City-Scale Localization with Telco Big Data [1]

EPL 646: Advanced Topics in Databases

Christodoulou Eirini, Loizou Marios

Department of Computer Science
University of Cyprus

30 November 2017

Overview

Overview

1. Introduction
    - MR-based positioning systems:
    - Global Positioning System (GPS)
    - GPS vs MR-based positioning systems
2. Paper Contributions
3. CCR Localization Models
4. Datasets and Measurements
5. Localization Performance

Introduction

MR-based positioning systems:

Accurate localization of mobile devices (MDs), using measurement report (MR) data

- MR data:
  - measurement parameters of radio signal strengths (RSSI) when MDs connect to base stations (BSs) in telecommunication (telco) networks
  - MRs are generated when MDs are making/receiving calls or use mobile broadband services (MBB)

MR-based positioning systems:

Accurate localization of mobile devices (MDs), using measurement report (MR) data

- MR data:
  - measurement parameters of radio signal strengths (RSSI) when MDs connect to base stations (BSs) in telecommunication (telco) networks
  - MRs are generated when MDs are making/receiving calls or use mobile broadband services (MBB)

Introduction

MR-based positioning systems:

Accurate localization of mobile devices (MDs), using measurement report (MR) data

- MR data:
  - measurement parameters of radio signal strengths (RSSI) when MDs connect to base stations (BSs) in telecommunication (telco) networks
  - MRs are generated when MDs are making/receiving calls or use mobile broadband services (MBB)

Introduction

Telco Big Data opportunity

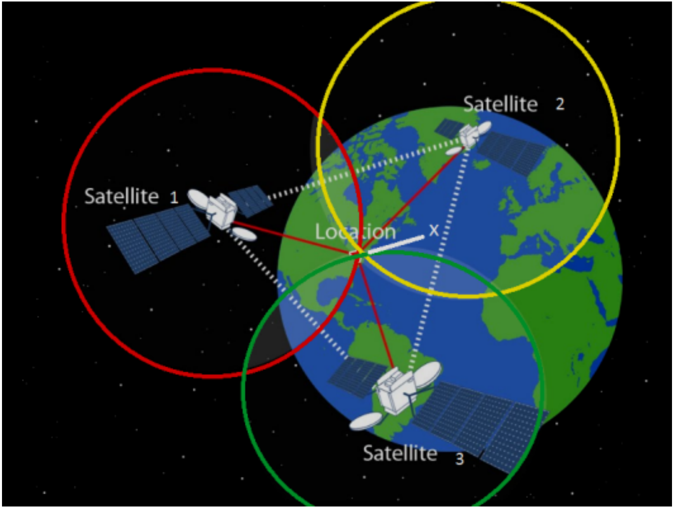
- Accumulation of many over-the-top (OTT) global positioning system (GPS) data in telco networks, from widely-used location based services (LBS)
- Can be used automatically as training labels for learning accurate MR-based positioning systems

Introduction

Telco Big Data opportunity

- Accumulation of many over-the-top (OTT) global positioning system (GPS) data in telco networks, from widely-used location based services (LBS)
- Can be used automatically as training labels for learning accurate MR-based positioning systems

Global Positioning System (GPS)



https://thesciencegeek.org/2017/01/29/gps/

Global Positioning System (GPS)

- A GPS device uses data from satellites to locate a specific point on Earth
- Using *trilateration*, a GPS receiver measures the distances to satellites using radio signals
- Trilateration is the process of determining absolute or relative locations of points by measurement of distances, using the geometry of circles, spheres or triangles

https://www.nationalgeographic.org/topics/gps/
https://en.wikipedia.org/wiki/Trilateration

Global Positioning System (GPS)

- A GPS device uses data from satellites to locate a specific point on Earth
- Using *trilateration*, a GPS receiver measures the distances to satellites using radio signals
- Trilateration is the process of determining absolute or relative locations of points by measurement of distances, using the geometry of circles, spheres or triangles

https://www.nationalgeographic.org/topics/gps/
https://en.wikipedia.org/wiki/Trilateration

Global Positioning System (GPS)

- A GPS device uses data from satellites to locate a specific point on Earth
- Using *trilateration*, a GPS receiver measures the distances to satellites using radio signals
- Trilateration is the process of determining absolute or relative locations of points by measurement of distances, using the geometry of circles, spheres or triangles

https://www.nationalgeographic.org/topics/gps/
https://en.wikipedia.org/wiki/Trilateration

GPS Pros & Cons

- ✓ Accurate localization of an MD outdoor with around $\pm 10m$ errors
- × Energy consuming
- × Unavailable in many MDs
- × Requires line-of-sight (LOS) to the satellites (degrades quickly indoor or underground
- × Being often turned off for private reasons

MR-based positioning systems Pros & Cons

- ✓ Energy efficient
- ✓ Available on most MDs
- ✓ Better network coverage and workable indoors and underground
- ✓ Active when making calls or MBB services
- × Localization error $50 \sim 1000m$ in different situations $\longrightarrow$ promise: next-generation telco networks (e.g. 5G) will have denser structures

Overview

1. Empirical Measurement of Big MR data

City-Scale Localization with Telco Big Data

* Analysis over four properties of Big MR data
  1. stability: Temporary change of MR data for the same location and MD
  2. sensitivity: Influence of small spatial changes on MR data
  3. uncertainty: MDs usually do not connect with their closest BS
  4. missing values: The amount of BSs and their corresponding RSSI value an MR contains
* This process provides practical guidance
  - how to update the localization models
  - determine the lower limit of localization errors

1. Empirical Measurement of Big MR data

City-Scale Localization with Telco Big Data

* Analysis over four properties of Big MR data
  1. stability: Temporary change of MR data for the same location and MD
  2. sensitivity: Influence of small spatial changes on MR data
  3. uncertainty: MDs usually do not connect with their closest BS
  4. missing values: The amount of BSs and their corresponding RSSI value an MR contains
* This process provides practical guidance
  - how to update the localization models
  - determine the lower limit of localization errors

1. Empirical Measurement of Big MR data

City-Scale Localization with Telco Big Data

- Analysis over four properties of Big MR data
  1. stability: Temporary change of MR data for the same location and MD
  2. sensitivity: Influence of small spatial changes on MR data
  3. uncertainty: MDs usually do not connect with their closest BS
  4. missing values: The amount of BSs and their corresponding RSSI value an MR contains
- This process provides practical guidance
  - how to update the localization models
  - determine the lower limit of localization errors

1. Empirical Measurement of Big MR data

* Analysis over four properties of Big MR data
  1. stability: Temporary change of MR data for the same location and MD
  2. sensitivity: Influence of small spatial changes on MR data
  3. uncertainty: MDs usually do not connect with their closest BS
  4. missing values: The amount of BSs and their corresponding RSSI value an MR contains
* This process provides practical guidance
  - how to update the localization models
  - determine the lower limit of localization errors

1. Empirical Measurement of Big MR data

* Analysis over four properties of Big MR data
  1. stability: Temporary change of MR data for the same location and MD
  2. sensitivity: Influence of small spatial changes on MR data
  3. uncertainty: MDs usually do not connect with their closest BS
  4. missing values: The amount of BSs and their corresponding RSSI value an MR contains

* This process provides practical guidance
  - how to update the localization models
  - determine the lower limit of localization errors

1. Empirical Measurement of Big MR data

* Analysis over four properties of Big MR data
  1. stability: Temporary change of MR data for the same location and MD
  2. sensitivity: Influence of small spatial changes on MR data
  3. uncertainty: MDs usually do not connect with their closest BS
  4. missing values: The amount of BSs and their corresponding RSSI value an MR contains
* This process provides practical guidance
  – how to update the localization models
  – determine the lower limit of localization errors

2. Context-aware coarse-to-fine regression (CCR) model

A machine learning-based positioning system from telco big data

1. Map-matching and interpolation algorithms to encode contextual information of road networks
2. Two-layer regression model to capture coarse-to-fine contextual features in a short time window for improved localization performance

Deployment in Spark/Hadoop-based telco big data platform for city-scale localization of MDs

2. Context-aware coarse-to-fine regression (CCR) model

A machine learning-based positioning system from telco big data

1. Map-matching and interpolation algorithms to encode contextual information of road networks

2. Two-layer regression model to capture coarse-to-fine contextual features in a short time window for improved localization performance

Deployment in Spark/Hadoop-based telco big data platform for city-scale localization of MDs

2. Context-aware coarse-to-fine regression (CCR) model

A machine learning-based positioning system from telco big data

1. Map-matching and interpolation algorithms to encode contextual information of road networks

2. Two-layer regression model to capture coarse-to-fine contextual features in a short time window for improved localization performance

Deployment in Spark/Hadoop-based telco big data platform for city-scale localization of MDs

2. Context-aware coarse-to-fine regression (CCR) model

A machine learning-based positioning system from telco big data

1. Map-matching and interpolation algorithms to encode contextual information of road networks
2. Two-layer regression model to capture coarse-to-fine contextual features in a short time window for improved localization performance

Deployment in Spark/Hadoop-based telco big data platform for city-scale localization of MDs

Overview

MR data generation

- MR data contain measurement parameters such as the Radio Signal Strength Indicator (RSSI) usually for 6 nearby sectors
- A mobile device connects to the sector with the strongest RSSI
- Only RSSI and engineering parameters of location-associated base stations are used to build the MR-based positioning system

MR data generation

- MR data contain measurement parameters such as the Radio Signal Strength Indicator (RSSI) usually for 6 nearby sectors
- A mobile device connects to the sector with the strongest RSSI
- Only RSSI and engineering parameters of location-associated base stations are used to build the MR-based positioning system

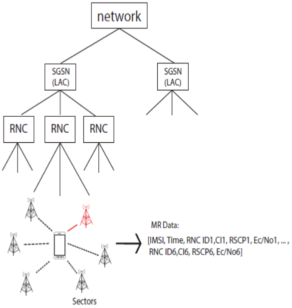
MR data generation

- MR data contain measurement parameters such as the Radio Signal Strength Indicator (RSSI) usually for 6 nearby sectors

- A mobile device connects to the sector with the strongest RSSI

- Only RSSI and engineering parameters of location-associated base stations are used to build the MR-based positioning system

MR data generation



Field	Example	Field	Example
MRTime	2015/12/16 08:00:00.000	IMSI	************058
SRNCID	2350	BestCellID	31171
SRNTI	4753	RAB	0,1,2,3,4,5
Event	0,1,2,3,5,8,-999	Delay	3
UE_TXPower	20	RNCID_1	2350
CellID_1	31171	EcNo_1	-6.5
RSCP_1	-85	RTT_1	1037
UE_Rx_Tx_1	1033	RNCID_2	2350
CellID_2	31171	EcNo_2	-6.5
RSCP_2	-85	RTT_2	1037
UE_Rx_Tx_2	1033	...	...
...	...	...	...
RNCID_6	2350	CellID_6	31171
EcNo_6	-6.5	RSCP_6	-85
RTT_6	1037	UE_Rx_Tx_6	1033

Figure 1: a) MR data generation procedure. b) Example of MR record

Data flow of CCR localization model in telco big data platform

Figure 2: a) OTT GPS locations generated by LBSs (low sampling rate - 60 seconds per record).
b) Two examples of MR data generated by call or MBB services (high sampling rate - 8 seconds per record)

Data flow of CCR localization model in telco big data platform

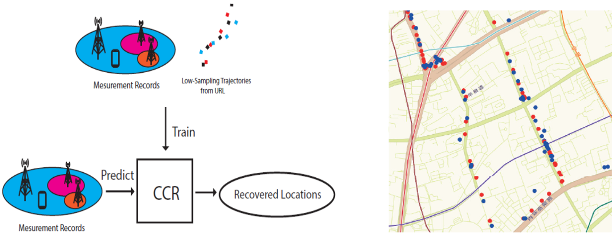


Figure 2: c) CCR training input - GPS data: training labels, MR data: training features; CCR input - MR records; CCR output - Recovered locations.
d) Predicted locations (blue dots), ground truth GPS locations (red dots)

Training Data Acquisition

1. Extract urls from OTT data, many of them contain GPS coordinates
2. Extract GPS coordinates from urls
   - Segmentation of long GPS trajectory
3. Map-matching method and most-frequent path interpolation to increase the sampling rate of GPS data
   3.1 Match each GPS coordinates in a short trajectory with the road network
   3.2 For each pair of adjacent matched points, compute the most-frequent-path between them and do a uniform interpolation on it, with a constant small time step
4. Match interpolated GPS data with MR data, by IMSI (International Mobile Subscriber Identity) and time stamp in a sliding window

Note: Steps 3.1 and 3.2 are implemented in parallel, in a distributed Hadoop/Spark system

Training Data Acquisition

1. Extract urls from OTT data, many of them contain GPS coordinates

2. Extract GPS coordinates from urls
   – Segmentation of long GPS trajectory

3. Map-matching method and most-frequent path interpolation to increase the sampling rate of GPS data
   3.1 Match each GPS coordinates in a short trajectory with the road network
   3.2 For each pair of adjacent matched points, compute the most-frequent-path between them and do a uniform interpolation on it, with a constant small time step

4. Match interpolated GPS data with MR data, by IMSI (International Mobile Subscriber Identity) and time stamp in a sliding window

Note: Steps 3.1 and 3.2 are implemented in parallel, in a distributed Hadoop/Spark system

Training Data Acquisition

1. Extract urls from OTT data, many of them contain GPS coordinates
2. Extract GPS coordinates from urls
   - Segmentation of long GPS trajectory
3. Map-matching method and most-frequent path interpolation to increase the sampling rate of GPS data
   3.1 Match each GPS coordinates in a short trajectory with the road network
   3.2 For each pair of adjacent matched points, compute the most-frequent-path between them and do a uniform interpolation on it, with a constant small time step
4. Match interpolated GPS data with MR data, by IMSI (International Mobile Subscriber Identity) and time stamp in a sliding window

Note: Steps 3.1 and 3.2 are implemented in parallel, in a distributed Hadoop/Spark system

Training Data Acquisition

1. Extract urls from OTT data, many of them contain GPS coordinates
2. Extract GPS coordinates from urls
   - Segmentation of long GPS trajectory
3. Map-matching method and most-frequent path interpolation to increase the sampling rate of GPS data
   - 3.1 Match each GPS coordinates in a short trajectory with the road network
   - 3.2 For each pair of adjacent matched points, compute the most-frequent-path between them and do a uniform interpolation on it, with a constant small time step
4. Match interpolated GPS data with MR data, by IMSI (International Mobile Subscriber Identity) and time stamp in a sliding window

Note: Steps 3.1 and 3.2 are implemented in parallel, in a distributed Hadoop/Spark system

Training Data Acquisition

1. Extract urls from OTT data, many of them contain GPS coordinates

2. Extract GPS coordinates from urls
   - Segmentation of long GPS trajectory

3. Map-matching method and most-frequent path interpolation to increase the sampling rate of GPS data
   3.1 Match each GPS coordinates in a short trajectory with the road network
   3.2 For each pair of adjacent matched points, compute the most-frequent-path between them and do a uniform interpolation on it, with a constant small time step

4. Match interpolated GPS data with MR data, by IMSI (International Mobile Subscriber Identity) and time stamp in a sliding window

Note: Steps 3.1 and 3.2 are implemented in parallel, in a distributed Hadoop/Spark system

CCR Training

Training of a two-layer regression model with the GPS-labeled MR data, using Random Forest (RF) for each layer

First layer:

- Input $\longrightarrow$ 258 dimensional coarse features

Features	Description
rssi	Received Signal Strength Indication
rscp	Received Signal Code Power
ecno	Ratio of energy per modulating bit to the noise spectral density
rncid	RNC id
ci	cell id
lon	longitude of a sector
lat	latitude of a sector
id	unique id for a BS
height	height of a BS
azimuth	azimuth of a BS
mdtilt	Mechanical Down Tilt
edtilt	Electrical Down Tilt
bs_type	type code of a BS (such Metrocell, Microcell, etc.)
company	producer of a BS (such as HUAWEI, Nokia)
n_sector	# associated sectors
n_bs	# associated BSs

- Output $\longleftarrow$ predicted GPS location for each training sample

CCR Training

Second layer:

- Input $\longrightarrow$ 34 dimensional fine-grained features

Features	Description
pred_lon	p_i's longitude from the output of the first layer
pred_lat	p_i's latitude from the output of the first layer
last_distance	distance between p_{i-1} and p_i
last_direction	direction from p_{i-1} to p_i
last_speed	average speed from p_{i-1} to p_i
last_time_gap	time gap between p_{i-1} and p_i
last_lon	p_{i-1}'s longitude from the output of the first layer
last_lat	p_{i-1}'s latitude from the output of the first layer
next_distance	distance between p_i and p_{i+1}
next_direction	direction from p_i to p_{i+1}
next_speed	average speed from p_i to p_{i+1}
next_time_gap	time gap between p_i and p_{i+1}
next_lon	p_{i+1}'s longitude from the output of the first layer
next_lat	p_{i+1}'s latitude from the output of the first layer
last_heading_change_angle	delta angle between p_{i-1} and p_i
last_speed_change	delta speed between p_{i-1} and p_i
next_heading_change_angle	delta angle between p_i and p_{i+1}
next_speed_change	delta speed between p_i and p_{i+1}
angle2azimuth	the angle between p_i and main BS's azimuth

- Output $\longleftarrow$ predicted GPS location with higher longitude/latitude accuracy

CCR Prediction

- CCR maps the RSSI feature vector of an MD to the predicted location coordinate
- Predicting a sequence of location coordinates $\Rightarrow$ recovery of the trajectory of the MD on the road networks

Note 1: Feature engineering and RF regression model are hand coded in Spark, based on Hive/Spark SQL and the parallel RF algorithm

Note 2: Scalability for big data without communication cost between machines, because each regression tree is trained independently

CCR Prediction

- CCR maps the RSSI feature vector of an MD to the predicted location coordinate
- Predicting a sequence of location coordinates $\Rightarrow$ recovery of the trajectory of the MD on the road networks

Note 1: Feature engineering and RF regression model are hand coded in Spark, based on Hive/Spark SQL and the parallel RF algorithm

Note 2: Scalability for big data without communication cost between machines, because each regression tree is trained independently

Overview

Dataset

Type	Number
Time	30 days
Area	12×11 square killometers
Data blocks	4×4 blocks
Number of BSs	$2,697$
Number of Sectors	$18,431$
Number of IMSIs	$17,699$
Number of Trajectories	$2,181,990$
Number of GPS-associated MR records before map-matching and interpolation	$4,749,150$
Number of GPS-associated MR records after map-matching and interpolation	$103,605,330$

Figure 3: Statistics of GPS-associated MR Dataset

Stability Measurement

Smaller distance to the sector lead to stronger RSSI values

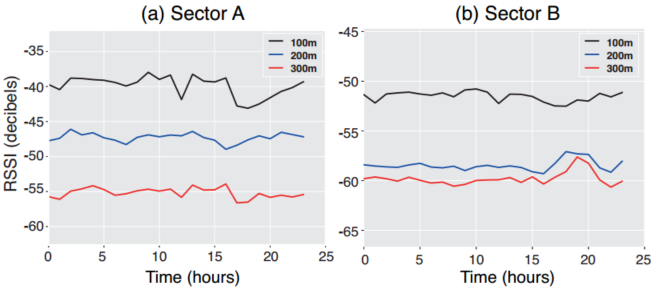


Figure 4: Stability measures over time (hours) in two typical sectors

Stability Measurement

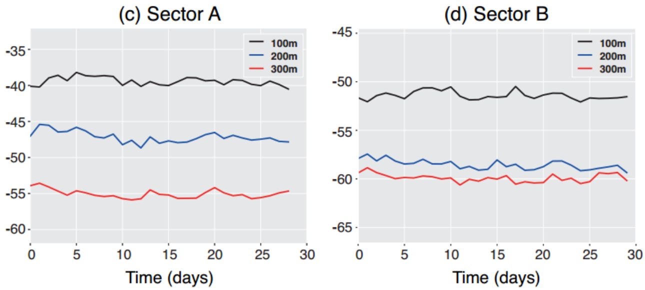


Figure 5: Stability measures over time (days) in two typical sectors

Stability

- Even if MD stays at the same position, it receives different RSSI
- The $|\Delta\text{RSSI}|$ over days is often less than $0.7db$
- Changes in RSSI less than ± 0.7 is indistinguishable because of instability

Key inspiration

- If the localization model is not updated by the new training data, the model may fail to capture the temporal change of RSSI for bad localization performance.
- Experiments confirm that update models by new coming data will enhance the overall localization performance

Stability

- Even if MD stays at the same position, it receives different RSSI
- The $|\Delta\text{RSSI}|$ over days is often less than $0.7db$
- Changes in RSSI less than ± 0.7 is indistinguishable because of instability

Key inspiration

- If the localization model is not updated by the new training data, the model may fail to capture the temporal change of RSSI for bad localization performance.
- Experiments confirm that update models by new coming data will enhance the overall localization performance

Sensitivity Measures

Select pairs where the $|\Delta\text{RSSI}| \leq 0.7$ (0.7 temporal fluctuation threshold)

Plots the histogram of each distance ΔGPS of those pairs

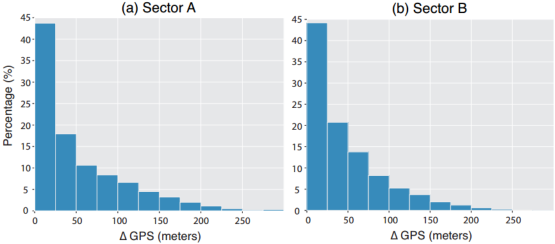


Figure 6: Sensitivity measurement over space of MR data (using ΔGPS when $|\Delta\text{RSSI}| \leq 0.7$ for two sectors A and B

Sensitivity

- The above distances cannot differentiated by the localization model
- More than 50% of pairs have the distance $\geq 25m$
- Lots of tested pairs with distance ΔGPS larger than $25m$ have almost zero change of RSSI.
- The sensitivity of MR data within the radius $25m$ is too low to be recognised in practice

To enhance performance

Improve the sensitivity of the sensor by designing new:

- Hardware
- Signal measurement methods

Sensitivity

- The above distances cannot differentiated by the localization model
- More than 50% of pairs have the distance $\geq 25m$
- Lots of tested pairs with distance ΔGPS larger than $25m$ have almost zero change of RSSI.
- The sensitivity of MR data within the radius $25m$ is too low to be recognised in practice

To enhance performance

Improve the sensitivity of the sensor by designing new:

- Hardware
- Signal measurement methods

Uncertainty Measurements

MDs do not always connect to the closest sector or BS

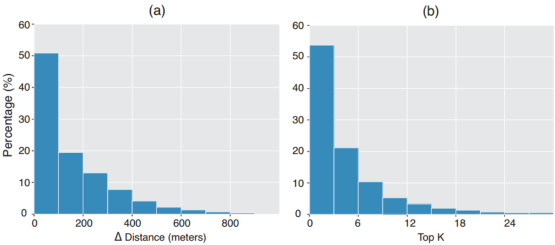


Figure 7: Uncertainty measurement of all MR data. (a) shows the distribution over ΔDistance between the connected BS and the closest BS. (b) shows the distribution over the top K closest BSs

Uncertainty

- Around 50% data connect to the BS within $100m$ of the closest one
- If ΔDistance $\geq 300m$ the localization performance is not satisfactory
- More than 50% MR data are yielded without connecting with top 6 closest BS $\longrightarrow$ worse localization performance

How to improve

Add more BSs in the cells, which will assure an MD always find the closest BS for stronger RSSI values and better service quality

Uncertainty

- Around 50% data connect to the BS within $100m$ of the closest one
- If ΔDistance $\geq 300m$ the localization performance is not satisfactory
- More than 50% MR data are yielded without connecting with top 6 closest BS $\longrightarrow$ worse localization performance

How to improve

Add more BSs in the cells, which will assure an MD always find the closest BS for stronger RSSI values and better service quality

Missing Value Measurement

MR records does not always have RSSI values from 6 nearby sectors

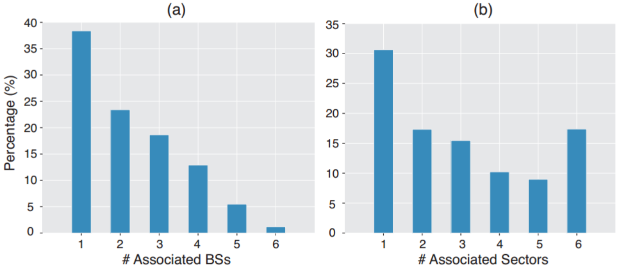


Figure 8: The distribution of MR data over the number of associated BSs (a) and sectors(b)

Missing Value

- RSSI values from different sectors in the same BS cannot improve the localization accuracy
- More than 60% of MR data have RSSI values from less than 2 location-associated BSs $\longrightarrow$ large localization error

How to overcome missing value problem

1. Add more BSs or improve sensor systems
2. Design missing value algorithm to estimate missing values

Missing Value

City-Scale Localization with Telco Big Data

- RSSI values from different sectors in the same BS cannot improve the localization accuracy
- More than 60% of MR data have RSSI values from less than 2 location-associated BSs $\longrightarrow$ large localization error

How to overcome missing value problem

1. Add more BSs or improve sensor systems
2. Design missing value algorithm to estimate missing values

Overview

Experiments details

Platform:

- Six Huawei RH2288 servers
- Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz
- 40 Cores and 189G Memory

Data:

- Training data: 24 days (around 80%)
- Testing data: 6 days (around 20%)

Training:

- 540 trees in Random Forest distributed in 6 machines
- Train a CCR localization model for each data block
- 16 CCR models in total

Experiments details

Platform:

- Six Huawei RH2288 servers
- Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz
- 40 Cores and 189G Memory

Data:

- Training data: 24 days (around 80%)
- Testing data: 6 days (around 20%)

Training:

- 540 trees in Random Forest distributed in 6 machines
- Train a CCR localization model for each data block
- 16 CCR models in total

Experiments details

Platform:

- Six Huawei RH2288 servers
- Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz
- 40 Cores and 189G Memory

Data:

- Training data: 24 days (around 80%)
- Testing data: 6 days (around 20%)

Training:

- 540 trees in Random Forest distributed in 6 machines
- Train a CCR localization model for each data block
- 16 CCR models in total

Evaluation metrics

- Localization error $\longrightarrow$ distance between predicted and actual GPS position
- Cumulative density function (CDF) of errors
- Area under curve (AUC) of CDF with the error less than $100m$

CCR Performance

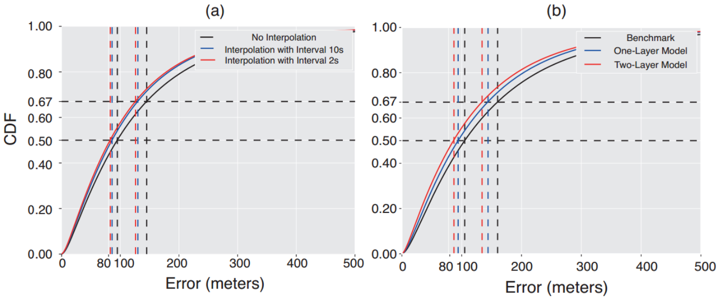


Figure 9: Comparisons of different CCR model settings: (a) Effectiveness of interpolation, and (b) Effectiveness of two-layer regression models

CCR on Stability

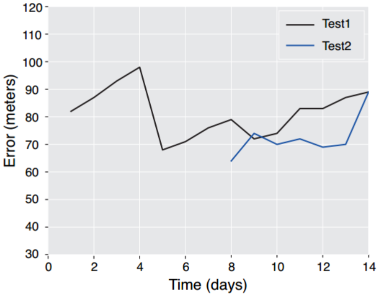


Figure 10: CCR on temporal stability of MR data. Test1 training: 1-7 days, Test2 training: 8-14 days

- Median error lower than that in Test1 $\longrightarrow$ updating CCR regularly captures the recent temporal variance in MR data

CCR on Sensitivity

- CCR produces a median error of $80m$
- The lower limit of the median error is $50m$
- Reason: The training data is insufficient
- Collection of more data by fine-grained walk with $0.5m$ per step
- Achieve median error around $46.1m$

Future work

- Increasing sensitivity of MR data
- Designing new sensors or new algorithms

CCR on Sensitivity

- CCR produces a median error of $80m$
- The lower limit of the median error is $50m$
- Reason: The training data is insufficient
- Collection of more data by fine-grained walk with $0.5m$ per step
- Achieve median error around $46.1m$

Future work

- Increasing sensitivity of MR data
- Designing new sensors or new algorithms

CCR on Uncertainty

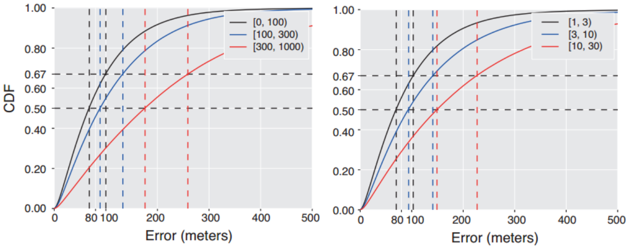


Figure 11: CCR performance on connection uncertainty of MR data: (a) ΔDistance and (b) Top K

- Bigger ΔDistance has the larger errors
- The average localization error is within 1/4 - 1/2 of the distance between the MD and the connecting BS
- Reduce the uncertainty by forcing the MD to connect to the closest BS

CCR on Missing value

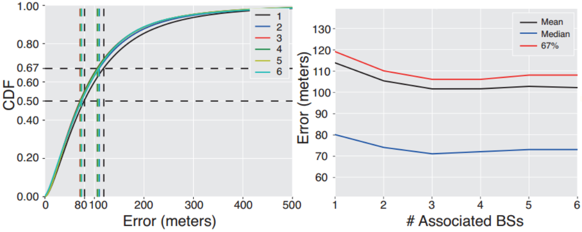


Figure 12: CCR localization performance on MR data with different missing number of BSs

- Less missing values lead to higher localization accuracy
- MR data with 3 associated BSs are enough to give satisfactory localization accuracy
- Improve by designing specific missing value imputation algorithms

Comparison and Error Analysis

1. Range-based method
   - Complex audio signal propagation patterns in urban areas
2. Cell* method
   - Uses some contextual knowledge through map-mapping
   - Does not use RSSI features
3. Fingerprinting method
   - Much better than Cell* method
4. Proposed CCR method
   - Designs a context-aware and coarse-to-fine features
   - Captures much more signal and context knowledge than previous methods

Comparison and Error Analysis

1. Range-based method
   - Complex audio signal propagation patterns in urban areas
2. Cell* method
   - Uses some contextual knowledge through map-mapping
   - Does not use RSSI features
3. Fingerprinting method
   - Much better than Cell* method
4. Proposed CCR method
   - Designs a context-aware and coarse-to-fine features
   - Captures much more signal and context knowledge than previous methods

Comparison and Error Analysis

1. Range-based method
   - Complex audio signal propagation patterns in urban areas
2. Cell* method
   - Uses some contextual knowledge through map-mapping
   - Does not use RSSI features
3. Fingerprinting method
   - Much better than Cell* method
4. Proposed CCR method
   - Designs a context-aware and coarse-to-fine features
   - Captures much more signal and context knowledge than previous methods

Comparison and Error Analysis

1. Range-based method
   - Complex audio signal propagation patterns in urban areas
2. Cell* method
   - Uses some contextual knowledge through map-mapping
   - Does not use RSSI features
3. Fingerprinting method
   - Much better than Cell* method
4. Proposed CCR method
   - Designs a context-aware and coarse-to-fine features
   - Captures much more signal and context knowledge than previous methods

Comparison and Error Analysis

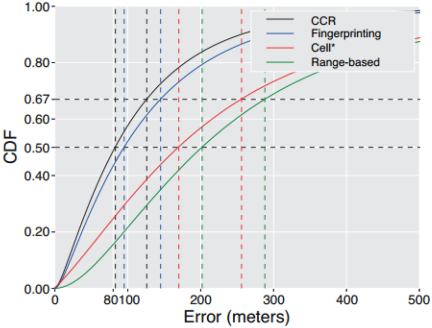


Figure 13: Comparison of CCR with the state-of-the-art localization methods: range based, Cell* and fingerprinting mehods

CCR Performance

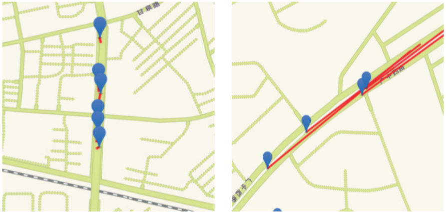


Figure 14: Localization error on CCR. Blue bubbles are the predicted values and red lines correct GPS location

- Small error on the same road
- Use knowledge of the road network
- Forward or backward error because training data is insufficient

CCR Performance

Figure 15: Localization error on CCR. Blue bubbles are the predicted values and red lines correct GPS location

- Close and parallel to road error
- Low temporal stability and spatial sensitivity of MR data
- Collect denser training data and design special algorithm in the areas with close parallel roads

CCR Performance

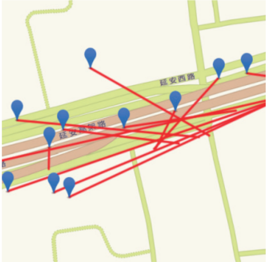


Figure 16: Localization error on CCR. Blue bubbles are the predicted values and red lines correct GPS location

- Large errors occurs in dense road network
- Collect more training data on areas with dense road networks to distinguish subtle detains among dense roads

The End

Thank you!!!

Bibliography

Fangzhou Zhu et al, **City-Scale Localization with Telco Big Data.** Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, p. 439-448, Indianapolis, Indiana, USA, 2016.
http://doi.acm.org/10.1145/2983323.2983345