



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

Comparação do uso de diferentes smartphones em técnicas de aprendizagem de máquina para localização indoor.

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Larissa Lages de Oliveira (llo@cin.ufpe.br)
Orientador: Daniel Carvalho da Cunha (dcunha@cin.ufpe.br)
Área: Localização Indoor e Aprendizagem de Máquina

Recife, 17 de abril de 2017, Brasil/PE

Resumo

Técnicas de localização indoor tem sido bastante estudadas nos últimos anos. Embora o GPS possa aproximar bem a posição do usuário em ambientes outdoor, em ambientes indoor a degradação do sinal é muito grande. Nesse cenário, algoritmos de Aprendizagem de Máquina, usando valores de força do sinal de WiFi (RSSI), têm sido largamente utilizados para localização de usuários. Entretanto pesquisas mostraram que smartphones diferentes mostram valores de RSSI diferentes para um mesmo ponto. O objetivo desse trabalho então é analisar a diferença na acurácia das técnicas de localização indoor para diferentes modelos de smartphones.

Introdução

O interesse por técnicas de localização indoor cresceu bastante nos últimos anos, muitas aplicações precisam saber a localização do usuário para prover seus serviços [1]. O GPS é uma boa solução para localização outdoor mas não funciona bem em ambientes indoor por causa da atenuação que o sinal sofre devido a estrutura da construção (telhado, paredes e outros objetos). Por esse motivo outras soluções foram propostas. Uma delas é o fingerprinting. Esta técnica usa valores de RSSI (Received Signal Strength Indicator) para localizar o usuário. Ela consiste de duas fases: offline e online. Na fase offline, é criada uma base de dados com medições de valores RSSI de pontos de acesso existentes, essa base de dados também inclui coordenadas do local da medição e pode incluir outras informações relevantes como número do andar e modelo do celular[2]. Já na fase online as medições do celular do usuário são comparadas com as que estão no banco de dados, os valores mais próximos indicam a posição dele.

Uma das maiores vantagens do fingerprinting é que ele não requer que nenhum hardware adicional seja usado além da infraestrutura de rede de área local sem fio (WLAN) já existente[1]. Entretanto existem alguns problemas na técnica de fingerprinting, diferentes smartphones podem enxergar número de pontos de acesso e valores de RSSI diferentes. A maioria dos algoritmos de fingerprinting desenvolvidos não levam isso em consideração [3]. Além disso, a seleção de algoritmos e seus parâmetros para comparação na fase online pode afetar a acurácia do método. Outro problema é que artigos que desenvolvem novas técnicas para fingerprinting geralmente utilizam sua própria base de dados e assim fica difícil fazer uma comparação entre os algoritmos já existentes.

O objetivo desse trabalho então é fazer um estudo comparativo para observar como a acurácia dos algoritmos de aprendizagem de máquina utilizados para fingerprinting mudam dependendo do modelo de smartphone utilizado. Para resolver o problema das diferentes bases de dados, neste trabalho será utilizada a base UJIIndoorLoc, disponível no site de acesso público UCI Machine Learning Repository¹. Ela foi criada para ser uma base de dados referência para comparar diferentes metodologias de localização indoor[1]. Diferentes técnicas de aprendizagem de máquina serão utilizadas e os resultados comparados entre si e com outros autores que utilizaram outras técnicas para a mesma base de dados, a fim de analisar a acurácia e a complexidade computacional.

¹ <http://archive.ics.uci.edu/ml/>

Objetivos

Este trabalho tem por objetivo analisar qual modelo de celular fornece os melhores resultados na hora de usar a técnica de fingerprinting. Os chipsets dos smartphones são diferentes, logo os pontos de acesso e os níveis de sinais que eles enxergam também são diferentes e isso vai influenciar no resultado dos algoritmos de predição. O trabalho também tem como objetivo comparar os resultados obtidos pelos algoritmos de aprendizagem de máquina implementados com os de outros autores, para a mesma base de dados, a fim de analisar a acurácia, a complexidade computacional e checar se os algoritmos implementados têm uma boa generalização.

Metodologia

A base de dados utilizada será a UJIIndoorLoc, ela é uma base de dados pública e tem 21049 amostras. Os dados foram coletados por mais de 20 usuários utilizando 25 modelos diferentes de smartphones. Os dados foram coletados dentro de três construções, na universidade Jaume I, usando 4 ou 5 andares, dependendo do prédio. A base de dados contém níveis de RSSI das WAPs (Wireless Access Points), as coordenadas geográficas das amostras, o ID do prédio e o ID da sala em que foi coletada a amostra, um ID para saber se o usuário está do lado de fora ou de dentro da sala, um ID do smartphone e o timestamp [1]. Essa base de dados foi escolhida justamente por ter muitas amostras e por ter sido construída a partir de medições de diferentes celulares.

A quantidade de amostras coletadas por cada celular não foi a mesma. Por isso os celulares escolhidos para o estudo são os que têm a maior quantidade de amostras. Os algoritmos de aprendizagem de máquina utilizados serão os que são mais utilizados hoje em dia na academia para resolver problemas de Wi-Fi fingerprinting. Primeiro será feita uma comparação entre os níveis de RSSI que cada celular coletou. Depois as máquinas de aprendizagem serão treinadas com a base de dados e testadas com as medições de cada um dos celulares. Com isso podemos saber qual celular vai nos dar uma confiabilidade maior na localização e o quanto a diferença dos níveis de sinais medidos vai impactar na acurácia dos métodos utilizados. O treinamento será feito com as medições de todos os celulares que não foram selecionados para a comparação mais algumas medições dos celulares selecionados. Os conjuntos de teste serão formados pelas amostras dos smartphones selecionados que não fizeram parte do conjunto de treinamento. Cada celular terá um conjunto de testes, e pela taxa de acertos de cada um deles, podemos saber qual celular nos dá o melhor resultado para o problema de localização. Localização indoor é um problema de classificação, portanto será considerado que o algoritmo encontrou o usuário se ele acertar o ID da sala em que o usuário está.

O código será feito em python com a ajuda da biblioteca scikit-learn para implementar os algoritmos. Os parâmetros dos algoritmos serão variados a fim de se obter os melhores resultados.

Cronograma

Atividade	Período														
	Março		Abril				Maio				Junho				Julho
Revisão bibliográfica	X	X													
Estudo dos algoritmos utilizados			X	X											
Implementação				X	X	X									
Experimentos							X	X	X	X					
Avaliação dos resultados											X	X			
Escrita do TG									x	x	x	x	x	x	
Preparação da apresentação												x	x	x	x

Referências

- [1] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, T. J., J. P., Avariento, M., Benedito-Bordonau, and J., Huerta, "UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint based Indoor Localization Problems", 5th International Conference on Indoor Positioning and Indoor Navigation, 2014.
- [2] S. Bozkurt, G. Elibol, S. Gunal, and U. Yayan, "A comparative study on machine learning algorithms for indoor positioning," in Innovations in Intelligent Systems and Applications (INISTA), 2015 International Symposium on, Sept 2015, pp.1–8.
- [3] G. Lui, T. Gallagher, B. Li, A. G. Dempster, and C. Rizos, "Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization," international Conference, 2011.

Possíveis Avaliadores

Prof. Um: George Darmiton da Cunha Cavalcanti

Prof. Dois: Cleber Zanchettin

Assinaturas

Recife, ___ de _____ de _____

Larissa Lages de Oliveira

(Aluno)

Daniel Carvalho da Cunha

(Orientador)