor* conhecidos, dentre eles um que já é considerado um marco na história da localização *indoor* e outro que representa o avanço possível com o uso de uma formulação mais elegante. Ambos são importantes por sua eficácia sem utilizar infraestrutura extra, são métodos que apenas requerem uma fase de treinamento e então permite localização com qualidade por um longo período, são apresentadas as coletas necessárias, a forma de tornar isso um problema de regressão. Além disso são mostradas melhorias possíveis nas duas fases que foram desenvolvidas em literatura mais recente.

O Capítulo 4 apresenta a base de dados que será utilizada nos testes, com alguns detalhes da coleta feita e explicando que é uma tentativa de padronizar os testes de algoritmos de localização *indoor* no futuro. Essa base foi criada já que a área sofre com uma grande variedade de testes com bases próprias que não são liberadas facilmente.

O Capítulo 5 é a explicação dos experimentos desenvolvidos e apresentação dos resultados obtidos. É a união do primeiro algoritmo explicado no capítulo 3 com a base de dados apresentada no capítulo 4. Ainda no capítulo 5 temos uma análise dos resultados obtidos nos experimentos.

O Capítulo 6 traz as conclusões sobre o resultado encontrado com uma discussão relacionada além de propostas de trabalhos futuros.

Conceitos Básicos

2.1 Contexto do Usuário

O contexto do usuário é a especificação do ambiente em que ele se encontra. Podemos classificar como contexto para um usuário: quantidade de pessoas próximas, quantidade de dispositivos próximos, funcionalidade dos dispositivos próximos, sinais de WiFi disponíveis, sinais de Bluetooth disponíveis, condições climáticas nas redondezas, localização do usuário, etc.

Uma utilidade possível desse conhecimento é monitoramento. Com o conhecimento contínuo do estado desse usuário é possível monitorar a movimentação de pessoas idosas ou com algum perigo maior de quedas, a qualidade de um sistema de conexão com a internet, a temperatura do ambiente ao redor do usuário ao longo do dia, o movimento de cargas em grandes depósitos de mercadorias ou o caminho percorrido por um atleta em treinamento para corridas.

Outro exemplo de utilidade surge devido à primeira. Com o conhecimento do contexto do usuário através do monitoramento é possível criar aplicações sensíveis ao contexto. Para os casos citados acima podemos criar um sistema para recomendar qual a melhor rede para se conectar, um aplicativo para descobrir se há redes conhecidas disponíveis, um sistema para sugerir uma mudança de posição ou de caminho de acordo com o tráfego no local ou na rota atual, uma plataforma para servir uma propaganda mais direcionada ao usuário baseado nas redondezas de onde ele se encontra.

Essas aplicações só são possíveis graças a esse conhecimento do que há em volta do usuário. E não se limitam à seres humanos já que o "usuário" pode ser um dispositivo simples ou até uma grande máquina. Para coletar esse conhecimento de contexto, possíveis fontes são os diversos sensores disponíveis em equipamentos.

2.2 Localização Indoor e Localização Outdoor

Entrando no contexto de localização, há uma divisão entre o estudo da localização *indoor* e a localização *outdoor*. A palavra *indoor* é do inglês e se refere a algo que está situado no interior de um estabelecimento [Oxford, 2018], algo que foi feito para ser usado no interior de um estabelecimento ou algo que ocorreu no interior de um estabelecimento. Temos como antônimo a palavra *outdoor*, que tem as definições similares às anteriores trocando o ambiente para o exterior de estabelecimentos, para áreas abertas. Logo localização *indoor* pode ser definida como a localização em espaços fechados e a localização *outdoor* é a localização em áreas abertas.

2.3 Localização Outdoor

O problema da localização outdoor já é bem resolvido através do uso da tecnologia de GPS. Foi criada nos Estados Unidos da América e pertence ao governo americano enquanto é operada pela força aérea americana. Esse sistema de navegação foi inicialmente lançado em 1978 e tem novos lançamentos de satélite periódicos como o de 1989 e de 2012 [Space, 2013]. GPS é a escolha padrão para resolver problemas de localização outdoor pois sua acurácia chega a poucos metros facilmente e é um sistema publicamente disponível.

A tecnologia GPS falha ao tratar de ambientes *indoor* pois ela exige que o sensor móvel de GPS esteja em um espaço aberto com pouca ou nenhuma barreira física, já que requer uma linha de visão desobstruída de pelo menos quatro satélites de GPS [Hoffman-Wellenhof, 2001]. Com paredes em vários cômodos e vários níveis de obstruções como o teto de cada andar e objetos, os sinais dos satélites não são se propagam como seria desejado.

2.4 Sistema de Localização Indoor

Dentre os dados fornecidos por dispositivos móveis, podemos destacar ondas de rádio, campos magnéticos, sinais acústicos [Curran, 2011]. Esses sensores têm atuação e presença em ambientes fechados e sua coleta é facilitada pela diversidade de sensores disponíveis num simples aparelho de celular atual.

Como mencionado anteriormente, podemos considerar a localização como um contexto do usuário, aplicações que aproveitam o contexto do usuário estão se proliferando e aumentando o interesse nos sistemas de localização *indoor*. A localização *indoor* continua um desafio que vem sendo atacado por vários métodos diferentes [Wu *et al.*, 2013], [Farid *et al.*, 2013], [Lymberopoulos *et al.*, 2015].

Sabemos que o desenvolvimento de sistemas de localização em tempo real se tornou um aditivo importante para muitos sistemas de localização. A capacidade de rastrear pessoas e objetos em um ambiente *indoor* tem diversas aplicações, como por exemplo segurança infantil, rastreamento de prisioneiros e redes de fornecimento [Curran, 2011].

2.5 Técnicas Baseadas em Radiofrequência

Na última década, houve um grande crescimento de estudos concernentes a localização *indoor* e o baseado em pontos de acesso tornou-se a base para muitos outros estudos. Os pontos de acesso são para redes de área local sem fio, conhecidas como WLAN (*Wireless LAN*). O crescimento dos estudos se dá principalmente pela proliferação de roteadores e de dispositivos móveis.

Hoje em dia WLAN's e celulares podem ser encontrados em qualquer lugar, e se tornaram indispensáveis em nossas vidas, e podemos seguramente esperar que a localização do telefone móvel é a mesma do usuário, sendo assim, pode ser utilizado para prever o comportamento humano e até mesmo interações sociais, entre outros estudos.

Sistemas de posicionamento baseados em WLAN são dependentes do Indicador de Força

do Sinal Recebido (*Received signal strength indication* ou RSSI), nele duas fases são necessárias: calibragem e operação. Na fase de calibragem, um mapa radial da área é construído para prever onde o usuário deve ser detectado. Após isso, na fase operacional, um usuário obtém um sinal indicando a força de todos os pontos de acesso de WLAN's que podem ser detectados dependendo da sua posição e assim cria um conjunto de teste. Esse conjunto é enviado ao servidor e comparado com outros conjuntos de treinamento do mapa radial. Basicamente, a localização do usuário corresponde a uma posição associada com maior similaridade do conjunto de teste com o conjunto do servidor.

Uma das maiores vantagens desses métodos é que eles não precisam de nenhuma instalação de hardware adicional, a não ser as já existentes dos pontos de acesso. No entanto, WLAN não foi idealizada para suportar a função de posicionamento e além do mais, há muitos obstáculos que devem ser levados em conta, como por exemplo a disseminação do sinal em locais fechados é muito difícil de se prever.

A utilização de smartphones cresce a cada ano e já é comum encontrá-los em cada etapa do nosso dia-a-dia. Os métodos baseados em radiofrequência para localização *indoor* então tomam força cada vez maior. O exemplo mais comum é o das redes sem fio baseados no padrão IEEE 802.11, com o termo Wi-Fi sendo frequentemente usado como sinônimo.

A potência recebida em um dispositivo de um sinal de Wi-Fi decai com a distância entre o dispositivo e o roteador. Numa área sem nenhum obstáculo entre o emissor e o receptor do sinal, a potência em certo ponto à uma distância d é mostrado na equação (2.1).

$$P(d)[dBm] = P(d_0)[dBm] - 10 \times n \times \log \frac{d}{d_0} \quad (2.1)$$

Em que d a distância até o local em que está o AP (ponto de acesso ou access point), d_0 a distância de um ponto de referência, n a taxa de perda no caminho com a distância e $P(x)$ a potência percebida numa distância x [Bahl & Padmanabhan, 2000].

Baseado nisso e com a existência de mais APs na área desejada, é possível localizar um dispositivo baseado nas potências percebidas em cada ponto. Isso porque o comportamento das potências tende a ser estável se a posição dos APs se mantém.

2.6 Fingerprinting

Supondo a existência de n diferentes pontos de acesso em um local, chamaremos a potência do i -ésimo AP de ss_i . Para a associação de todas as medidas $\langle ss_1, ss_2, ss_3, \dots, ss_n \rangle$ com as coordenadas $\langle x, y \rangle$, daremos o nome de fingerprint.

2.7 Métricas de Comparação de Fingerprints

Para comparação de medidas das potências coletadas em um local, é criada uma medida de similaridade entre elas. Podemos usar algumas métricas conhecidas na literatura, como a distância euclideana, a distância Manhattan e a distância Chebyshev.

2.7.1 Distância Euclideana

A distância euclidiana é a raiz quadrada da soma das diferenças dos atributos ao quadrado. Isso de forma mais explicada na equação (2.2) e com mais síntese na equação (2.3):

$$SS = \langle ss_1, ss_2, \dots, ss_n \rangle$$

$$SS' = \langle ss'_1, ss'_2, \dots, ss'_n \rangle$$

$$euclidean(SS, SS') = \sqrt{(ss_1 - ss'_1)^2 + (ss_2 - ss'_2)^2 + \dots + (ss_n - ss'_n)^2} \quad (2.2)$$

$$euclidean(SS, SS') = \sqrt{\sum_{k=1}^n (ss_k - ss'_k)^2} \quad (2.3)$$

2.7.2 Distância Manhattan

A distância Manhattan é a soma do módulo das diferenças de cada atributo. Isso pode ser visto de forma mais descritiva na equação (2.4) e de forma mais sintética na equação (2.5).

$$SS = \langle ss_1, ss_2, \dots, ss_n \rangle$$

$$SS' = \langle ss'_1, ss'_2, \dots, ss'_n \rangle$$

$$manhattan(SS, SS') = |ss_1 - ss'_1| + |ss_2 - ss'_2| + \dots + |ss_n - ss'_n| \quad (2.4)$$

$$manhattan(SS, SS') = \sum_{k=1}^n |ss_k - ss'_k| \quad (2.5)$$

2.7.3 Distância Chebyshev

A distância Chebyshev é o máximo entre o módulo das diferenças de cada atributo. Isso pode ser visto de forma mais descritiva na equação (2.6) e como fórmula mais sintética na equação (2.7).

$$SS = \langle ss_1, ss_2, \dots, ss_n \rangle$$

$$SS' = \langle ss'_1, ss'_2, \dots, ss'_n \rangle$$

$$chebyshev(SS, SS') = \max(|ss_1 - ss'_1|, |ss_2 - ss'_2|, \dots, |ss_n - ss'_n|) \quad (2.6)$$

$$chebyshev(SS, SS') = \max_{k=1}^n |ss_k - ss'_k| \quad (2.7)$$

Escolher a melhor métrica para calcular similaridade em um conjunto de dados é um desafio encontrado por profissionais e pesquisadores ao utilizar algum algoritmo de comparação num

conjunto de dados e podem trazer variações efetivas nos resultados como citado em literatura [Shirchorshidi *et al.*, 2015].

2.8 Inteligência Artificial

2.8.1 Método k-NN

Com uma base de *fingerprints*, é possível usar as instâncias desejadas e associar a algum método de inteligência artificial para realizar a regressão.

Amplamente utilizado para problemas de classificação e regressão, temos o *k-nearest neighbors algorithm* (k-NN). É um método não-paramétrico que consiste em encontrar, para uma certa base e uma instância de entrada x , quais as k instâncias de treinamento mais próximas no espaço dos atributos.

2.8.2 Estratégia de União de Vizinhos com Peso Constante

Para regressões, a estratégia clássica é escolher a média dos k vizinhos, chamando o vetor de atributos de cada vizinho de $v_i, i \in \{1, \dots, k\}$:

$$v_{resultante} = \frac{\sum_{i=1}^k v_i}{k}$$

Essa estratégia considera um peso constante para a contribuição de todos os vizinhos.

2.8.3 Estratégia de União de Vizinhos com Peso Inverso da Distância

Outra estratégia é escolher a média ponderada pelo inverso da distância dos k vizinhos, chamando o vetor de atributos da instância em teste de v e o de cada vizinho de $v_i, i \in \{1, \dots, k\}$:

$$v_{resultante} = \frac{\sum_{i=1}^k \frac{1}{c+d(v,v_i)} \times v_i}{\sum_{i=1}^k \frac{1}{c+d(v,v_i)}}$$

Essa estratégia considera um peso maior para a contribuição de vizinhos mais próximos, tentando aliviar o ruído das instâncias mais distantes, o valor c é usado para manter os valores dentro de um limite.

2.8.4 Estratégia de União de Vizinhos com Peso Exponencial na Distância

A última estratégia apresentada é escolher a média ponderada por uma exponencial na distância dos k vizinhos, chamando o vetor de atributos da instância em teste de v e o de cada vizinho de $v_i, i \in \{1, \dots, k\}$:

$$v_{resultante} = \frac{\sum_{i=1}^k e^{-\frac{d(v,v_i)}{c}} \times v_i}{\sum_{i=1}^k e^{-\frac{d(v,v_i)}{c}}}$$

Essa estratégia considera um peso maior para a contribuição de vizinhos mais próximos e é mais suave que usar apenas o inverso da distância, o valor c é usado para manter os valores dentro de um limite.

Trabalhos Relacionados

3.1 RADAR

O algoritmo RADAR [Bahl & Padmanabhan, 2000] é considerado o canônico quando se fala de localização *indoor*, ele é um sistema que utiliza redes Wi-Fi de forma eficaz para conseguir uma regressão de qualidade como na Figura 3.1.

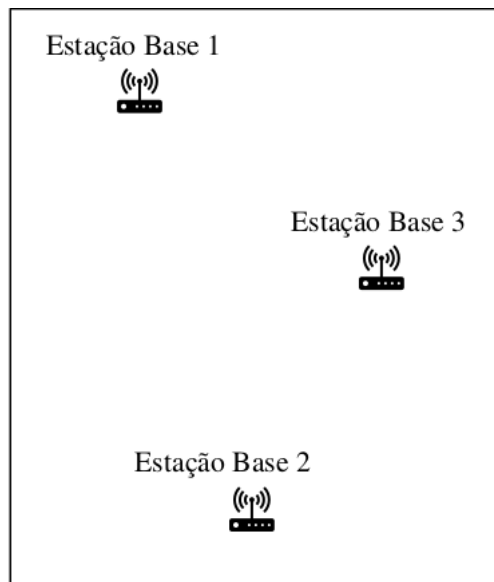


Figura 3.1 Exemplo de localizações das estações base para o método RADAR

A premissa do trabalho é que o estudo da potência do sinal é capaz de inferir a localização atual do usuário. Em gráficos dos resultados é possível ver claramente como a potência decai quando o usuário se afasta e como a potência cresce quando o usuário se aproxima das estações base. Logo é apresentada a ideia principal do artigo do RADAR, triangulação da posição baseado na entrada de força do sinal *SS* (*signal strength*):

$$SS = \langle ss_1, ss_2, \dots, ss_n \rangle$$

Em que ss_i é a potência do sinal da estação i .

Há o dado coletado em vários pontos em volta das estações base e essa fase de coleta é chamada de fase *offline*, já que é feita como um preparo e é feita apenas uma vez. Cada SS é

guardado associado a suas coordenadas $\langle x, y \rangle$, criando um *fingerprint*. A base feita na fase *offline* passa a ser a fonte para posteriores consultas na fase online.

Já na fase online do algoritmo, existe uma consulta para descobrir para um usuário, as coordenadas $\langle x, y \rangle$, tendo apenas a medida dos sinais que estão visíveis para ele. Logo a entrada seria um SS' :

$$SS' = \langle ss'_1, ss'_2, \dots, ss'_n \rangle$$

E esse SS' é comparado com a base de *fingerprints* montada na fase *offline*. O algoritmo foi chamado de NNSS (Nearest Neighbor in Signal Space), seu objetivo é encontrar o vizinho mais próximo de SS' , a métrica usada foi a euclidiana. Logo a distância entre dois sinais SS e SS' é:

$$SS = \langle ss_1, ss_2, \dots, ss_n \rangle$$

$$SS' = \langle ss'_1, ss'_2, \dots, ss'_n \rangle$$

$$d(SS, SS') = \sqrt{(ss_1 - ss'_1)^2 + (ss_2 - ss'_2)^2 + \dots + (ss_n - ss'_n)^2}$$

O *fingerprint* desejado então é o que possui o SS que minimiza a função $d(SS, SS')$.

Os autores sugerem também uma possível melhora utilizando n vizinhos mais próximos. Considerando esses vizinhos é possível corrigir a limitação de que os *fingerprints* estão em posições fixas. Na fase online um usuário pode estar entre dois ou três *fingerprints*. Um exemplo seriam três *fingerprints* com coordenadas $\langle x, y \rangle$ que envolvem o ponto de teste, como mostrado na Figura 3.2.

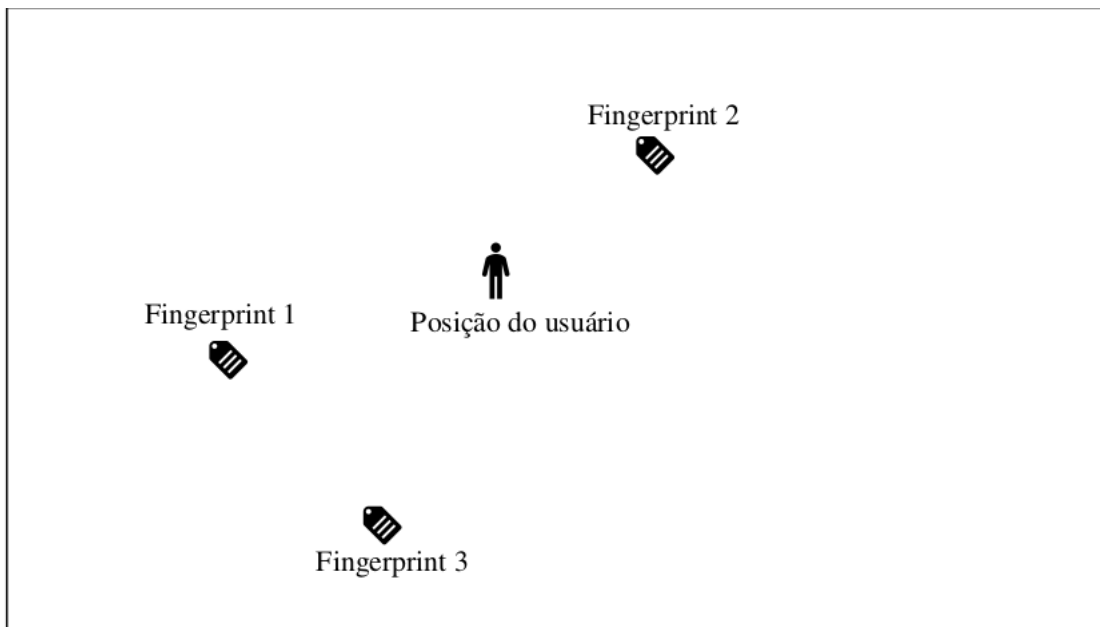


Figura 3.2 Posição problema para algoritmo do RADAR

Com a ideia de usar três vizinhos, por exemplo, é possível corrigir casos como o da Figura 3.2, em que apenas o vizinho mais próximo iria colocar o usuário na posição do Fingerprint 2 quando ele ficaria melhor no centroide dos três *fingerprints*.

A fase *offline* mencionada requer a coleta dos sinais em diferentes pontos do ambiente desejado. No artigo também há a sugestão de criação dos *fingerprints* de forma automática, através do estudo do ambiente para evitar o trabalho manual que é necessário para o método proposto. Foram testados três modelos diferentes, mas apenas um foi considerado apto para casos gerais.

Foi considerado os método *Rayleigh fading model* [Hashemi, 1993]. Apesar de elegante matematicamente, esse método não se encaixa no caso *indoor* comum. A distribuição usada exige que sinais cheguem a usuários com mesma potência, exigência que já falha devido à forte característica de atenuação com cada parede ou obstáculo na linha de visão entre AP e usuário. Uma alternativa era utilizar o *Rician distribution model* que foi descartado devido à sua complexidade, já que exige muitos parâmetros e a idéia era criar um modelo para testes de forma simples e replicável [Bahl & Padmanabhan, 2000]. Finalmente, o modelo escolhido foi o modelo de propagação *Floor Attenuation Factor* (FAF). Esse modelo considera a atenuação devido a obstáculos entre o AP e o usuário. O modelo *Wall Attenuation Factor* (WAF) então permite o cálculo da potência:

$$P(d)[dBm] = \begin{cases} P(d_0)[dBm] - 10 \times n \times \log \frac{d}{d_0} - nW \times WAF & \text{se } nW < C \\ P(d_0)[dBm] - 10 \times n \times \log \frac{d}{d_0} - C \times WAF & \text{se } nW \geq C. \end{cases} \quad (3.1)$$

Sendo d a distância até o local em que está o AP, nW o número de obstáculos físicos entre o AP e o usuário, C o limite de paredes até onde o fator de atenuação faz diferença, WAF o próprio Wall Attenuation Factor citado anteriormente, d_0 a distância de um ponto de referência, n a taxa de perda no caminho com a distância e $P(d_0)$ a potência percebida numa distância d_0 [Bahl & Padmanabhan, 2000].

O valor do WAF é determinado experimentalmente no ambiente a partir de algumas medidas. Então o modelo é usado para popular um grid de *fingerprints* artificial. Esse grid se torna o equivalente aos dados coletados na fase *offline* manual.

3.2 Horus

O Horus [Youssef & Agrawala, 2005] foi o primeiro grande salto de qualidade ao atacar o problema da localização *indoor*. Buscando uma formulação bem mais avançada matematicamente. Para utilizar o Horus é necessária uma extensão do conceito de *fingerprinting*.

Sendo L o conjunto de localizações dos *fingerprints* coletados e sendo ss uma medida das potências lidas dos pontos de acesso disponíveis, o objetivo do Horus é encontrar o ponto $l \in L$ que faz mais sentido de ser associado com ss . Para isso a abordagem usada é a probabilística, a posição l desejada é a que maximiza a probabilidade em (3.2).

$$P(l|ss) \quad (3.2)$$

A probabilidade em (3.2) pode ser manipulada utilizando o teorema de Bayes como pode ser visto em (3.3).

$$P(l|ss) = \frac{P(ss|l) \times P(l)}{P(ss)} \quad (3.3)$$

Como é discutido no trabalho, se o desejado é encontrar o l que maximiza a probabilidade citada, o valor de $P(ss)$ não varia de acordo com o l . Chegamos então a uma definição do valor que queremos alcançar em (3.4).

$$l_{\text{correto}} = \operatorname{argmax}_i (P(ss|l) \times P(l)) \quad (3.4)$$

Para o caso em que temos o dado histórico dos usuários atualizado, o $P(l)$ pode ser definido em função da última ou das últimas localizações disponíveis para aquele usuário. Isso resolve uma parte da probabilidade, a outra parte é a $P(ss|l)$, que envolve a probabilidade de cada medida ter exatamente aquela potência para cada AP, supondo que a medida da potência para o k -ésimo AP seja ss_k , como a propagação de ondas eletromagnéticas é independente, chegamos na equação (3.5).

$$P(ss|l) = \prod_k P(ss_k|l) \quad (3.5)$$

Em que cada probabilidade independente $P(ss_k|l)$ pode ser aproximada para uma gaussiana (conforme comportamento analisado no trabalho original). Assim, para criar uma base para utilizar o método Horus, é preciso coletar várias medidas de potência dos sinais em cada posição l . Com sucesso nessa fase de coleta mais complexa, é possível encontrar parâmetros para a função densidade de probabilidade semelhante a da equação (3.6).

$$fdp(p) = \frac{e^{-(p-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}} \quad (3.6)$$

Em que p seria a potência para um AP específico, μ é a média da distribuição e σ é o desvio padrão da distribuição. Se há várias medidas num único ponto é possível estimar esses valores e então podemos encontrar a probabilidade independente de cada termo da equação (3.6) usando a equação (3.7).

$$P(ss_k|l) = \int_{ss_k-0,5}^{ss_k+0,5} fdp(x)dx \quad (3.7)$$

Em que a integral avança 0,5 para ambos os sentidos em relação a ss_k para englobar todas as possíveis medidas da potência do sinal (o smartphone arredonda a medida da potência para o inteiro mais próximo).

O Horus mostra comparações com o RADAR apresentando uma melhora com apenas o custo extra de múltiplas coletas em cada ponto, sua abordagem probabilística elevou o nível da área de localização *indoor*.

3.3 Variações do RADAR

Outros trabalhos continuam a evoluir o método apresentado no RADAR e no Horus. As melhorias em geral vêm de variações nas várias etapas que constituem os métodos que já estavam consolidados na literatura.

Uma variação possível partindo de métodos já existentes é o uso de métricas alternativas à euclideana. Já que os métodos precisam de uma forma de comparar a similaridade entre *fingerprints*, são possibilidades de uso as métricas Manhattan, Bray-Curtis, Chi-Squared e Mahalanobis [Corte-Valiente *et al.* , 2009].

Trabalhos mais recentes também criam variações apenas na forma de calcular similaridade, como exemplo temos o PiLoc [Luo *et al.* , 2014] que propõe uma métrica específica para comparação de medidas de potências de APs.

A principal contribuição em produções mais atuais é a de como criar uma infraestrutura de forma mais independente e com menos custo como o próprio PiLoc [Luo *et al.* , 2014] e o UnLoc [Wang *et al.* , 2012].

Base de Treinamento e Base de Teste

4.1 Falta de Bases Comuns Para Localização Indoor

O campo de pesquisa de localização *indoor* é um dos principais no estudo da computação móvel já que permite aplicação direta nos diversos sensores existentes. Além disso existe uma escassez de bases de dados para comparação no estado da arte. É comum encontrar trabalhos em que a base utilizada é construída pelos próprios autores ou privada. Essa particularização de bases impede uma comparação completa com os diferentes métodos, já que as condições de criação são específicas demais para replicação contínua.

Em campos como reconhecimento de padrões, uma nova proposta pode realizar testes comparando com uma base já bem conhecida ou até liberando sua base. A criação de uma base comum é uma forma de comparar de forma justa as diferentes metodologias existentes na literatura.

Múltiplas soluções têm sido criadas atacando o problema de localização *indoor* baseadas em *fingerprints*. Para criar um padrão e tentar aumentar a qualidade das pesquisas do tipo surgiu a base de dados UJIIndoorLoc. Essa base foi desenvolvida para sanar esta deficiência. Esperando assim que se torne uma referência de base de dados para se comparar metodologias de localização *indoor*.

4.2 UJIIndoorLoc

UJIIndoorLoc [Torres-Sospedra *et al.*, 2014] possui medidas coletadas em três construções diferentes, em cinco andares (térreo mais quatro). Existe um total de 520 APs visíveis no local, espalhados em vários pontos.

Até o presente momento, as principais características desse conjunto de dados são:

- Cobre uma área de aproximadamente cem mil metros quadrados. Inclui três prédios variando de quatro a cinco andares, dependendo da construção.
- O número de lugares diferentes (pontos de referência) que aparecem na base é 933.
- Foram capturados 21.049 *fingerprints*, sendo 19.938 para treinamento e aprendizado e 1111 para validação e teste.
- Independência dos dados foi garantida coletando o conjunto de validação quatro meses antes de coletar os *fingerprints* para treinamento.

- O número de pontos de acesso visíveis englobando a totalidade de medidas de treinamento e de teste é 520.
- Os dados foram coletados por mais de 20 usuários usando 25 modelos de dispositivos móveis diferentes (alguns usuários usaram mais de um modelo).

Dois aplicativos Android foram utilizados para gerar a base de dados, a saber CaptureLoc e ValidationLoc. Ambas as aplicações usam o mesmo mapa como referência que é publicado na ArcGIS. Esses serviços contêm informação geográfica de partes interna de prédios.

No trabalho atual será considerado o primeiro andar dos edifícios. Dessa forma o foco estará no erro apenas nas coordenadas $\langle x, y \rangle$. Nesse pavimento há um total de 4.369 instâncias de treinamento e um total de 132 instâncias de teste.

Experimentos e Análise

5.1 Proposta

Nesse trabalho, o objetivo é utilizar a base de dados UJIIndoorLoc para testar um método de localização *indoor*. Há uma limitação na base que é a coleta de medidas. A coleta no UJIIndoorLoc foi feita apenas focando na quantidade de pontos, as medidas são de qualidade por possuírem muitos detalhes das potências, localização no plano, andar, identificação do usuário e do dispositivo utilizado para cada instância. Porém métodos como o do Horus e o proposto no PiLoc precisam de uma coleta mais complexa com várias medições no mesmo ponto, o UJIIndoorLoc não é adaptado para esses algoritmos. Já que a base ainda permite uma variação entre métodos baseados apenas numa coleta básica, a aplicação a seguir será do próprio RADAR. Serão testadas várias combinações de métricas diferentes, várias escolhas de número de vizinhos e também de vários algoritmos para definição das coordenadas a partir dos vizinhos.

Os experimentos foram realizados com as instâncias do conjunto de treinamento localizadas no térreo, dessa forma podemos focar no cálculo do erro apenas em duas coordenadas (latitude e longitude). Como foram considerados os três edifícios que compõem a base, temos uma grande área para realizar os testes. A análise é feita em cima dos percentis de erro para cada experimento. A definição de percentil é a seguinte: se o percentil X é de valor y , isso significa que $X\%$ dos erros possuem valor igual ou inferior a y . A métrica de interesse será, no erro euclidiano das medições, a **mediana** ou **percentil 50**. Essa medida significa que se um método tem mediana $7m$ no erro, então metade das medições obtiveram um erro igual ou inferior a $7m$.

5.2 Experimentos com a métrica Manhattan

5.2.1 Utilizando a União de Vizinhos com Peso Constante

Há uma diminuição da mediana do erro ao aumentar a quantidade de vizinhos utilizada. Isso ocorre até um limite quando com 14 vizinhos há o percentil 50 mínimo. A variação pode ser vista com mais clareza na Figura 5.1.

Esse comportamento é esperado já que permite mais flexibilidade para resolver casos como o apontado na Figura 3.2, também é possível perceber que o excesso de vizinhos começa a trazer mais ruído do que melhora a partir de certo ponto pois são considerados vizinhos muito distantes.

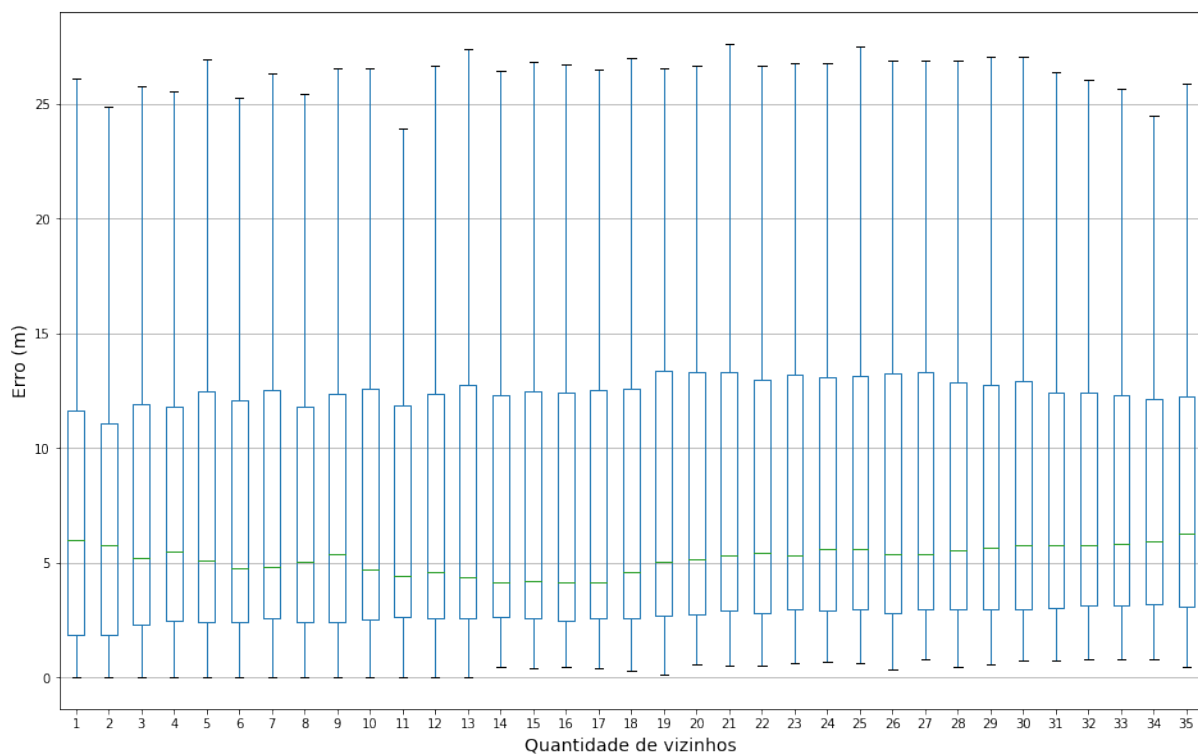


Figura 5.1 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso constante.

Com mais de 14 vizinhos há um aumento cada vez mais visível do erro, com o aumento consistente visível na Figura 5.1. Esse crescimento é justificado pela escolha do peso igual para todos os vizinhos, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.1.

Manhattan e União com Peso Constante			
vizinhos	25%	50%	75%
1	1.864	5.969	11.627
2	1.853	5.782	11.080
3	2.275	5.195	11.912
4	2.480	5.485	11.790
5	2.389	5.066	12.445
6	2.424	4.770	12.088
7	2.583	4.821	12.518
8	2.405	5.027	11.805
9	2.401	5.387	12.360
10	2.510	4.678	12.574
11	2.616	4.412	11.851
12	2.583	4.594	12.346
13	2.585	4.389	12.771
14	2.609	4.169	12.319
15	2.556	4.206	12.476
16	2.464	4.126	12.391
17	2.551	4.138	12.525
18	2.587	4.605	12.602
19	2.677	5.057	13.335
20	2.721	5.122	13.330
21	2.885	5.323	13.280
22	2.802	5.433	12.987
23	2.970	5.315	13.220
24	2.902	5.618	13.078
25	2.990	5.599	13.132
26	2.818	5.388	13.267
27	2.981	5.374	13.315
28	2.969	5.557	12.837
29	2.993	5.634	12.730
30	2.966	5.761	12.925
31	3.013	5.743	12.399
32	3.112	5.772	12.399
33	3.155	5.842	12.320
34	3.211	5.948	12.124
35	3.101	6.282	12.257

Tabela 5.1 Erro em metros para Manhattan e estratégia de união de k vizinhos com peso constante, k variando de 1 a 35.

5.2.2 Utilizando a União de Vizinhos com Peso Inverso da Distância

Há uma diminuição da mediana do erro ao aumentar a quantidade de vizinhos utilizada. Isso ocorre até um limite quando com 17 vizinhos há o percentil 50 mínimo. A variação pode ser vista com mais clareza na Figura 5.2.

É possível perceber que o excesso de vizinhos começa a trazer mais ruído do que melhora a partir de certo ponto pois são considerados vizinhos muito distantes.

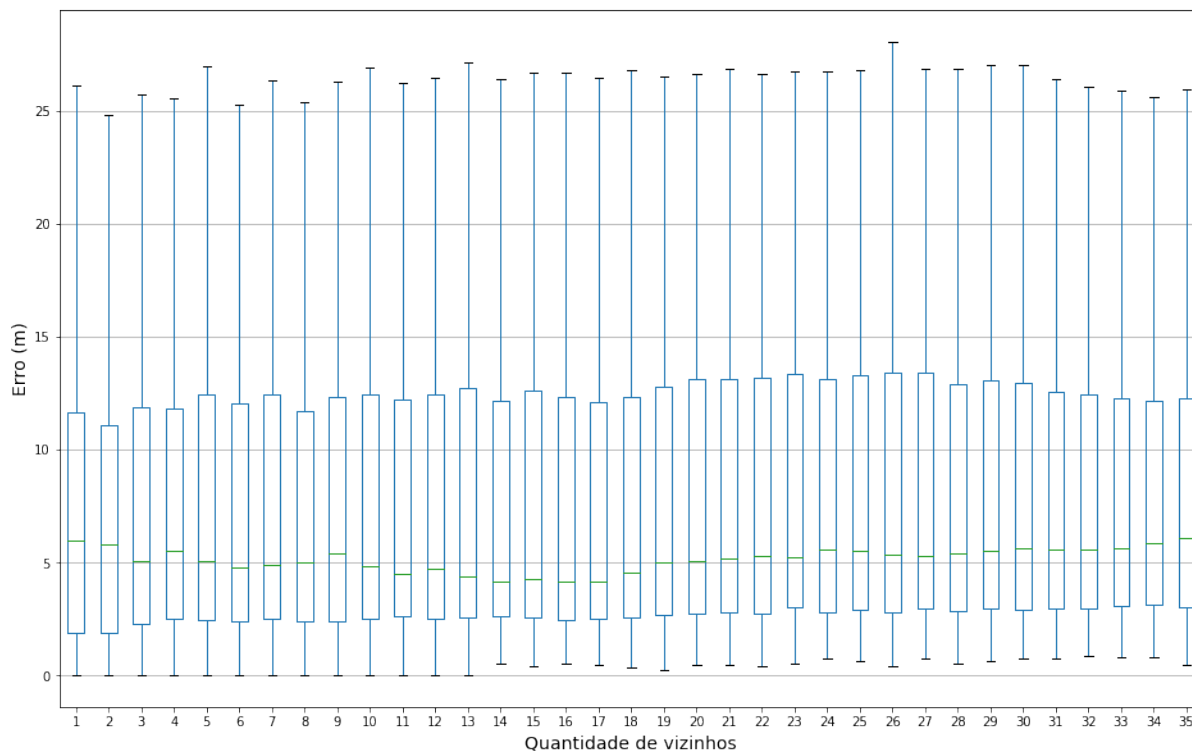


Figura 5.2 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso inverso da distância.

Com mais de 17 vizinhos há um aumento do erro, com o aumento suave visível na Figura 5.2. Esse crescimento é justificado pela adição de ruído para cada novo vizinho, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.2.

Manhattan e União com Peso Inverso da Distância			
vizinhos	25%	50%	75%
1	1.864	5.969	11.627
2	1.854	5.782	11.052
3	2.251	5.031	11.856
4	2.498	5.498	11.790
5	2.425	5.081	12.464
6	2.416	4.801	12.044
7	2.489	4.863	12.416
8	2.372	5.012	11.687
9	2.371	5.374	12.299
10	2.520	4.804	12.457
11	2.631	4.507	12.233
12	2.514	4.704	12.451
13	2.549	4.392	12.743
14	2.625	4.136	12.161
15	2.568	4.240	12.583
16	2.455	4.144	12.338
17	2.526	4.134	12.116
18	2.554	4.542	12.319
19	2.646	5.007	12.782
20	2.719	5.030	13.091
21	2.764	5.176	13.117
22	2.754	5.308	13.176
23	2.994	5.204	13.360
24	2.805	5.559	13.139
25	2.923	5.514	13.268
26	2.788	5.331	13.377
27	2.929	5.305	13.385
28	2.840	5.384	12.869
29	2.934	5.498	13.040
30	2.885	5.639	12.951
31	2.980	5.550	12.525
32	2.980	5.550	12.451
33	3.075	5.618	12.295
34	3.110	5.841	12.179
35	3.024	6.104	12.273

Tabela 5.2 Erro em metros para métrica Manhattan e estratégia de união de k vizinhos com peso inverso da distância, k variando de 1 a 35.

5.2.3 Utilizando a União de Vizinhos com Peso Exponencial na Distância

Há uma diminuição da mediana do erro ao aumentar a quantidade de vizinhos utilizada. Isso ocorre até um limite quando com 16 vizinhos há o percentil 50 mínimo. A variação pode ser vista com mais clareza na Figura 5.3.

É possível perceber que o excesso de vizinhos começa a trazer mais ruído do que melhora a partir de certo ponto pois são considerados vizinhos muito distantes.

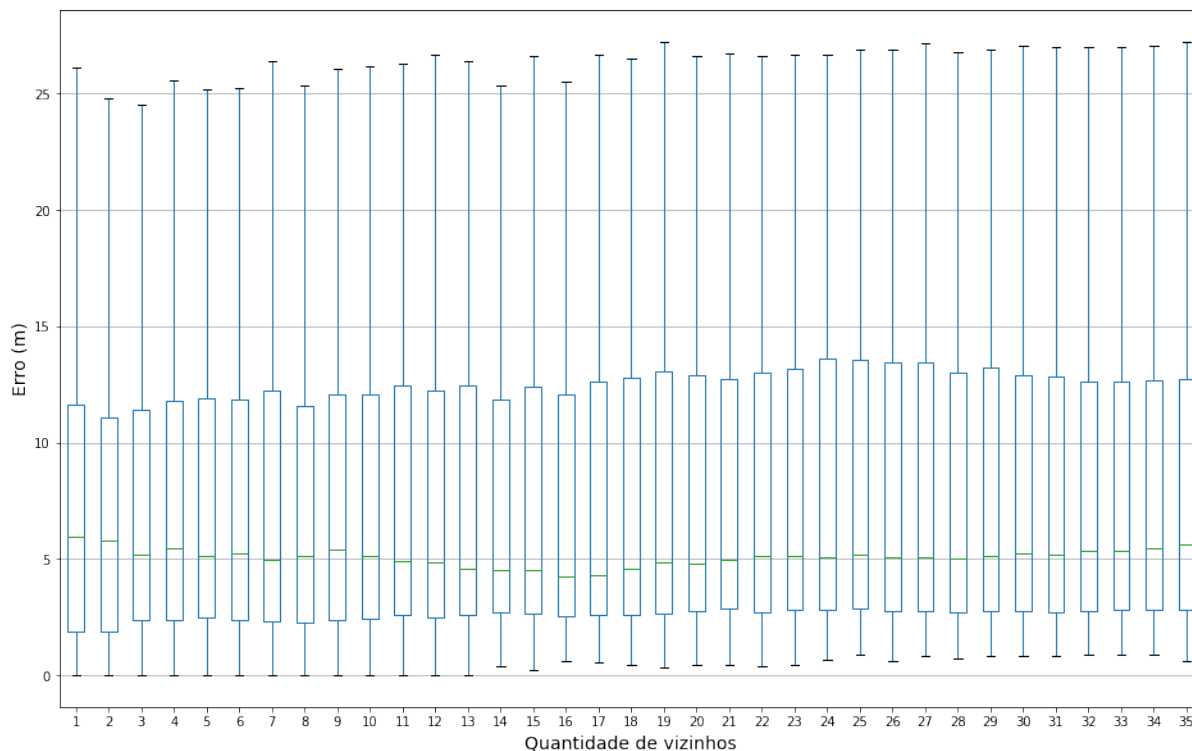


Figura 5.3 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso exponencial na distância.

Esse crescimento é justificado pela adição de ruído para cada novo vizinho, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.3.

Manhattan e União com Peso Exponencial na Distância			
vizinhos	25%	50%	75%
1	1.864	5.969	11.627
2	1.857	5.774	11.085
3	2.363	5.196	11.418
4	2.401	5.461	11.790
5	2.477	5.127	11.921
6	2.352	5.213	11.870
7	2.335	4.983	12.214
8	2.272	5.134	11.553
9	2.353	5.418	12.049
10	2.419	5.107	12.051
11	2.580	4.923	12.453
12	2.498	4.862	12.227
13	2.581	4.585	12.450
14	2.678	4.546	11.861
15	2.646	4.513	12.416
16	2.561	4.254	12.042
17	2.618	4.298	12.595
18	2.576	4.575	12.768
19	2.651	4.832	13.072
20	2.778	4.778	12.919
21	2.863	4.951	12.721
22	2.717	5.110	12.998
23	2.802	5.112	13.153
24	2.796	5.085	13.609
25	2.851	5.160	13.552
26	2.783	5.053	13.449
27	2.752	5.072	13.438
28	2.700	5.038	13.032
29	2.762	5.142	13.201
30	2.738	5.224	12.881
31	2.693	5.186	12.847
32	2.774	5.332	12.603
33	2.808	5.371	12.617
34	2.835	5.453	12.689
35	2.825	5.651	12.757

Tabela 5.3 Erro em metros para métrica Manhattan e estratégia de união de k vizinhos com peso exponencial na distância, k variando de 1 a 35.

5.3 Experimentos com a métrica euclideana

5.3.1 Utilizando a União de Vizinhos com Peso Constante

Há uma variação da mediana do erro ao aumentar a quantidade de vizinhos utilizada com crescimentos e decrescimentos. Essa variação se estabiliza com 24 vizinhos onde há o percentil 50 mínimo. A variação pode ser vista com mais clareza na Figura 5.4.

Esse comportamento é esperado já que permite mais flexibilidade para resolver casos como o apontado na Figura 3.2, também é possível perceber que o excesso de vizinhos começa a trazer mais ruído do que melhora a partir de deste ponto pois são considerados vizinhos muito distantes.

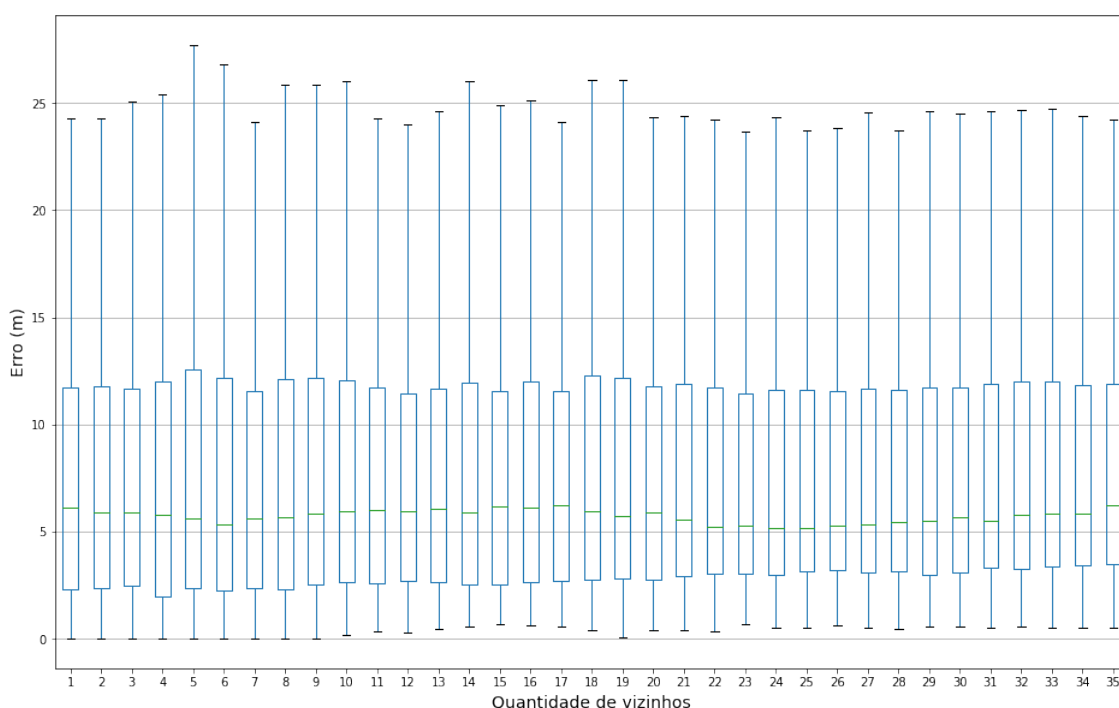


Figura 5.4 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso constante.

Com mais de 24 vizinhos, o aumento do erro é contínuo, com o aumento consistente até o limite testado. Esse crescimento é justificado pela escolha do peso igual para todos os vizinhos, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.4.

Euclideana e União com Peso Constante			
vizinhos	25%	50%	75%
1	2.331	6.135	11.721
2	2.357	5.871	11.790
3	2.495	5.901	11.667
4	1.996	5.795	11.983
5	2.344	5.616	12.542
6	2.272	5.358	12.188
7	2.357	5.615	11.580
8	2.325	5.690	12.119
9	2.513	5.859	12.173
10	2.668	5.933	12.079
11	2.560	6.007	11.740
12	2.672	5.972	11.443
13	2.634	6.047	11.667
14	2.535	5.897	11.940
15	2.546	6.168	11.541
16	2.658	6.100	11.982
17	2.681	6.231	11.564
18	2.742	5.944	12.305
19	2.799	5.725	12.156
20	2.736	5.907	11.806
21	2.936	5.565	11.901
22	3.028	5.209	11.701
23	3.014	5.289	11.461
24	2.965	5.161	11.612
25	3.119	5.179	11.588
26	3.202	5.264	11.544
27	3.087	5.314	11.684
28	3.133	5.442	11.640
29	3.001	5.478	11.702
30	3.095	5.657	11.738
31	3.319	5.522	11.886
32	3.276	5.777	12.014
33	3.361	5.837	12.000
34	3.448	5.865	11.848
35	3.486	6.222	11.885

Tabela 5.4 Erro em metros para métrica euclideana e estratégia de união de k vizinhos com peso constante, k variando de 1 a 35.

5.3.2 Utilizando a União de Vizinhos com Peso Inverso da Distância

Há uma variação da mediana do erro ao aumentar a quantidade de vizinhos utilizada com crescimentos e decrescimentos. Essa variação se estabiliza com 24 vizinhos onde há o percentil 50 mínimo. A variação pode ser vista com mais clareza na Figura 5.5.

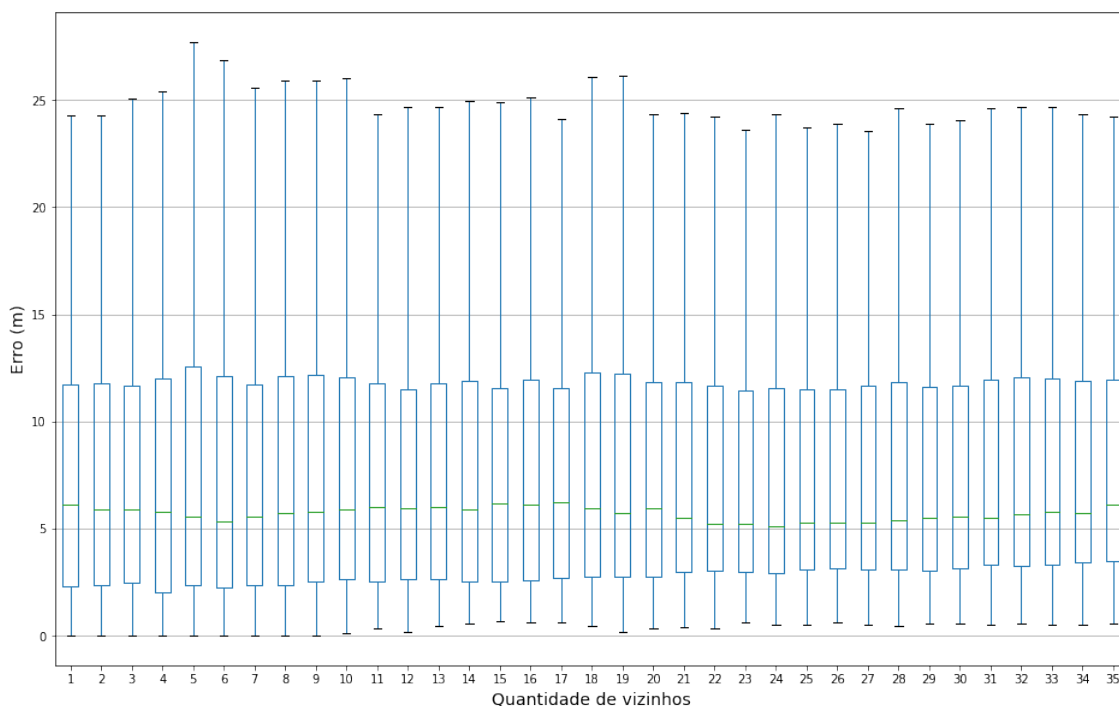


Figura 5.5 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica euclidiana usando a união de vizinhos com peso inverso da distância.

Com mais de 24 vizinhos, o aumento do erro é contínuo, com o aumento consistente até o limite testado. Esse crescimento é justificado pela adição de ruído para cada novo vizinho, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.5.

Euclideana e União com Peso Inverso da Distância			
vizinhos	25%	50%	75%
1	2.331	6.135	11.721
2	2.366	5.871	11.790
3	2.466	5.878	11.667
4	1.999	5.800	11.984
5	2.343	5.562	12.554
6	2.253	5.352	12.091
7	2.362	5.552	11.699
8	2.365	5.708	12.129
9	2.516	5.806	12.148
10	2.629	5.888	12.080
11	2.556	6.003	11.766
12	2.666	5.922	11.522
13	2.618	6.020	11.786
14	2.529	5.897	11.906
15	2.536	6.165	11.574
16	2.573	6.104	11.956
17	2.684	6.231	11.551
18	2.748	5.972	12.262
19	2.741	5.739	12.209
20	2.748	5.928	11.827
21	2.953	5.509	11.841
22	3.034	5.226	11.658
23	2.958	5.239	11.426
24	2.916	5.099	11.562
25	3.082	5.249	11.486
26	3.149	5.253	11.488
27	3.103	5.248	11.655
28	3.099	5.370	11.822
29	3.035	5.476	11.607
30	3.124	5.532	11.663
31	3.314	5.482	11.946
32	3.265	5.690	12.081
33	3.310	5.777	12.000
34	3.401	5.747	11.916
35	3.462	6.097	11.956

Tabela 5.5 Erro em metros para métrica euclideana e estratégia de união de k vizinhos com peso inverso da distância, k variando de 1 a 35.

5.3.3 Utilizando a União de Vizinhos com Peso Exponencial na Distância

Há uma variação da mediana do erro ao aumentar a quantidade de vizinhos utilizada, essa variação alterna entre negativa e positiva. O erro estabiliza com 24 vizinhos onde há o percentil 50 mínimo. A variação pode ser vista com mais clareza na Figura 5.6.

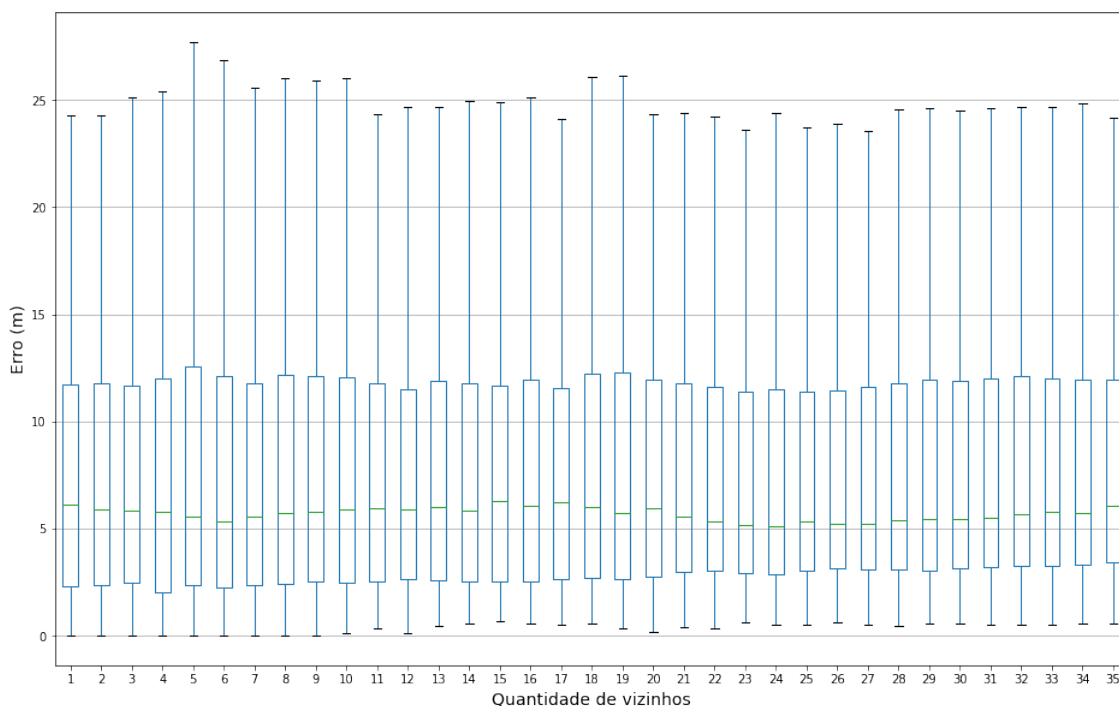


Figura 5.6 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica euclidiana usando a união de vizinhos com peso inverso da distância.

Com mais de 24 vizinhos, o aumento do erro é contínuo até o limite testado. Esse crescimento é justificado pela adição de ruído para cada novo vizinho, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.6.

Euclideana e União com Peso Exponencial na Distância			
vizinhos	25%	50%	75%
1	2.331	6.135	11.721
2	2.375	5.871	11.790
3	2.456	5.850	11.667
4	2.001	5.802	11.983
5	2.340	5.538	12.564
6	2.244	5.348	12.102
7	2.364	5.545	11.802
8	2.421	5.743	12.166
9	2.516	5.799	12.112
10	2.483	5.878	12.082
11	2.535	5.967	11.779
12	2.640	5.920	11.514
13	2.599	5.984	11.876
14	2.523	5.838	11.807
15	2.536	6.257	11.667
16	2.502	6.078	11.934
17	2.647	6.230	11.539
18	2.671	6.004	12.246
19	2.641	5.750	12.301
20	2.765	5.942	11.925
21	2.960	5.569	11.788
22	3.023	5.316	11.590
23	2.940	5.178	11.375
24	2.889	5.109	11.488
25	3.028	5.355	11.372
26	3.137	5.203	11.458
27	3.103	5.220	11.616
28	3.080	5.378	11.806
29	3.023	5.466	11.969
30	3.139	5.468	11.880
31	3.223	5.518	11.987
32	3.241	5.656	12.112
33	3.241	5.757	11.990
34	3.315	5.727	11.950
35	3.423	6.050	11.964

Tabela 5.6 Erro em metros para métrica euclideana e estratégia de união de k vizinhos com peso exponencial na distância, k variando de 1 a 35.

5.4 Experimentos com a métrica Chebyshev

5.4.1 Utilizando a União de Vizinhos com Peso Constante

Há uma variação da mediana do erro ao aumentar a quantidade de vizinhos utilizada com crescimentos e decrescimentos. Para essa métrica, o aumento de quantidade de vizinhos apenas deteriora o resultado após o mínimo em 3 vizinhos. A variação pode ser vista com mais clareza na Figura 5.7.

Por ser uma métrica que considera vizinhos considerando o peso de apenas o maior atributo, é visível a piora com a adição de vizinhos. O ruído passa a aumentar a partir ponto mínimo em 4 vizinhos pois são considerados vizinhos muito distantes.

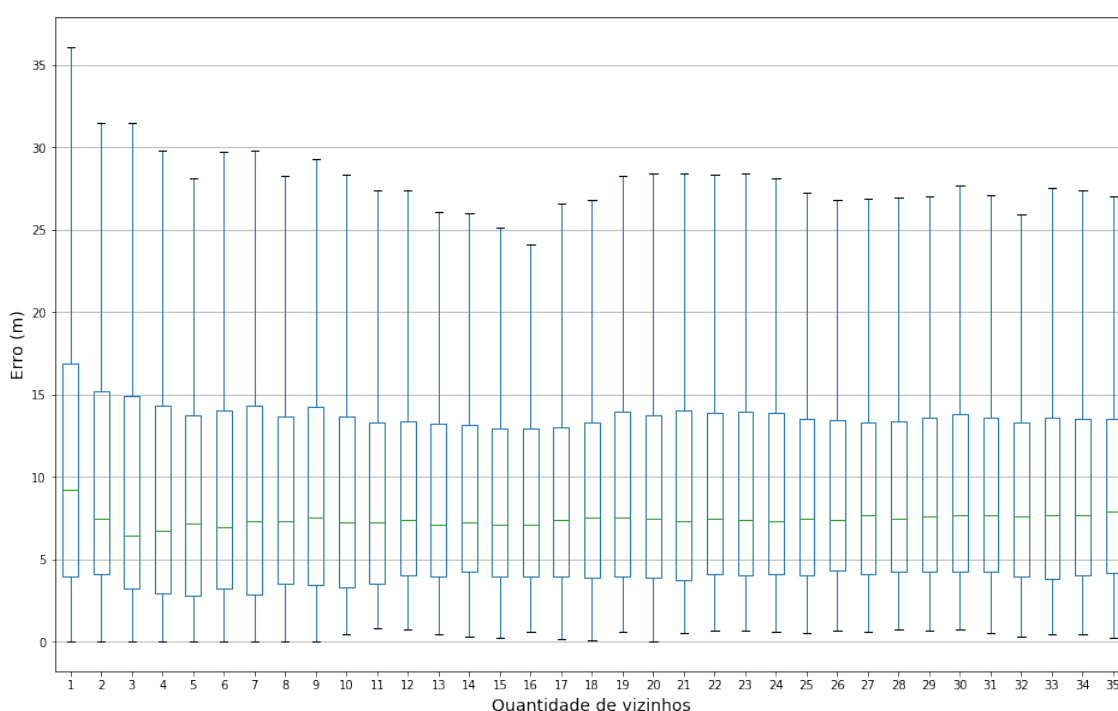


Figura 5.7 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Manhattan usando a união de vizinhos com peso constante.

Com mais de 3 vizinhos, o aumento do erro é contínuo, com o aumento consistente até o limite testado. Esse crescimento é justificado pela escolha do peso igual para todos os vizinhos, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.7.

Chebyshev e União com Peso Constante			
vizinhos	25%	50%	75%
1	3.961	9.175	16.870
2	4.091	7.476	15.197
3	3.224	6.412	14.884
4	2.960	6.758	14.285
5	2.768	7.168	13.714
6	3.214	6.938	14.039
7	2.875	7.303	14.354
8	3.526	7.311	13.638
9	3.453	7.510	14.265
10	3.307	7.237	13.647
11	3.544	7.208	13.302
12	4.028	7.379	13.387
13	3.968	7.062	13.231
14	4.211	7.221	13.140
15	3.942	7.076	12.942
16	3.946	7.075	12.915
17	3.936	7.361	13.036
18	3.891	7.522	13.281
19	3.951	7.543	13.967
20	3.872	7.487	13.757
21	3.725	7.328	14.051
22	4.081	7.424	13.858
23	3.992	7.375	13.986
24	4.063	7.335	13.853
25	3.992	7.418	13.492
26	4.301	7.358	13.427
27	4.079	7.670	13.296
28	4.207	7.460	13.367
29	4.256	7.614	13.615
30	4.212	7.679	13.794
31	4.218	7.656	13.564
32	3.946	7.613	13.292
33	3.773	7.696	13.559
34	3.998	7.704	13.546
35	4.200	7.900	13.539

Tabela 5.7 Erro em metros para métrica Chebyshev e estratégia de união de k vizinhos com peso constante, k variando de 1 a 35.

5.4.2 Utilizando a União de Vizinhos com Peso Inverso da Distância

Há uma variação da mediana do erro ao aumentar a quantidade de vizinhos utilizada com crescimentos e decréscimos. Para essa métrica, o aumento de quantidade de vizinhos apenas deteriora o resultado após o mínimo em 3 vizinhos. A variação pode ser vista com mais clareza na Figura 5.8.

Por ser uma métrica que considera vizinhos considerando o peso de apenas o maior atributo, é visível a piora com a adição de vizinhos. O ruído passa a aumentar a partir ponto mínimo em 4 vizinhos pois são considerados vizinhos muito distantes.

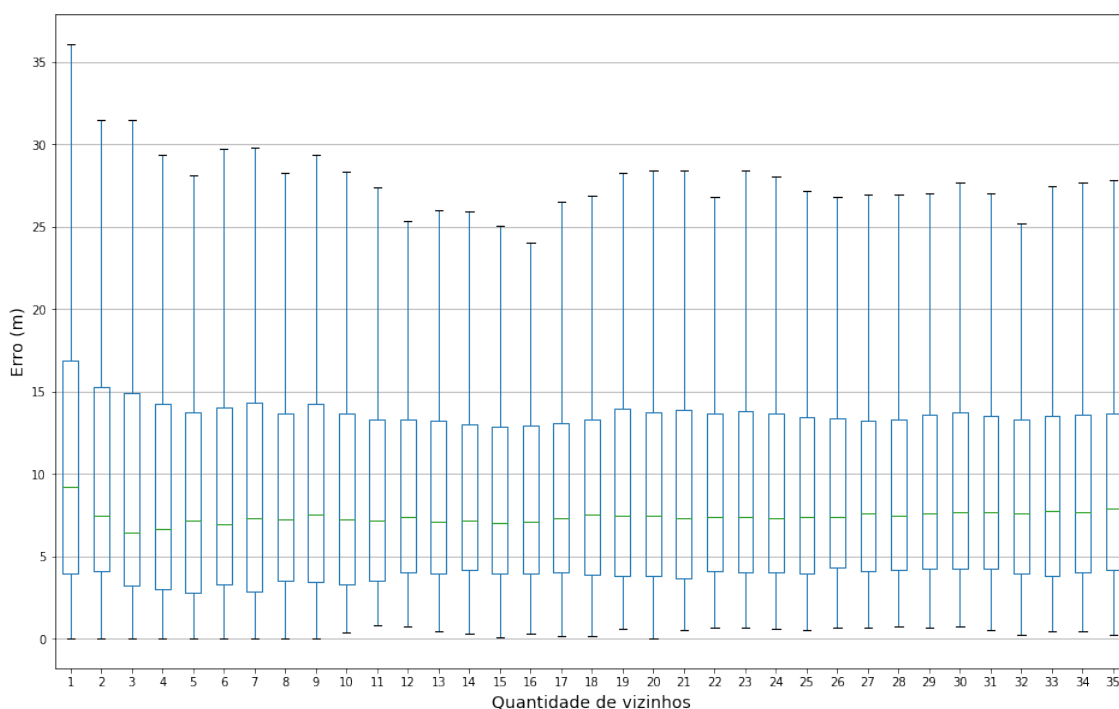


Figura 5.8 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Chebyshev usando a união de vizinhos com peso inverso da distância.

Com 4 ou mais vizinhos, o erro só atinge valores superiores ao obtido com 3 vizinhos, com o aumento consistente até o limite testado. Esse crescimento é justificado pela adição de ruído para cada novo vizinho, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.8.

Chebyshev e União com Peso Inverso da Distância			
vizinhos	25%	50%	75%
1	3.961	9.175	16.870
2	4.091	7.476	15.258
3	3.220	6.412	14.884
4	2.983	6.668	14.276
5	2.765	7.168	13.716
6	3.265	6.962	14.046
7	2.873	7.317	14.354
8	3.521	7.261	13.655
9	3.452	7.501	14.236
10	3.322	7.204	13.642
11	3.544	7.187	13.317
12	4.054	7.357	13.297
13	3.961	7.122	13.201
14	4.183	7.190	12.986
15	3.941	7.033	12.832
16	3.921	7.071	12.940
17	3.997	7.300	13.054
18	3.865	7.498	13.287
19	3.832	7.483	13.954
20	3.809	7.481	13.739
21	3.657	7.328	13.870
22	4.079	7.409	13.676
23	4.058	7.412	13.834
24	4.007	7.332	13.668
25	3.935	7.360	13.409
26	4.296	7.350	13.359
27	4.079	7.580	13.233
28	4.201	7.472	13.302
29	4.256	7.615	13.588
30	4.207	7.680	13.765
31	4.220	7.684	13.520
32	3.943	7.595	13.277
33	3.798	7.716	13.530
34	3.991	7.668	13.594
35	4.169	7.919	13.646

Tabela 5.8 Erro em metros para métrica Chebyshev e estratégia de união de k vizinhos com peso inverso da distância, k variando de 1 a 35.

5.4.3 Utilizando a União de Vizinhos com Peso Exponencial na Distância

Há uma variação da mediana do erro ao aumentar a quantidade de vizinhos utilizada com crescimentos e decréscimos. Para essa métrica, o aumento de quantidade de vizinhos apenas deteriora o resultado após o mínimo em 3 vizinhos. A variação pode ser vista com mais clareza na Figura 5.9.

Por ser uma métrica que considera vizinhos considerando o peso de apenas o maior atributo, é visível a piora com a adição de vizinhos. O ruído passa a aumentar a partir ponto mínimo em 4 vizinhos pois são considerados vizinhos muito distantes.

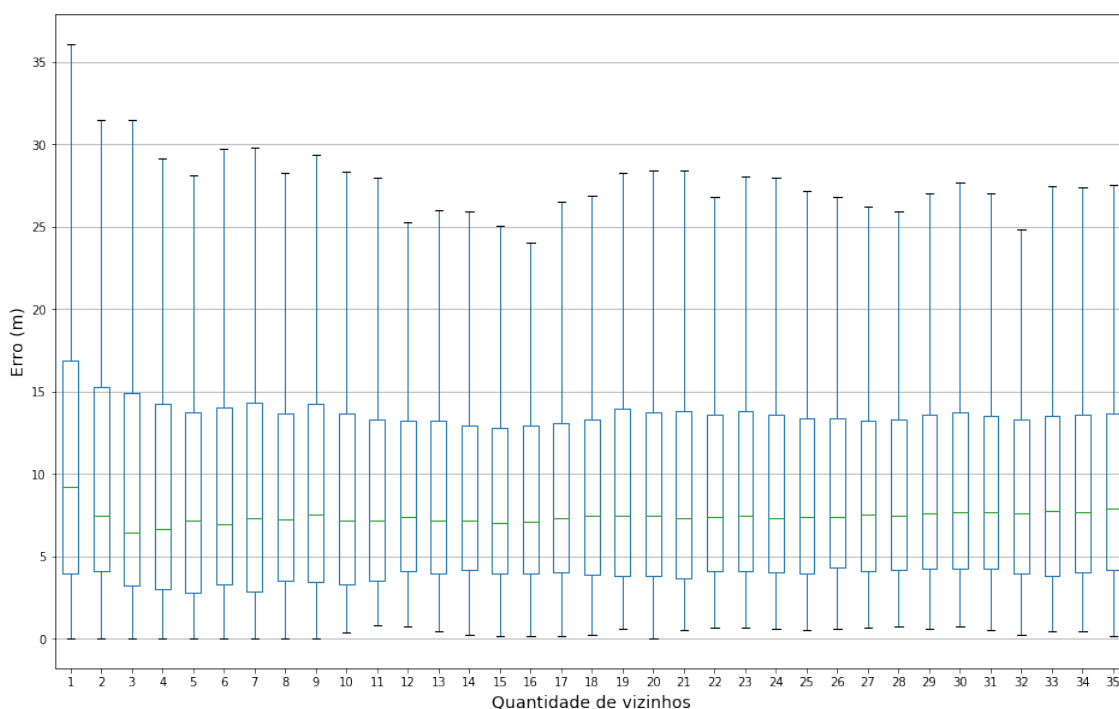


Figura 5.9 Visão em Box-Plot da evolução do erro com o crescimento do número de vizinhos para a métrica Chebyshev usando a união de vizinhos com peso inverso da distância.

Com 4 ou mais vizinhos, o erro só atinge valores superiores ao obtido com 3 vizinhos, com um aumento consistente até o limite testado. Esse crescimento é justificado pela adição de ruído para cada novo vizinho, que causa uma movimentação do resultado final para cada vez mais longe do ideal.

A variação dos valores e o valor ótimo podem ser vistos em detalhes na Tabela 5.9.

Chebyshev e União com Peso Exponencial na Distância			
vizinhos	25%	50%	75%
1	3.961	9.175	16.870
2	4.091	7.476	15.270
3	3.220	6.412	14.884
4	2.996	6.668	14.273
5	2.803	7.168	13.718
6	3.267	6.969	14.048
7	2.869	7.320	14.354
8	3.521	7.247	13.671
9	3.440	7.498	14.225
10	3.327	7.191	13.638
11	3.544	7.178	13.328
12	4.064	7.350	13.221
13	3.959	7.127	13.202
14	4.177	7.190	12.941
15	3.941	7.038	12.795
16	3.923	7.069	12.950
17	3.997	7.291	13.088
18	3.859	7.489	13.293
19	3.778	7.475	13.941
20	3.789	7.473	13.727
21	3.632	7.328	13.813
22	4.078	7.403	13.619
23	4.066	7.428	13.775
24	4.000	7.329	13.607
25	3.917	7.360	13.373
26	4.280	7.350	13.334
27	4.079	7.556	13.204
28	4.199	7.477	13.274
29	4.256	7.615	13.554
30	4.206	7.680	13.721
31	4.221	7.683	13.500
32	3.947	7.580	13.260
33	3.813	7.727	13.511
34	3.989	7.660	13.598
35	4.160	7.922	13.670

Tabela 5.9 Erro em metros para métrica Chebyshev e estratégia de união de k vizinhos com peso exponencial na distância, k variando de 1 a 35.

5.5 Análise dos Resultados

Nos trabalhos da literatura, é comum utilizar o percentil 50 (a mediana) do erro como medida mais importante [Bahl & Padmanabhan, 2000] [Youssef & Agrawala, 2005] [Luo *et al.*, 2014], já que os piores casos geralmente levam o erro a valores bem distintos do caso geral. A média então passa a ser um valor que não pode ser visto isoladamente.

A base de dados se mostrou bastante rica em instâncias de treinamento. O que permitiu ganhos na qualidade da regressão com o uso de vários vizinhos próximos ao invés da estratégia básica de escolher o mais próximo.

Os melhores resultados de erro mediano ficaram em 4,126m. Esse resultado é 40% pior do que o RADAR em sua base própria, que chegou a 2,94m. Ainda assim é um erro que pode ser aceito para uma solução amadora de entrega baseada em contexto. Não é um erro aceitável para um sistema de rastreamento de cargas em um armazém, já que o percentil 75 supera 10m.

O erro sobe vertiginosamente devido ao uso de várias construções na base, fazendo destoa-rem os valores quando há um erro na regressão que leva ao edifício errado.

Uma comparação direta dos melhores resultados pode ser vista na Tabela 5.10.

Métrica e Estratégia para Unir Vizinhos	vizinhos	50%
Manhattan e União com Peso Constante	16	4.126
Manhattan e União com Peso Inverso da Distância	17	4.134
Manhattan e União com Peso Exponencial na Distância	16	4.254
Euclideana e União com Peso Constante	24	5.161
Euclideana e União com Peso Inverso da Distância	24	5.099
Euclideana e União com Peso Exponencial na Distância	24	5.109
Chebyshev e União com Peso Constante	3	6.412
Chebyshev e União com Peso Inverso da Distância	3	6.412
Chebyshev e União com Peso Exponencial na Distância	3	6.412

Tabela 5.10 Comparação dos melhores resultados no percentil 50 das várias métricas e estratégias, em metros.

A métrica Manhattan teve uma performance boa para vários parâmetros diferentes. O percentil 50 ficou abaixo de 6m para quase todas as estratégias aplicadas, só ficou com valores superiores quando usados 35 vizinhos e apenas nas estratégias de união dos vizinhos com peso constante e com o inverso da distância como peso. Além de chegar a valores de percentil 50 de 4,1m nas melhores estratégias.

A distância euclidiana serviu bem como distância, trazendo um erro nos melhores casos com mediana de 5,2m da medida original. Isso demonstra a qualidade da métrica canônica em problemas de comparação entre medidas multidimensionais.

A métrica de Chebyshev foi a que demonstrou os resultados com maior erro. Essa é uma métrica que acaba por simplificar o cálculo de distância, já que considera apenas o máximo das diferenças entre atributos. Logo ela é uma boa escolha para problemas com poucas dimensões ou com grandes diferenças apenas em poucos atributos muito significativos. Por ser um problema multidimensional, há diversos atributos que são as potências percebidas de cada fonte

de sinal de Wi-Fi. Além disso, há uma importância individual de cada sinal para cada região, esses pontos fizeram a métrica não ser tão bem sucedida.

Conclusão e Trabalhos Futuros

6.1 Conclusão

O objetivo deste trabalho foi o de apresentar um estudo sobre contexto de usuário e a importância da localização como um desses contextos, a diferenciação das localizações indoor e outdoor e a apresentação de técnicas de localização indoor utilizando fingerprints. Dando seguimento ao estudo, foi feita uma aplicação de um método de localização indoor em uma base especificamente desenvolvida para a área. Mostrando a capacidade de criar um sistema de forma direta apenas com um tipo de dado (o mapa de fingerprints).

A literatura na área de localização indoor está evoluindo e vários métodos já foram desenvolvidos com soluções novas ou ajustes em soluções existentes. O problema específico da localização utilizando sinais de radiofrequência já permite aplicações em tempo real utilizando os métodos apresentados como o RADAR, Horus e com as mudanças sugeridas na literatura.

O conjunto de testes utilizado, o UJIIndoorLoc foi apresentado como uma possibilidade de padrão para testes de novos métodos que não precisem de alta complexidade da coleta, sendo uma opção válida para métodos baseados em fingerprint. Foi explicado como os diversos trabalhos dessa área de pesquisa têm ambientes dificilmente replicáveis, sendo um problema para a comparação justa das criações posteriores.

A aplicação do algoritmo do RADAR na base foi capaz de aproximar as coordenadas com suas variações de métricas de similaridade, esse trabalho permitiu mostrar uma aplicação real do UJIIndoorLoc.

Nos experimentos foram descobertas variações no uso das similaridades, sendo a Manhattan a mais eficaz para o caso mediano, o que faz sentido já que é um problema multidimensional em que a diferença em cada uma das medidas de potência dos APs pode influenciar. Esse resultado foi seguido de perto da métrica euclideana, também capaz de considerar a diferença em cada coordenada. A métrica Chebyshev mostrou um erro mais elevado sendo pior que as alternativas em todos os casos, mostrando que não é possível considerar demais apenas uma das diferenças de potência.

O teste também mostrou algumas limitações do método. Para áreas muito grandes o erro pode crescer muito no pior caso, já que não existe uma filtragem intermediária de fingerprints, as instâncias de treinamento podem ser separadas em várias pequenas áreas ou por edifícios e uma fase intermediária pode encontrar essa área restrita e carregar apenas os fingerprints internos à área, o que pode melhorar o erro.

A métrica Manhattan se mostrou a de melhor qualidade para a base escolhida, com um erro menor para a maior parte dos casos e também por ser menos custosa computacionalmente. A métrica euclidiana chegou a resultados da mesma qualidade e a métrica de Chebyshev acabou

sendo descartada nesse método escolhido devido à alta taxa de erro.

As estratégias apresentadas para união dos vizinhos mais próximos em uma localização não causaram uma grande variação. Esse fato pode ser explicado pela quantidade de instâncias de treinamento bem distribuídas pelo mapa do local. Dessa forma, a escolha dos vizinhos influenciou bem mais que a estratégia usada para extrair a localização final. Ainda assim o efeito foi perceptível, as métricas que garantem um peso maior para instâncias mais próximas suavizaram o erro em casos com mais de 20 vizinhos. Além disso a estratégia com o peso baseado na exponencial da distância conseguiu manter todas as aplicações na métrica euclidiana com percentil 50 abaixo de $6m$.

6.2 Trabalhos Futuros

É possível quebrar a localização em duas fases. Como a base é muito grande, as instâncias de treinamento de cada edifício poderiam ser separadas em bases menores. Então uma fase de seleção inicial teria o objetivo de descobrir o edifício correto da busca. A segunda fase seria então a regressão possivelmente mais precisa com um conjunto de treinamento apenas para de instâncias internas à construção correta.

Outra possibilidade é a seleção de atributos por área. Como a presença do sinal de cada dispositivo emissor tem um alcance limitado, é possível fazer uma localização considerando apenas os atributos que tem alta influência em cada área. Essa estratégia também teria maior proveito em duas fases.

Possíveis expansões podem ser criadas com novas bases de dados, bases com várias coletas no mesmo ponto para permitir o uso de estratégias probabilísticas similares ao Horus e o uso de métricas mais complexas.

Referências Bibliográficas

- [Bahl & Padmanabhan, 2000] Bahl, P., & Padmanabhan, V.N. 2000. RADAR: an in-building RF-based user location and tracking system. *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, **2**, 775–784. Institute of Electrical & Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/infcom.2000.832252>.
- [Corte-Valiente *et al.* , 2009] Corte-Valiente, A., Gómez-Pulido, J. Manuel, & Gutiérrez-Blanco, O. 2009. Efficient Techniques and Algorithms for Improving Indoor Localization Precision on WLAN Networks Applications. *International Journal of Communications, Network and System Sciences*, **2**(7), 645–651.
- [Curran, 2011] Curran, Kevin. 2011. An evaluation of indoor location determination technologies. *Journal of Location Based Services*, **5**(2), 775–784. Institute of Electrical & Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/infcom.2000.832252>.
- [Farid *et al.* , 2013] Farid, Zahid, Nordin, Rosdiadee, & Ismail, Mahamod. 2013. Recent Advances in Wireless Indoor Localization Techniques and System. *Journal of Computer Networks and Communications*, **2013**, 1–12.
- [Gartner, 2017] Gartner. 2017. *Gartner Says Worldwide Sales of Smartphones Grew 7 Percent in the Fourth Quarter of 2016*.
- [Hackernoon, 2017] Hackernoon. 2017. *How Much Time Do People Spend on Their Mobile Phones in 2017?*
- [Hashemi, 1993] Hashemi, H. 1993. The indoor radio propagation channel. *Proceedings of the IEEE*, **81**(7), 943–968.
- [Hoffman-Wellenhof, 2001] Hoffman-Wellenhof. 2001. *Global positioning system: theory and practice*. Springer-Verlag. 382p.
- [InLoco, 2018] InLoco. 2018. *A tecnologia de localização indoor mais precisa do mercado*.
- [LIPS, 2018] LIPS. 2018. *Developer-friendly indoor RTLS for real-time accurate positioning*.
- [Luo *et al.* , 2014] Luo, C., Hong, H., & Chan, M. C. 2014 (April). PiLoc: A self-calibrating participatory indoor localization system. *Pages 143–153 of: IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*.

- [Lymberopoulos *et al.* , 2015] Lymberopoulos, Dimitrios, Liu, Jie, Yang, Xue, Choudhury, Romit Roy, Handziski, Vlado, & Sen, Souvik. 2015. A realistic evaluation and comparison of indoor location technologies. *In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks - IPSN '15*. ACM Press.
- [Oxford, 2018] Oxford. 2018. *Definition of indoor in English*.
- [Shirkhorshidi *et al.* , 2015] Shirkhorshidi, Ali Seyed, Aghabozorgi, Saeed, & Wah, Teh Ying. 2015. A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data. *PLOS ONE*, **10**(12), 1–20.
- [Space, 2013] Space. 2013. *Navstar: GPS Satellite Network*.
- [Torres-Sospedra *et al.* , 2014] Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J. P., Arnau, T. J., Benedito-Bordonau, M., & Huerta, J. 2014 (Oct). UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. *Pages 261–270 of: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.
- [Wang *et al.* , 2012] Wang, He, Sen, Souvik, Elgohary, Ahmed, Farid, Moustafa, Youssef, Moustafa, & Choudhury, Romit Roy. 2012. No Need to War-drive: Unsupervised Indoor Localization. *Pages 197–210 of: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*. MobiSys '12. New York, NY, USA: ACM.
- [Wu *et al.* , 2013] Wu, Chenshu, Yang, Zheng, Liu, Yunhao, & Xi, Wei. 2013. WILL: Wireless Indoor Localization without Site Survey. *IEEE Transactions on Parallel and Distributed Systems*, **24**(4), 839–848.
- [Youssef & Agrawala, 2005] Youssef, Moustafa, & Agrawala, Ashok. 2005. *The Horus WLAN location determination system*. Ph.D. thesis, New York, NY, USA.