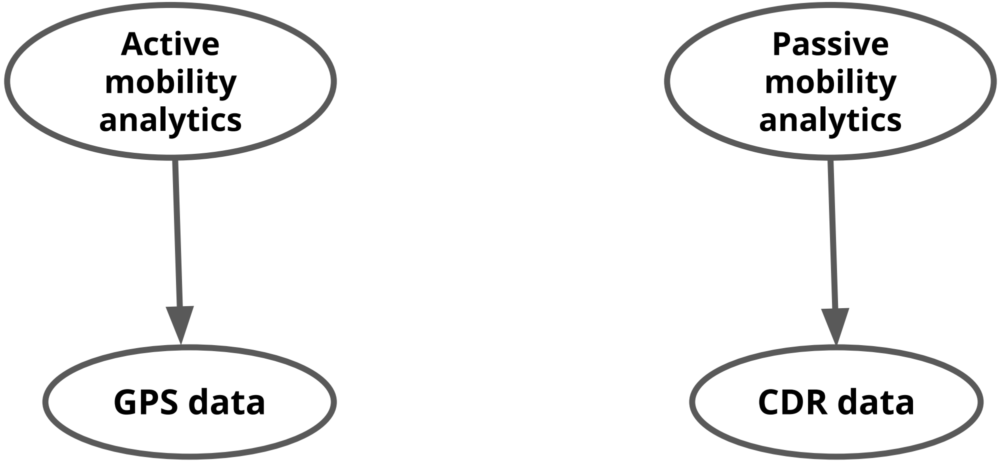
Mobile Positioning and Trajectory Reconstruction Based on Mobile Phone Network Data: A Tentative Using Particle Filter

Salijona Dyrmishi, Amnir Hadachi

Introduction



GPS data

- \+ High accuracies
- \+ High sampling rate

- \- Not always available
- \- High extraction cost

CDR data

- \- High spatial uncertainties
- \- High temporal uncertainties

- \+ Available for all mobile users
- \+ Low extraction cost

Introduction

Call detail records (CDR) :

- Collected for billing purposes
- SMS, Call, 4G/5G events

CDR fields:

- Calling party
- Called party
- Billed number
- Time
- Type
- Cell global identity (CGI)

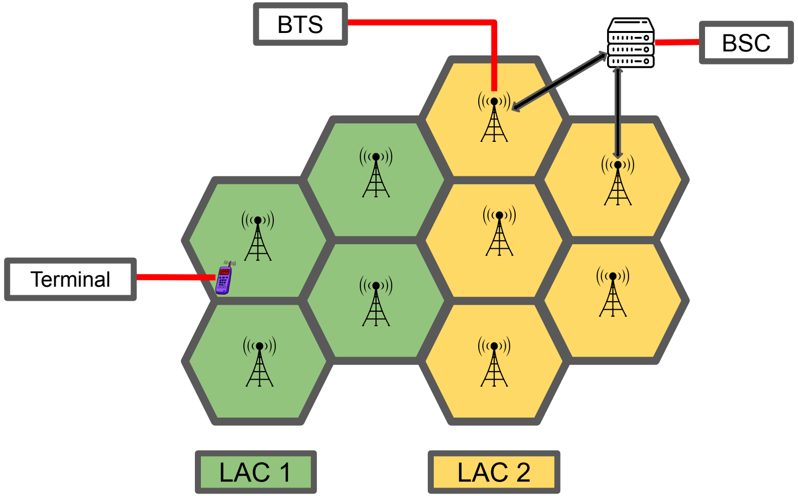


Figure 1: Components of mobile network

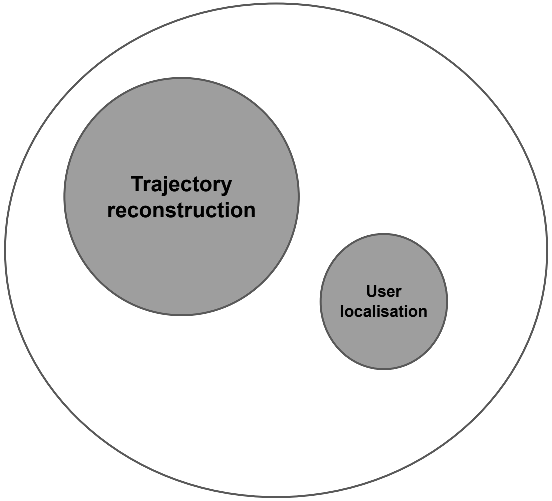
Related work

Trajectory reconstruction:

- Aims at reducing the temporal uncertainties
- Methods:
  - Interpolation techniques[1] (*Hoteit et al.*)
  - Tensor factorization[2] (*Chen et al.*)
  - Shortest path[3] (*Vajakas et al.*)

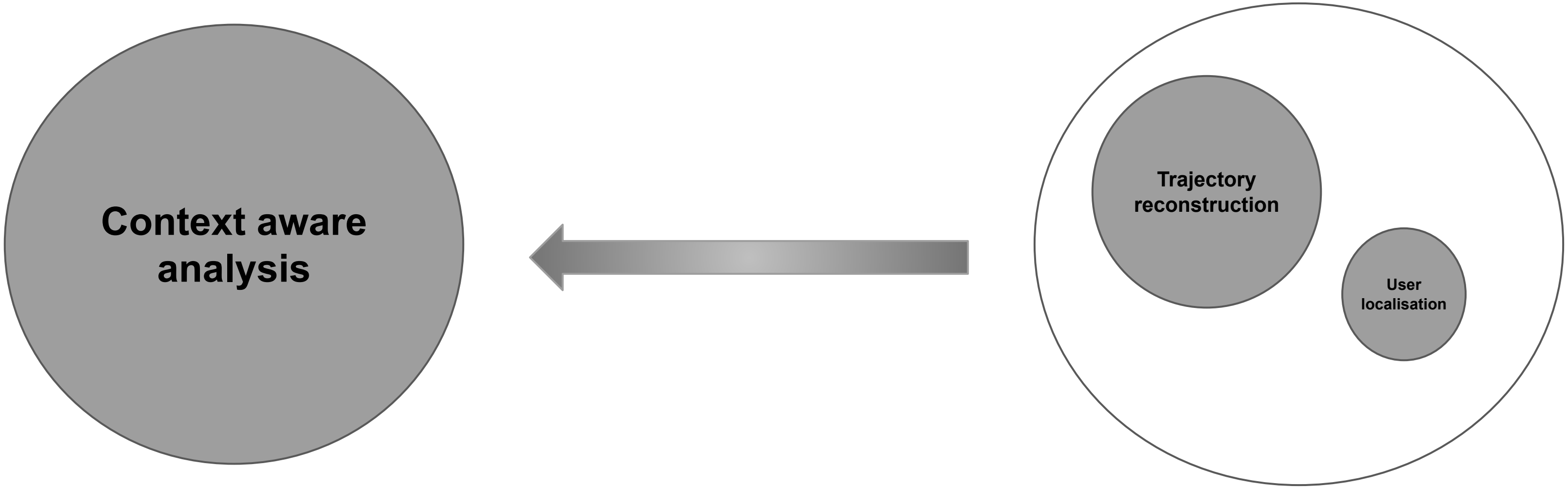
User localisation:

- Aims at reducing the spatial uncertainties
- Methods:
  - Switching Kalman Filter[4] (*Lind et al.*)



Motivation

Aim: Explore the possibility of using a ***new non-linear*** method compared to ***existing linear*** method.



Methodology: Particle filtering

Task: Model phenomena using observations

- 1. Kalman Filter
- 2. Hidden Markov Models
- 3. Particle Filtering

Particle filtering

- Presented in 1993 as bootstrap filtering
- Not normal noise, non-linear models
- Approximate solution

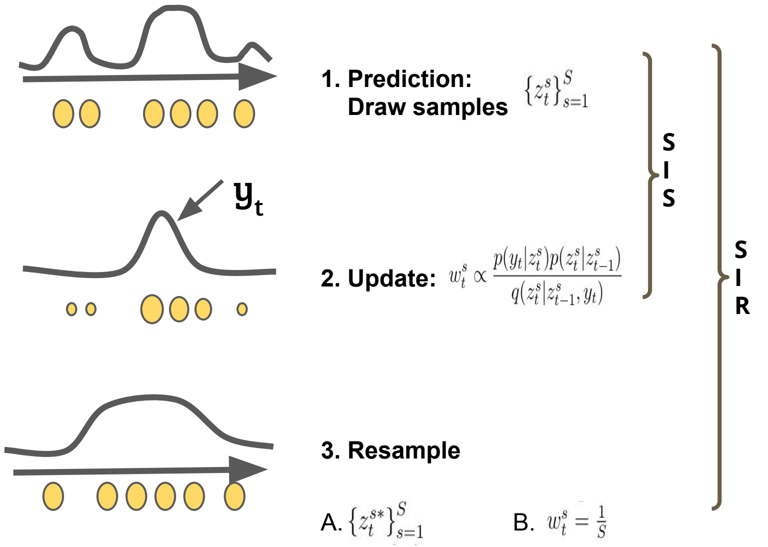


Figure 2. Particle filter schematic view

Methodology: Particle filtering for user positioning

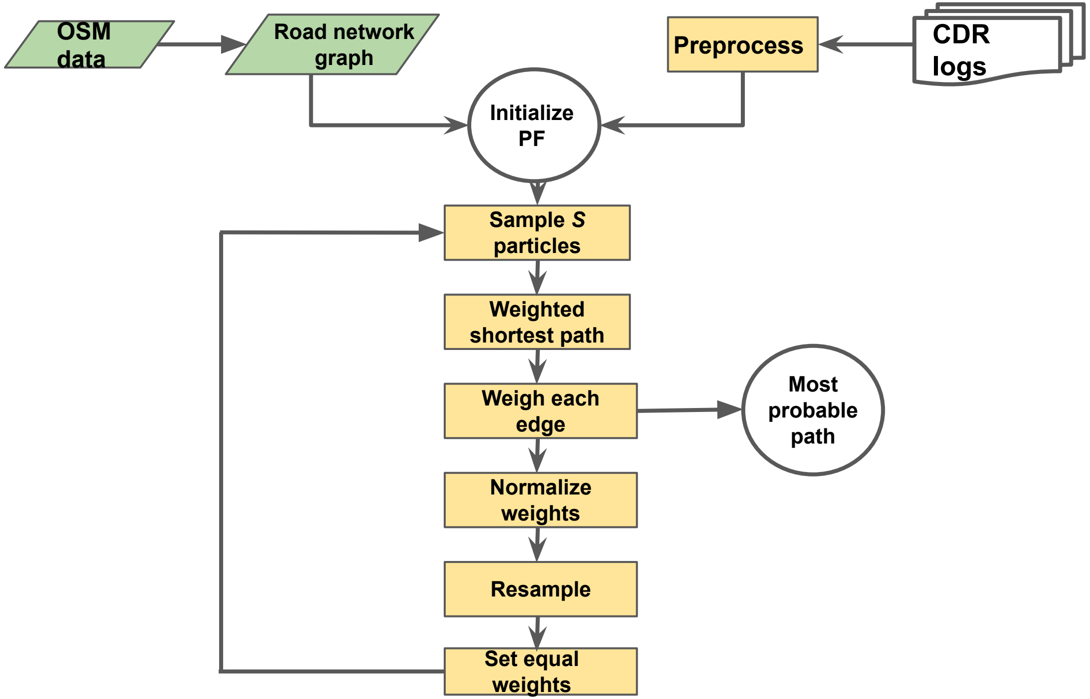
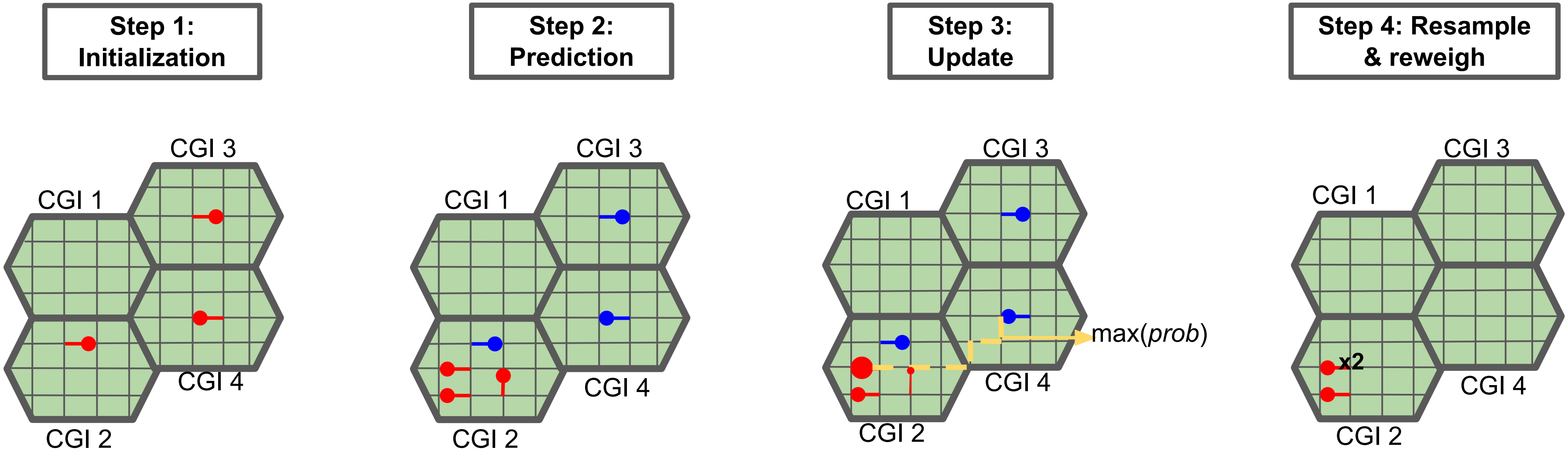


Figure 3. Mobile positioning flow chart

Methodology: Particle filtering for user positioning



$\mathbf{y}_t \longrightarrow$ *(CGI 2; Timestamp t)*

$$\text{Prob} = \frac{|(T_2 - T_1) \times TD|}{(T_2 - T_1)}$$

T_2 - timestamp of present event
T_1 - timestamp of past event
TD - estimated travel time between two locations

Figure 4. User positioning using mobile data

Experiments: Synthetic data

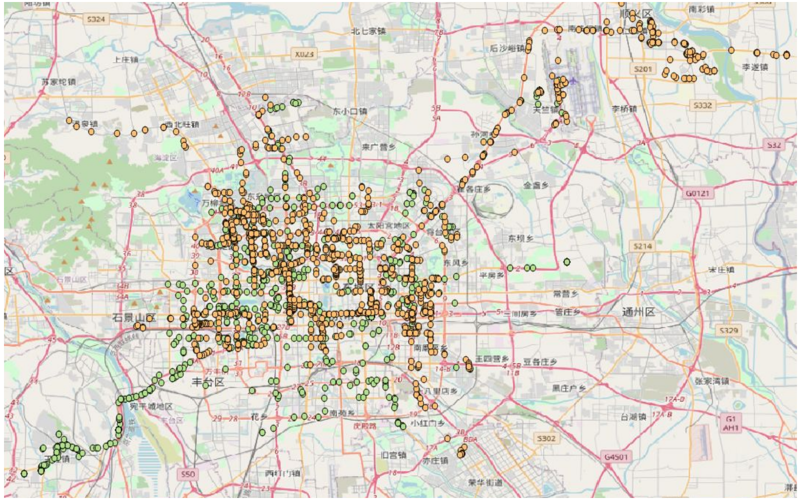


Figure 5. GPS locations from T-drive dataset [5]

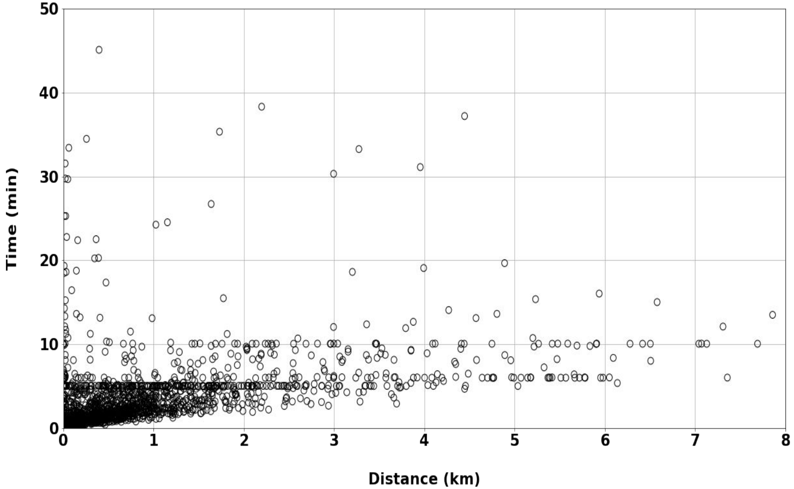


Figure 6. Time and distance relation in T-drive dataset

Experiments: Synthetic data

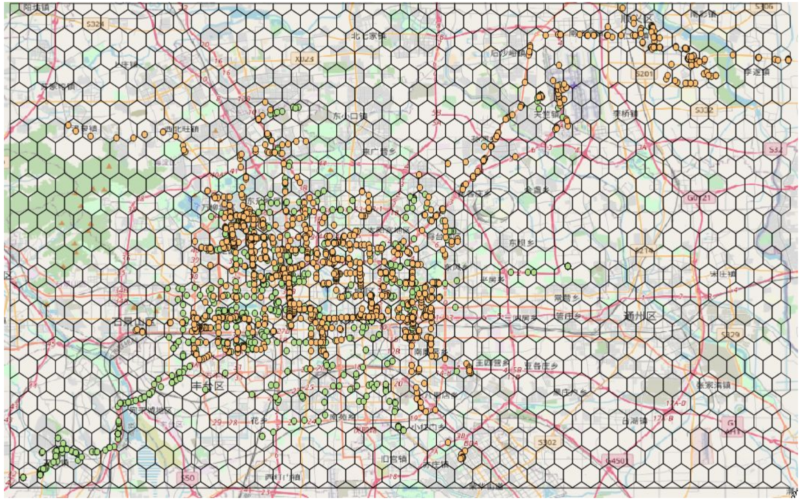


Figure 5. GPS locations from T-drive dataset [5]

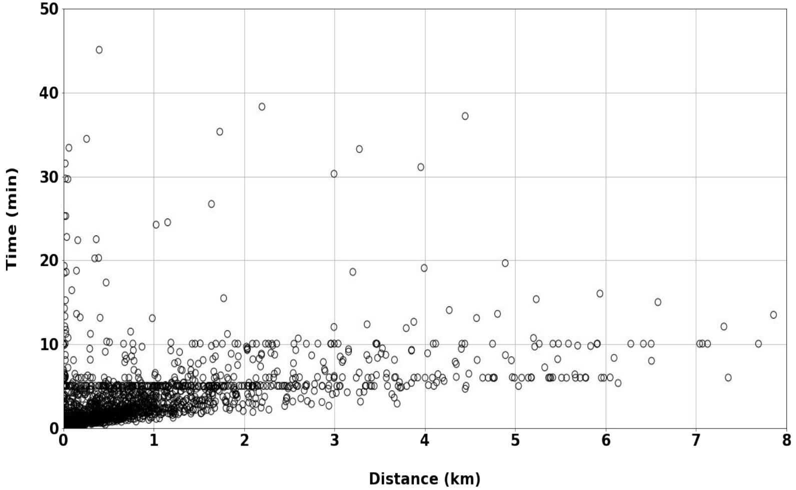


Figure 6. Time and distance relation in T-drive dataset

Experiments: Synthetic data results

The ***error*** between PF prediction and GPS position is calculated using ***Haversine distance***

$$a = sin^2(\frac{\Delta\Phi}{2}) + cos(\Phi_1) \cdot cos(\Phi_2) \cdot sin^2(\frac{\Delta\Lambda}{2})$$

$$d = R \cdot 2 \cdot atan2(\sqrt{a}, \sqrt{1-a})$$

Φ: latitude
Λ: longitude,
R: Earth's radius (mean radius = 6,371 km)

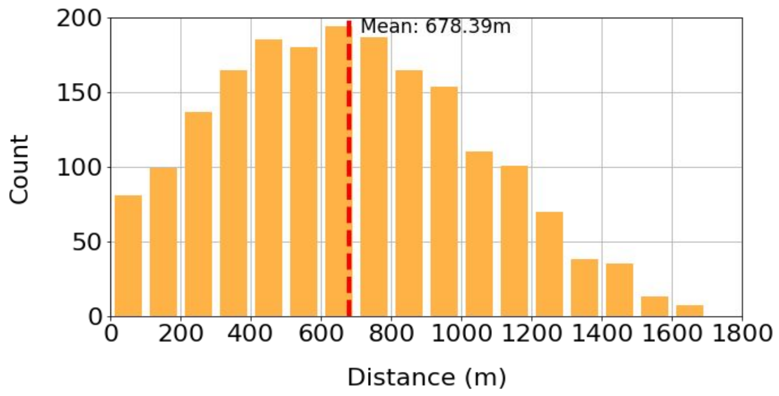


Figure 7. Error distribution using 50 particles in generated CDR events

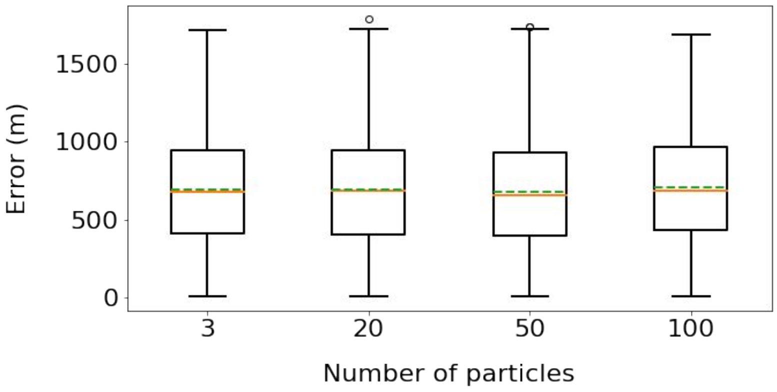


Figure 8. The effect of sample size

Experiments: Synthetic data results

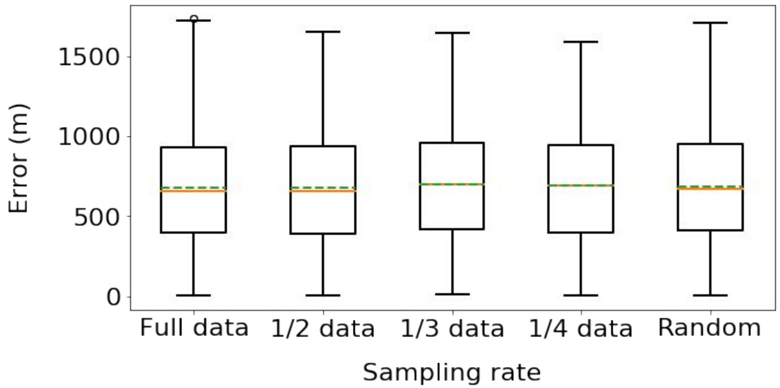


Figure 9. The effect of time granularity

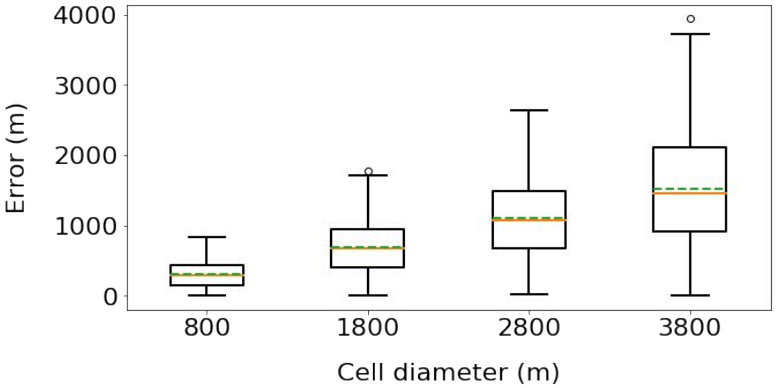


Figure 10. The effect of cell coverage surface

Experiments: Real use case

CDR data of five mobile owners in Estonia + GPS locations

Data extraction period is between April and August 2015

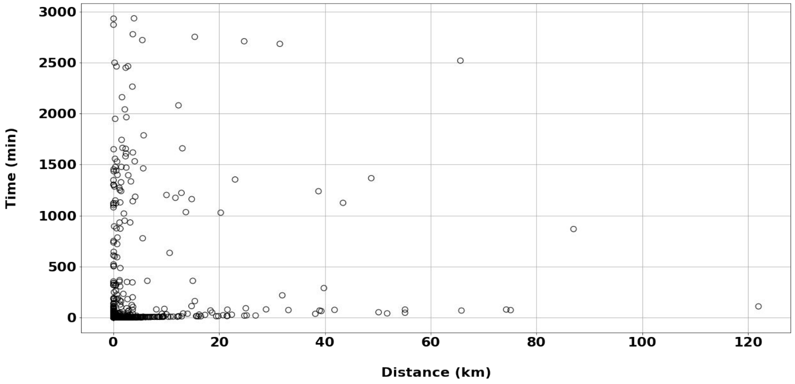


Figure 11. Time and distance relation in real CDR events

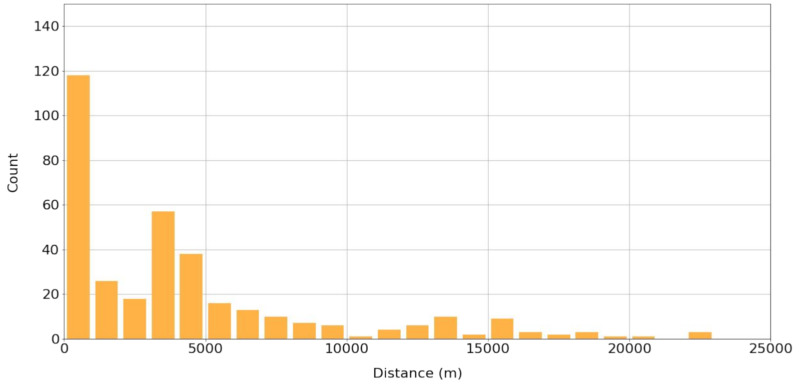


Figure 12. Cell diameter distribution in real CDR events

Experiments: Real use case results

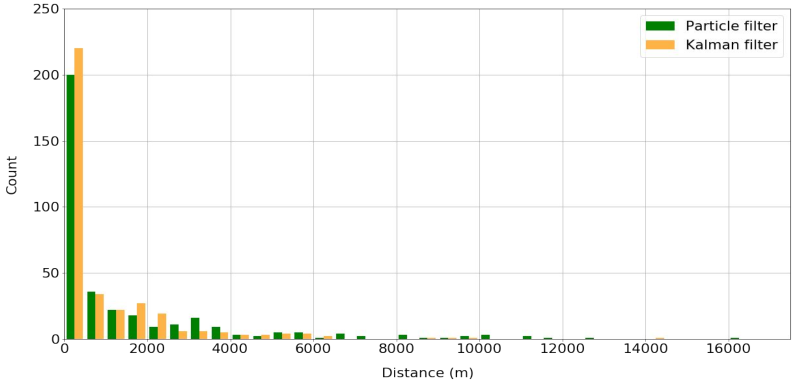


Figure 13. Distribution of errors for Particle Filter and Switching Kalman Filter in real CDR data

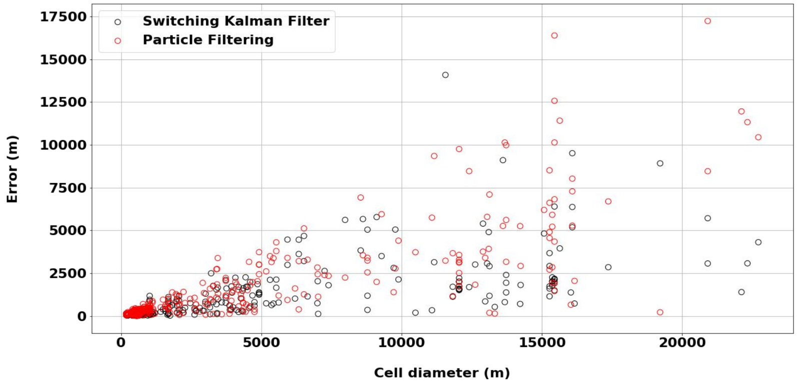


Figure 14. The relation between the accuracy and the diameter size in real data

Experiments: Real use case path results

$$Acc = \frac{len(E_{gps} \cap E_{pf})}{len(E_{gps})} \cdot 100$$

$$E_{gps} = e_1, e_2, ...e_k$$

$$E_{pf} = e_1, e_2, ...e_j$$

Average acc for PF: **17%**
Average acc for SKF: **14%**

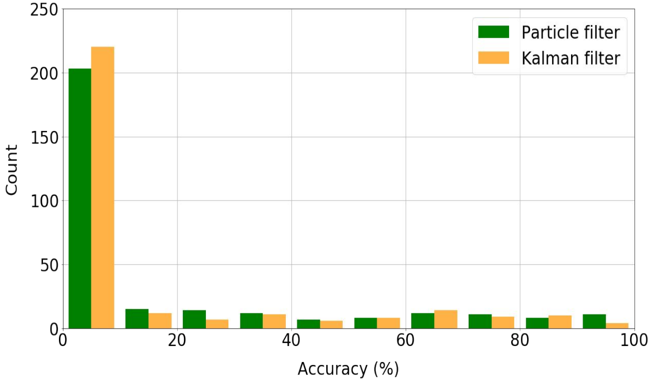


Figure 15. Path accuracy in real data

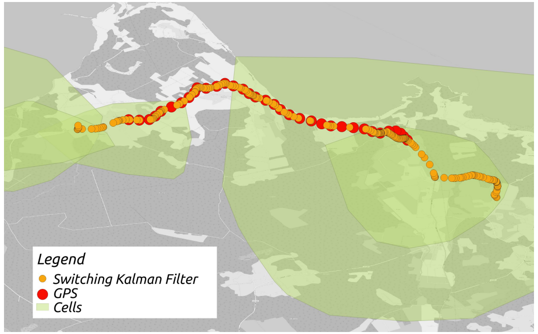


Figure 16. Path with accuracy 85% (SKF)

Figure 17. Path with accuracy 87% (PF)

Discussions

Travel time estimation between two nodes is one major factor that affects our accuracy

- ***FW:*** *Sophistication in the method we use for travel time estimation can take into consideration the traffic situation, most common paths, traffic rules, etc.*

The second important factor is coverage areas information

- ***FW:*** *Preprocessing step like coverage area optimization or overlap detection*

Conclusions

Based on the analysis we could conclude our approach was achieving similar results but not better compared to the previous linear technique like Switching Kalman Filter. Although, our method had a 3% higher accuracy in the path evaluation.

References

[1] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle, “Estimating human trajectories and hotspots through mobile phone data,” Computer Networks, vol. 64, pp. 296–307, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128614000656

[2] G. Chen, A. Viana, M. Fiore, and C. Sarraute, “Complete Trajectory Reconstruction from Sparse Mobile Phone Data,” EPJ Data Science, vol. 8, 2019.

[3] T. Vajakas, J. Vajakas, and R. Lillemets, “Trajectory reconstruction from mobile positioning data using cell-to-cell travel time information,” International Journal of Geographical Information Science, vol. 29, no. 11, pp. 1941–1954, 2015. [Online]. Available: https://doi.org/10.1080/13658816.2015.1049540

[4] A. Lind, A. Hadachi, and O. Batrashev, “A new approach for mobile positioning using the cdr data of cellular networks,” in 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS).

[5] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. Driving with knowledge from the physical world. In The 17th ACM SIGKDD international conference on Knowledge Discovery and Data mining, KDD’11, New York, NY, USA, 2011. ACM.

[6] Montella, C. (2011). The Kalman Filter and Related Algorithms: A Literature Review.

Final discussions

1. History of the trajectory as influence
2. Running times
    a. Particle Filtering time complexity can scale up to $O(M^n)$[6]
    b. Kalman Filter time complexity can scale up to $O(n^3)$
3. Static number of particles for different cell sizes
4. Sampling and resampling in the algorithm is based on the uniform distribution: map matching proposal
5. 3 hours for breaking the trips