TELECOM PARIS

Wireless sensor network simulation and localization

Author:
Yukun LIU

Email:
yukun.liu@telecom-paris.fr

Supervisor:
Pascal Bianchi
Ons Jelassi Ben Atallah

pascal.bianchi@telecom-paris.fr
ons.jelassi@telecom-paris.fr

Acknowledgements

I consider myself as a very lucky person to find and complete this internship, and I would like to express my deep appreciation to the wonderful people who led me through the internship period.

I thank Mme. Marie-Hélène Piovano, Professional Project and Internships Manager of Télécom Paris, for helping me out in internship searching.

I would like to express my very great appreciation to Dr. Pascal Bianchi for generously offering this opportunity to carry out the work. I am also grateful to him and Dr. Ons Jelassi Ben Atallah for their precious guidance during the planning and development of this research work which were extremely valuable for my study both theoretically and practically.

I also express deep thanks to the PhD students Arturp, Marc in the same office, for passing normal daily moments together.

I perceive this opportunity as a big milestone in my career development, as a beginning of my research journey. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement.

Abstract

Localization is one of the fundemental problems in wireless sensor networks (WSNs), and it enables numerous applications. Localizing based on channel state information (CSI) becomes a research hotspot because it can provide fine-grained information. Besides, simulation is accompanied with localization techniques to validate the algorithm and save cost.

This project is proceeded under the supervision of Professor Pascal Bianchi and Professor Ons Jellais. It consists of two independent parts. The first part of the project is to deploy a simulator for WSNs in outdoor environment. To achieve this, the communication module in MATLAB is chosen. Based on it, a simulator allows multipath modelling is built and simulations are run in several cities with chosen transmitters and receivers. Meanwhile, the fingerprinting databases storing encoded features of channel state measurements are created.

The second part is to explore fingerprint-based localization methods with generated database. Ensemble methods are chosen to map fingerprints to position. Meanwhile, the impact of grid size of transmitters is studied. Then a architecture for localizing using fingerprint-based method on different maps is proposed. A series of experiments have been conducted to illustrate its performance in various conditions. The effectiveness of the proposed method is valideted by comparing with state-of-art localization techniques.

Keywords: Localization, channel state information, wireless sensor networks

Contents

1 Introduction

Under the rapid development of Internet-of-things (IoT), location is becoming one important feature in the IoT field and localization brings high value service [1]. Originally, the key drivers to localize an agent in a network have been governmental force for sake of emergency. The official support for emergency service has motivated the standardization of localization and the development of localization strategies [2]. Moreover, the exploitation of the location information has also drawn attention from operators due to its potentiality for commercialization. On the user side, the location of a device can generate massive civil applications, so-called location-based services (LBS), such as navigation, geo-marketing, social networking and advertising, etc. On the network side, the location information can also help in network optimization, namely location-aware communication, which can improve network efficiency and capacity, such as intelligent transportation systems (ITS), self-organizing networks (SON), etc.

For indoor localization, there are abundant medias and techniques for positioning as shown in Table. 1 according to [3].

Media	Examples	Pros	Cons
Mechanical	IMU	high precision, high energe efficiency	invasive
acoustic	sonar	cheap, low power consuming	short range, synchronization
Optical	Infrared, camera	cheap, resistance to multipath	LOS only, short range, privacy
radio frequency	WiFi	infrastracture support	Los if using signal strength

Table 1: Overview of indoor localization media

Localising a device in an outdoor environment typically utilises the Global Positioning System (GPS). This system is the standard for outdoor localisation and provides metre-level accuracy during normal operations. However, it provides high positioning accuracy at the expense of additional hardware and storage space, while it can quickly deplete the battery on the device, and the location information is not necessarily shared with network operator. For networks with low energy cost sensors, e.g. Low-Power Wide Area Network (LPWAN), GPS will be no longer available. While LPWAN enjoys the benefits that transmitting low frequency signal of large wavelength allows for signals to pass through walls and obstacles, and thus signals can reach further. This project focuses on localizing under LPWAN schema with tranmisson signal at a frequency about 1GHz. It developed a pipeline from WSN simulation to outdoor localization, and proposed a novel method for localizing with transfer scenarios. There is no physical experiment involved in this work and the localization methods are tested based on data generated from simulation.

2 Simulation

The use of Wireless Sensor Networks involves conception, designing and test phase [4]. While for the test phase, building a testbed and running experiments on it are costly, time consuming, and difficult on technical aspect. Moreover, since real experiments involve many uncertainties and factors affecting the result, experimental repeatability is largely compromised [5]. Thus, a simulator is essential for WSNs development and application that it allows to tune configurable parameters and reproduce experimental results. Gaining insights and drawing conclusions from a simulation study is not a trivial task. Nowadays, with the huge variety of available simulators, it is important to identify which simulator suits the most for a particular scenario.

The radio frequency (RF) propagation module in MATLAB describes the behavior of electromagnetic radiation from a point of transmission as it travels through the surrounding environment based on path loss model [6]. It is capable to calculate the coverage, the strength of signal and has great flexibility in setting up the environment and builds the site on open street map. Moreover, it provides functions for channel state information, with operational frequency from 100 MHz to 100 GHz, which fits the need ($\sim$ 1 GHz). Therefore, it is chosen as our ideal simulator tool and all the latter analysis will be drawn from it.

2.1 Simulator settings

To build a complete simulator, three components are required to be set up: transmitters, receivers and a propagation model. In our study, we focus on a simple uplink process, without concerning protocol and encoding/decoding process. Thus transmitters and receivers correspond to sensors and base stations respectively. They can be built up with location (latitude, longitude, elevation) and transmitting frequency. The propagation model describes the transmitting channel, and predicts the propagation and attenuation of radio signals as the signals travel through the environment. After the three components are established, the simulator can load open map information and execute the simulation with the support of the map.

In our implementation, the key characteristics of sites are listed in Table. 2. For simplicity, transmitters and receivers are synchronized. Every transmitter and receiver contains one antenna only (Single-input single-output). Besides, the antenna is assumed to be isotropic that the transmitter radiates uniformly in all directions and the receiver captures signals in all directions. The geolocation and height of sites depend on the spatial scenarios. Some specific cases we used in simulation are presented in Section. 2.2.

Property	Parameter
Antenna type	Isotropic
Antenna size	1
Transmitter power*	10 W
Transmitter frequency*	1 GHz
Receiver sensitivity#	-100 dBm

* - transmitter property only, # - receiver property only

Table 2: Configuration of sites

For propagation, a shooting-bouncing-rays (SBR) method with ray tracing model is chosen. The SBR method launches many rays from transmitters, and the reflection between rays and surroundings are modeled (other interaction types like diffraction, refraction, scattering are ignored). It models both line-of-sight (LOS) and none-line-of-sight (NLOS) conditions, and supports calculation of propagation paths for up to ten path reflections.

Property	Parameter
Ray tracing method	shooting-bouncing-rays
Max number of reflections	5
Building material	concrete
Surface material	glass

Table 3: Configuration of propagation model

2.2 Spatial scenarios

Spatial scenarios refer to a schema of structure arranging transmitters and receivers in simulation. In general, a square map frame is chosen and the frame is divided into meshes. One example is made in Figure 2.1. The planar distance is 50m between adjacent transmitters and 200m between adjacent receivers. The height of a transmitter/receiver antenna is set as the geodesic elevation with an offset (1.5m/30m for transmitter/receiver) in our simulation.

The transmitters/receivers sit separately on mesh nodes and transmitters emit signals to all reachable receivers. Subsequently features or measurements will be derived at each of the receiver for further processing.

Figure 2.1: An example of spatial scenarios: a 20×20 transmitter mesh in red and a 5×5 receiver mesh in blue are evenly distributed in a 1km×1km square in London city.

2.3 Feature representation

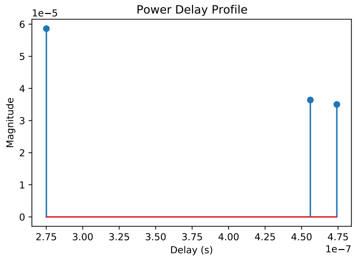
The ultimate objective of the project is to localize a transmitter with the assistance of the network composed by receivers. It is also named network-based localization, determining the device location by using signal measurements performed by the network with respect to the device. The classic measurements such as Time of Arrival (ToA), Received Signal Strength Indicator (RSSI), suffer from significant performance degragation in complex scenarios as being susceptible to multipath effect. Differently, the PHY layer feature, channel state information such as channel impulse response (CIR) can discriminate multipath effect and achieve pervasive localization [7]. Considering our simulation in urban environment, CIR will be an appropriate measurement.

One way to represent the CIR of a multipath channel is by discrete number of impulses shown in Eq. 2.1

$$h(t,\tau) = \sum_{i=1}^{N} c_i(t)\delta(\tau - \tau_i) \tag{2.1}$$

where $h(t)$ is the CIR representation, $c_i(t)$ are the complex attenuation coefficients varying with time, and $\delta(\cdot)$ denotes the dirac function. There are N arrival paths and τ_i is delay corresponding to each path. In implementation, each impulse describes a trace from a transmitter to a receiver, and N is the total number of traces. An example is illustrated in Figure 2.2.

(a) Ray tracing scenario



(b) Channel profile

Figure 2.2: An example of CIR generation: (a) illustration of signal propagation between sites. Every connected line indicates one ray trace. The site in red color is a transmitter with geolocation (51.5134, -0.0901); The blue site is a receiver located in (51.5132, -0.0909); (b) shows the impulse profile where each impulse corresponds to one trace in (a).

CIR measurements bring rich location information, and also depend largely on environments. It is probable that a receiver captures no signal from a transmitter where the surroundings block the signal. Furthermore, sites in distinct location with diverse surrounding complexity will reflect different number of ray traces and thus different impulses. All these factors may concern data standardization and localization.

3 Localization

3.1 An overview for outdoor localization techniques

Localization is a process to determine the coordinates of a device in a system, which is one of most important subject under the wide diffusion of WSN, and has a variety of applications. There is a rich literature on localization strategies. This study concerns network with low-power sensors in wide area, where radio frequency (RF) is one of the conventional media for positioning [8]. A taxonomy of RF-based localization system is shown in Figure 3.2. Depending on fundamental principles, the RF-based localization systems can be categorized into range-based and range-free type [2]. Range-based methods base on geometric mapping techniques such as trilateration, triangulation for range estimation and target localization. The range estimation normally relies on measurements such as RSSI, ToA, TDoA from different terminals. It requires at least 3 devices to determine one sensor's position and is vulnerable to multipath effect.

Unlike Range-based algorithms, range-free methods do not require absolute distance estimation. It operates on connectivity information between nodes, They are basically implemented from fingerprinting or proximity techniques.

In fingerprint approach, each available node is represented by a unique channel fingerprint extracted from preprocessing of channel measurements [8]. The pre-collected fingerprints are stored in a database. Newly acquired fingerprints are compared against the database and applied for localization based on positioning algorithm as shown in Figure 3.1.

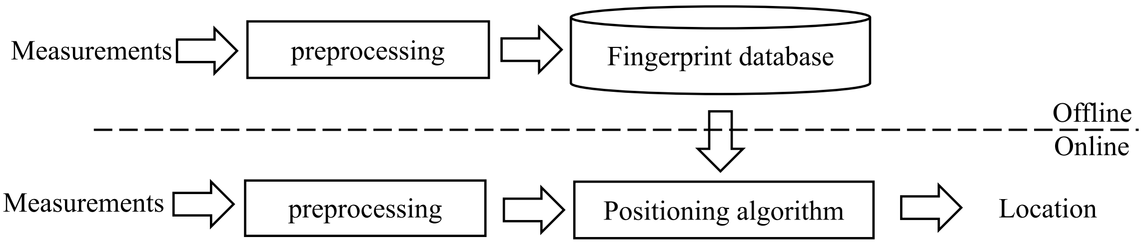


Figure 3.1: Fingerprint approach schema

For proximity technique, it applies radio communication range to establish the nodes that are proximal. It bases on the idea that the distance of nodes is within the transmission range once they can communicate [9] . Some foremost proximity based techniques such as Centroid Localization Algorithm (CLA) [10], Approximate Point In triangulation (APIT) [11], DV-Hop algorithm [12], depends on reference nodes or anchors for localization, attempt to select the anchors with the most significant characteristics for location estimation. Since it requires no extra hardware for signal transmission, it is cost-effective but also less accurate.

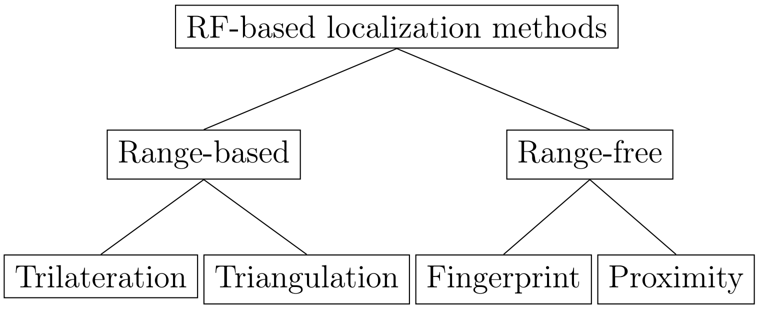


Figure 3.2: A simplified taxonomy of RF-based localization methods

3.2 Fingerprinting-based outdoor localization

Since the concerning scenario is outdoor wide area with low power sensors, a fingerprinting-based localization method is employed, which exhibits multipath effect resistance and can achieve high positioning resolution. The fingerprints database is generated from CIR measurements as stated in Section. 2.3.

3.2.1 Preprocessing

A CIR measurement sample is a complex 2D array, composed by time and amplitude features. It consists of multiple impulses, and each impulse marks a transmission trace. The fingerprinting database will be represented as a cell, where each entry in the cell corresponds to a CIR measurement between a transmitter and a receiver, and its label will be the geo-location of devices to locate:

$$CIR\ cell = X = \begin{bmatrix} H_{1,1}, & H_{1,2}, & \cdots & H_{1,R} \\ H_{2,1}, & H_{2,2}, & \cdots & H_{2,R} \\ \vdots & \vdots & \cdots & \vdots \\ H_{T,1}, & H_{T,2}, & \cdots & H_{T,R} \end{bmatrix}, H_{i,j} = \begin{bmatrix} t_0 & \cdots & t_{n_{i,j}} \\ a_0 & \cdots & a_{n_{i,j}} \end{bmatrix} \tag{3.1}$$

where T, R denote the number of Tx and Rx, $H_{i,j}$ indicates the CIR measurements from transmitter i to receiver j, t and a corresponds to time and amplitude in the CIR and $n_{i,j}$ is the number of impulses in $H_{i,j}$.

Due to a variety of environment complexity, the shapes of $H_{i,j}$ may vary. To standardize the form of data, we choose to take the magnitude of each measurement

$\|H_{i,j}\|$ and pad 0 to the same size M, as the max number of traces in each sample. The full preprocessing schema is illustrated in Figure 3.3.

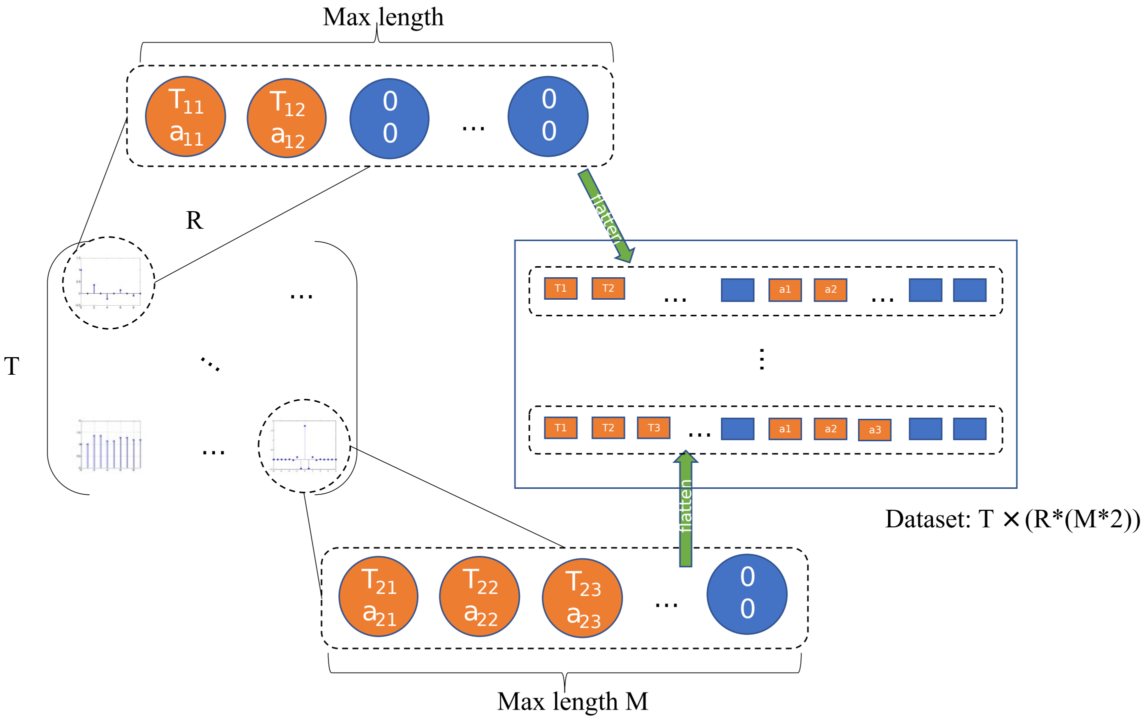


Figure 3.3: Preprocess CIR measurements: each entry is padded with 0 to the same length and flattened

For training purpose, the measurements are flattened as vectors. Consequently, the fingerprinting dataset will be of shape $T \times (R * M * 2)$. The labels of fingerprints are the set to be the geo-locations of devices to locate:

$$Y = \begin{bmatrix} lat_1, & lon_1 \\ lat_2, & lon_2 \\ \vdots, & \vdots \\ lat_T, & lon_T \end{bmatrix} \tag{3.2}$$

where (lat_i, lon_i) notes the latitude and longitude of Tx_i.

Following the schema, one dataset instance created by simulation in London city is shown in Figure 3.4. It shows the spectrum of CIR values, and it is evidently sparse since each entry are padded with zeros for standardization, which results in unnecessary training load. There are still room left for feature fusion.

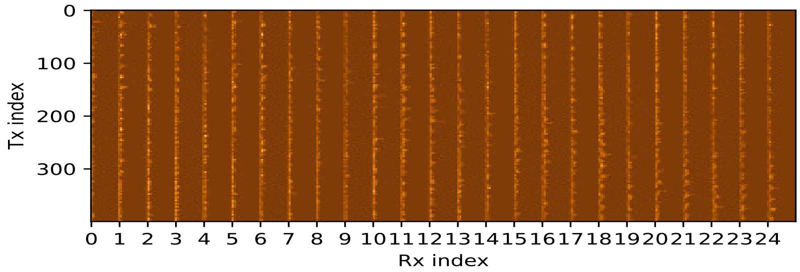


Figure 3.4: An overview of fingerprinting database for London city: 400 Tx and 25 Rx during simulation with 5 max reflections

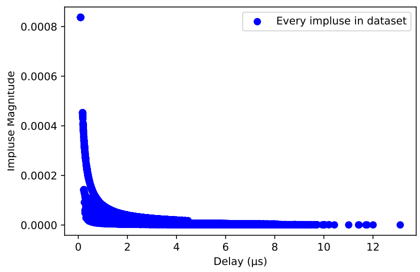
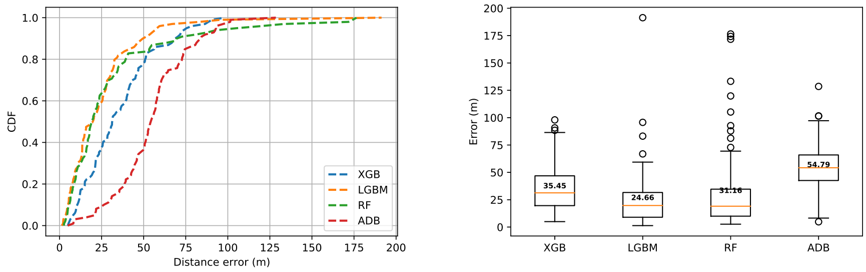


Figure 3.5: Overview of impulse responses: 9239 samples simulated between 400 Tx and 25 Rx in London city

Besides, it is verified the magnitude is highly correlated with time delay, which is logic. The lower delay results in lower attenuation. It may provide a possibility to fuse time and amplitude features into one feature. But the risk is that the time delay is not correlated with the magnitude exactly. Fusing the two features will inevitably loss information and may reduce the accuracy of the localization algorithm.

3.2.2 Experiments

On the technical aspect, localization is to find a function able to map measurements to location space. To do this, four ensemble learning methods are chosen for non-linear mapping between the measurements and locations: Random Forest (RF), ExtremeGradientBoost (XGB), LightGradientBoostMachine (LGBM), AdaptiveBoost (ADB). The testbed for data simulation is chosen in a 1km$\times$1km square area, latitude from 51.5108 to 51.5198, longitude from -0.0988 to -0.0844, located



(a) CDF of localization errors (b) Boxplot of localization errors

Figure 3.6: Performance evaluation for ensemble methods

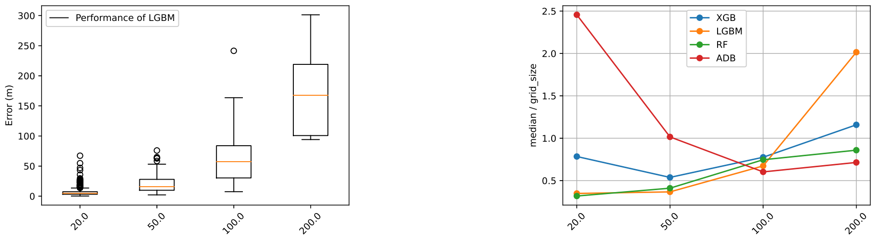
in London. The Tx and Rx are uniformly distributed in a 20×20 and 5×5 mesh in fixed positions as shown in Figure 2.1, with maximum reflection of 5. The distance between tx mesh, namely grid size, is 50m, which is a criteria to evaluate the algorithms. In addition, the evaluation metric is defined as the distance between prediction and label.

The performance of each method after parameter tuning is illustrated in Figure 3.6. It shows LGBM method generally performs better that it has lowest mean distance error. While its variance is large that it can severely under-positioning. On the other hand, XGB performs slightly worse in general, but it is relatively robust that its regression error is constrained in a small region. The performance of other two methods wander between LGBM and XGB. AdaBoost least performs in the localization task that its mean distance error is 54.79m, greater than the grid size (50m).

Besides the learning method, the impact of grid size is studied. Within the same 1km×1km area and rx setup, the tx mesh is controlled from grid size of 200m to more fine grained: 100m, 50m, 20m. The experiment result is listed in Figure 3.8. It is logic the positioning error increases as the grid gets coarse. What interests more is the trend that the relative error goes with grid granularity.

3.3 Localization with transferred scenarios

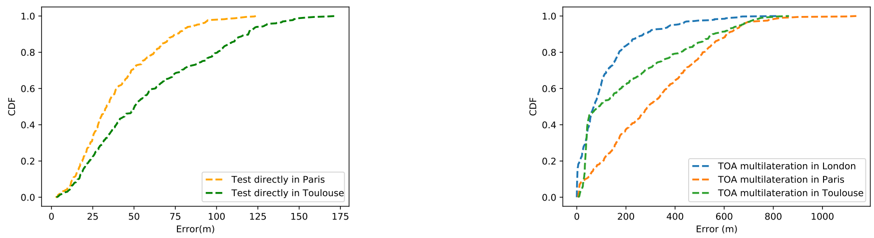
The formal experiments focuses on training model to determine the device position homologous as the database, e.g. training with fingerprinting database from London and test with measurements in London. However, when the test comes across new data, the trained model is applied to different scenarios, the precision may dramatically drop [13]. Especially in our case, if the model trained with data from London is used to test in Paris or Toulouse, its performance can be largely



(a) LGBM with varying grid size (b) Mean/Grid size error w.r.t grid size

Figure 3.7: Performance evaluation w.r.t grid size

reduced as shown in Figure 3.8a. Nevertheless, it is known range-based methods rely on geometric mapping instead of training model with data. It is resistant to data variations and can be naturally used for transferring scenarios. Therefore, one of range-based methods - Time of Arrival based multilateration is implemented for comparison as well. This method provides an algebraic solution based on ToA, its performance is shown in Figure 3.8b.



(a) Position estimation with LGBM trained on London city (b) Time-of-Arrival multilateration positioning

Figure 3.8: Benchmark methods for proposed scenarios

The performance for this range-based method is not decent but it is expected. Range-based methods can largely be affected by multipath effects which happens frequently for transmission in urban environment. Neither of these methods present a convincing result facing transferred scenarios. While generalizing a model is imperative. In practice only small portion of data can be collected and trained, a model inevitably faces challenges of coming data. It is needed to enhance the transferability of the model to process the heterologous signal. For example a model trained by fingerprinting database from London, processes the data from Paris. Apart from captured features, some prior information about heterologous signal can be obtained in advance, such as the street map, the position of the station etc. These information can also be merged in the model to help improve model's generalization performance.

To tackle this issue, a novel framework is proposed.

3.3.1 Transferring framework

Transferring is to train a model with source data and employ it in target data. Classical transfer learning method demands massive data for training and adapting which is not available in this project. Since different distributions result in severely reduced performance, we decide to create a model to learn some features that possess similar distributions no matter where the data is from. We decide to make the model to learn the distance between transmitters and receivers firstly, because the distance between them should be within a certain range. Thus, a novel framework for resolving heterlogous data is proposed, combining two phases:

Point-to-point distance estimation: Finding a mapping f that maps CIR feature of i-th transmitter $H_i = [H_{i,1}, H_{i,2}, \cdots, H_{i,R}]$ to distances between the transmitter and all the receivers $D_i = [D_{i,1}, D_{i,2}, \cdots, D_{i,R}] \in \mathcal{R}^R$

$$f : H_i \to D_i$$

In the implementation, the mapping f for distance estimation is replaced by ensemble methods for regression.

Point searching: Finding a function that searches i-th transmitter position whose distance with all Rx matches most as the estimated distances D_i, with knowing Rx positions P.

$$\begin{aligned} &\arg\min_{z \in \mathcal{R}^2} \sum_{i=1}^{R} (d(z, P_i) - D_i)^2 \\ s.t. \quad & z(0) \in (-90, 90),\ z(1) \in (0, 180) \end{aligned}$$

where $P \in \mathcal{R}^{2\times R}$, is a 2D array for Rx positions, $d(\cdot,\cdot)$ is a function measuring euclidean distance based on two geolocations, and z is the coordination of the searching point, consisting of latitude and longitude, which is limited within $[-90, 90]$ and $[0, 180]$. This is a deterministic method and requires no prior training.

The overall architecture of transferring a model to a known scenario is illustrated in Figure 3.9. It includes two parts: offline training and online testing. Firstly a source database is used to train a model for distance estimation in advance. Then the model is directly taken to estimate transmitter-wise distance, and then employ an optimizer, with addition of receivers' location information to find the matching position.

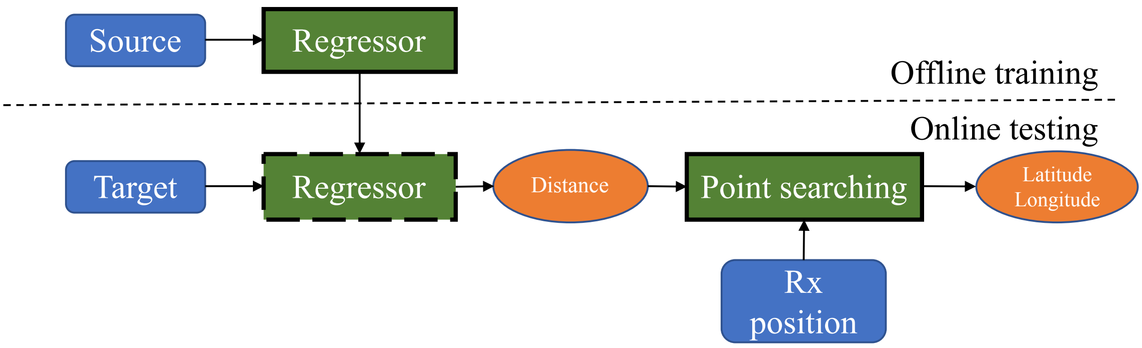


Figure 3.9: Transferring structure

3.3.2 Experiments

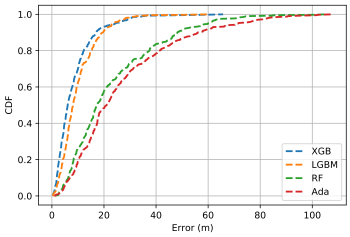
For experiments, London is chosen the data source testbed, and Paris and Toulouse are transferred scenarios, specifications are listed in Table 4. The setup of sites are the same: 20×20 Tx mesh and 5×5 Rx mesh uniformly dispersed in a 1km×1km square area with 50m grid size. There are 400 samples for each fingerprinting database.

City	Latitide range	Longitude range	Networks
London	51.5108, 51.5198	-0.0988, -0.08440	Transmitter mesh: 20×20
Paris	48.8296, 48.8386	2.3132, 2.3268	Receiver mesh: 5×5
Toulouse	43.6048, 43.6138	1.4393, 1.4517	Grid size: 50m

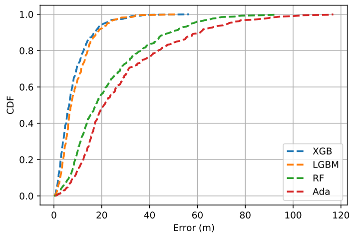
Table 4: WSN configurations in experiments

To explore the proposed framework, the effect of regressor is tested firstly. Four ensemble methods are selected with positioning tasks in Paris and Toulouse. Following the workflow in Figure 3.9, the regressor is feed by London fingerprinting database, then tested with target data and followed by a optimizer to determine the position. The result is presented in Figure 3.10. XGB outperforms on other methods. Considering the grid size of 50m, the results of XGB, LGBM and AdaBoost are acceptable.

The performance of the proposed method outperforms estimating position directly with ensemble methods, not to mention the range-based method. A quantified result is shown in Table 5.



(a) Trained by London database, positioning with Paris database



(b) Trained by London database, positioning with Toulouse database

Figure 3.10: Positioning with different scenarios and regressors

Mean error (m)		Methods		
		Direct positioning with LGBM	ToA multilateration	**Proposed method**
City	Paris	39.8	315.6	10.6
	Toulouse	59.5	205.9	13.8

Table 5: Performance comparison

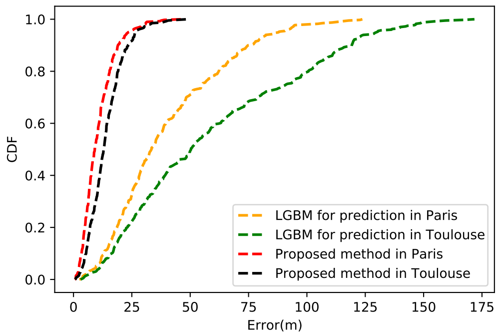
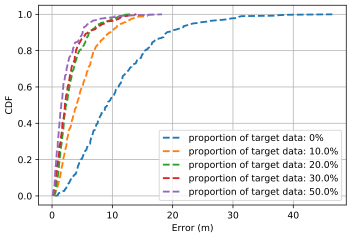


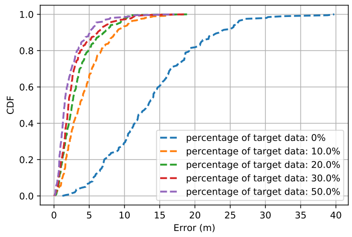
Figure 3.11: Performance of positioning with different methods in Paris and Toulouse: orange & green curve: predicting position with LGBM model trained with data from London, tested on Paris and Toulouse; red & black curve: predicting position with proposed framework, trained from London, tested on Paris and Toulouse

3.3.3 Transferring scenarios with supplemented data

Besides, the impact of supplemented target data is explored. Different proportion of target data is to source fingerprinting database in training in order to augment the database. The result is shown below with LGBM as the regressor, under the condition the supplemented data are of same weight as the source data.



(a) Trained by London data and supplemented Paris data, positioning with non-trained Paris data



(b) Trained by London data and supplemented Toulouse data, positioning with non-trained Toulouse data

Figure 3.12: Positioning with proportional supplemented target data

From Figure 3.12, it is interesting to witness the evident improvement with supplemented target data. The more target data are supplied, the more accurate the positioning will be. Following this experiment, another test is conducted that more weights are assigned to supplemented data in Training stage. While it turns out the performance has only tiny changed (See Appendix 4 for more details).

Lastly, the impact of heterogeneity of data is analyzed. It is vital to verify if heterogenous data will moderate the performance. 75% of data from London (Source) is taken for training the regressor (set as LGBM), and the other 25% is left as target data for testing and comparing with database from Paris and Toulouse. Although it should be noted the test size is unbalanced among the three groups. There are 100 samples for London test, and 400 for Paris and Toulouse. The result is demonstrated in Figure 3.13. A quantitative summation is introduced in Tabel 6. It turns out the model performs better on its homologous data indeed. Overall the localization error is well controlled considering a granularity of 50m. It is also interesting to find the model predicts better in Paris than Toulouse., which may be due to the fact that Paris has a more similar map condition with London than Toulouse. Moreover, comparing the result in Table 6 and Table 5, it is found the latter performance is slightly better than the former. The reason is 100% London data was consumed for training in the latter case while 75% for the former.

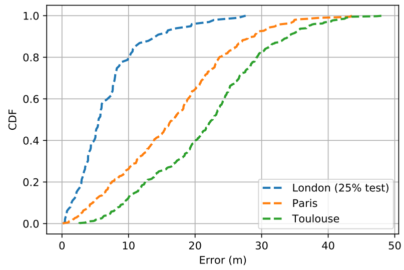


Figure 3.13: Positioning for different cities

City	Mean Error (m)	Maximum Error (m)	$p(\|e\| \leq 20m)$
London	6.8	27.3	95.1%
Paris	16.6	43.6	64.0%
Toulouse	21.9	48.2	40.6%

Table 6: Localization performance with proposed framework: 75% of data from London for training, tested on Paris, Toulouse and 25% of London data

4 Conclusion

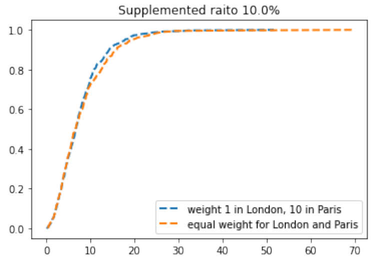
In this project, a simulator for outdoor WSNs is built based on Matlab communication module. Then several fingerprinting databases are created from simulations while channel impulse response are chosen as the measurements to take in. Then we explored about range-based and range-free positioning methods in the localization part. Due to the resistance for multipath effect, fingerprinting techniques are chosen as positioning techniques. Then ensemble learning methods are adpated to train the fingerprint database. Besides, the relation between localization precision and grid size is also explored. Furthermore, considering the challenge of data coming from different resources, a novel localization scheme is proposed to deal with transferring scenarios. Specifically, a regressor is trained offline for distance estimation, and an optimal point is searched by optimization. Meanwhile, a ToA multilateration localization method is implemented as baseline. The experimental results demonstrate that the proposed schema outperforms the baseline and can achieve satisfactory localization precision.

Nevertheless, the work is far from finished. Here are some notes about the limitation of proposed schema and possible future improvements:

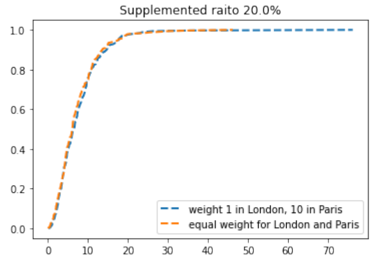
- Feature engineering: Preprocessing could be done to extract more time, amplitude, phase, angle information from CIR
- Simulation data: Current analyses are conducted based on simulation data, which is too naive and ideal. It worths considering to add noise to the measurements to make simulation more realistic.
- Fusing models: Range-based and range-free methods all have their own advantages. There may be a way to fuse these two methods which can merge their merits.

Appendix A

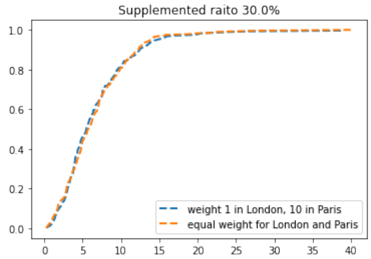
Performance of proposed method, Trained by London data and supplemented Paris data, weight ratio=1:10, testing on Paris data



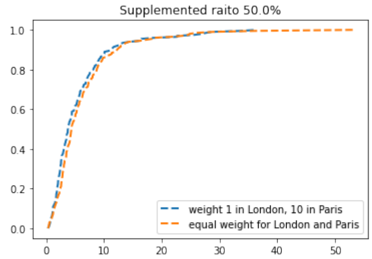
(a) Trained with supplemented data, supplemented ration: 10%



(b) Trained with supplemented data, supplemented ration: 20%



(c) Trained with supplemented data, supplemented ration: 30%



(d) Trained with supplemented data, supplemented ration: 50%

References

[1] Quoc Duy Vo and Pradipta De. A survey of fingerprint-based outdoor localization. *IEEE Communications Surveys Tutorials*, 18(1):491–506, 2016.

[2] José A. del Peral-Rosado, Ronald Raulefs, José A. López-Salcedo, and Gonzalo Seco-Granados. Survey of cellular mobile radio localization methods: From 1g to 5g. *IEEE Communications Surveys Tutorials*, 20(2):1124–1148, 2018.

[3] George Oguntala, Raed Abd-Alhameed, Stephen Jones, James Noras, Mohammad Patwary, and Jonathan Rodriguez. Indoor location identification technologies for real-time iot-based applications: An inclusive survey. *Computer Science Review*, 30:55–79, 2018.

[4] Michel Bakni, Luis Manuel Moreno Chacón, Yudith Cardinale, Guillaume Terrasson, and Octavian Curea. WSN simulators evaluation: an approach focusing on energy awareness. *CoRR*, abs/2002.06246, 2020.

[5] Fei Yu. A survey of wireless sensor network simulation tools. 2011.

[6] Inc. The MathWorks. *Communication Toolbox*. Natick, Massachusetts, United State, 2022.

[7] Zheng Yang, Zimu Zhou, and Yunhao Liu. From rssi to csi: Indoor localization via channel response. *ACM Comput. Surv.*, 46(2), dec 2013.

[8] Josyl Mariela Rocamora, Ivan Wang-Hei Ho, Wan-Mai Mak, and Alan Pak-Tao Lau. Survey of csi fingerprinting-based indoor positioning and mobility tracking systems. *IET Signal Processing*, 14(7):407–419, 2020.

[9] Adeniran Ademuwagun and Verdicchio Fabio. Reach centroid localization algorithm. *Wireless Sensor Network*, 09:87–101, 01 2017.

[10] N. Bulusu, J. Heidemann, and D. Estrin. Gps-less low-cost outdoor localization for very small devices. *IEEE Personal Communications*, 7(5):28–34, 2000.

[11] Tian He, Chengdu Huang, Brian M. Blum, John A. Stankovic, and Tarek Abdelzaher. Range-free localization schemes for large scale sensor networks. MobiCom '03, page 81–95, New York, NY, USA, 2003. Association for Computing Machinery.

[12] Nath-B. Niculescu, D. Dv based positioning in ad hoc networks. *Telecommunication Systems*, 22, 2003.

[13] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer Learning*. Cambridge University Press, 2020.