Uncovering Mobile Infrastructure in Developing Countries with Crowdsourced Measurements

Mah-Rukh Fida and Mahesh K. Marina
School of Informatics
The University of Edinburgh

Other than network operators, knowledge about deployment patterns of a communication infrastructure enable a variety of third party services for customers, regulators and the wider eco-system. Common applications of which include getting insight on a network's service areas, identifying rouge transmitters, cheap end-device localization and guidance on potential EMF radiation areas. Network operators, however, treat their infrastructure related information sensitive from their market position standpoint and generally do not disclose it, except to regulators and policy makers and that too with a non-disclosure agreement. Measuring from the outside (say, from user devices) and making inferences about the infrastructure is therefore the only means to estimate this information.

We therefore focus on localizing cell towers with crowdsourced measurements that are the samples contributed by users from their end-devices and crowdsourced performance reporting applications. From literature we find a number of algorithms proposed for measurement based infrastructure localization [1–5], most of which have emanated from the Wi-Fi AP localization context. Achieving accuracy and robustness by applying these algorithms on the crowdsourced samples, however, becomes challenging due to the lack of control over the measurement process on the device side. These measurements come with heterogeneous characteristics such as differences in sampling layout, density, size, location inaccuracy, correlation between RSS and response rate of samples to distance from the cell tower and noise in samples' RSS values.

The heterogeneity of measurement scenarios results in none of the localization algorithms being consistently superior to others. To address this issue, some studies recommend either pre-processing [1]. Other recent approach, exemplified by [6] which proposes "Filtered" Weighted Centroid (FWC), is to retain only a *predictive* subset of samples within each cell (measurement scenario); the *predictive* measurement set means the subset of samples with which a chosen localization algorithm can estimate a cell tower location fairly accurately.

Our approach [7] to deal with heterogeneity in measurements is fundamentally different in that that we do not fix the localization algorithm nor we perform any pre-processing steps and limit ourselves to predictive samples. We rather adaptively select an algorithm, among a set of five commonly used algorithms belonging to geometric, RSS and path loss based localization categories, that is expected to most accurately locate the cell tower for the given measurement scenario. Our proposal is based on a supervised learning model that we call as Adaptive Algorithm Selection (AAS). AAS uses a number of features derived from the measurement samples of cells as independent variables. For response variable it uses a label indicating the localization algorithm that most accurately estimates a cell tower location while using its related measurement samples. Under different classification schemes we find Random Forest to achieve highest classification accuracy in generating an AAS model. A trained AAS model thus, adaptively, chooses an algorithm from the suite of five commonly used localization algorithm that it expects to arrive at the location of a test cell relatively accurately.

Through an extensive measurement based evaluation, using the OpenCelliD dataset, we show that AAS approach significantly outperforms FWC by more than 65% in median error and reduces the mean error by more than half. At the same time, AAS achieves median error within 20% of the Oracle scheme that always picks the best performing algorithm. Other than cellular networks we validate good performance of AAS over single algorithms in WLAN AP localization settings. Moreover, we apply AAS to our driving use case of mapping mobile infrastructure in developing countries, where usually this information is not publicly available. Even when using a model trained on data from separate region and carrier, AAS results in better accuracy than when using a single fixed algorithm, especially when features' distribution of trained and test datasets match.

References

[1] J. Yang et al. Accuracy Characterization of Cell Tower Localization. In *Proc. ACM UbiComp*, 2010.

[2] E. Neidhardt et al. Estimating Locations and Coverage Areas of Mobile Network Cells based on Crowdsourced Data. In *Proc. WMNC'13*, 2013.

[3] P. Nurmi, S. Bhattacharya, and J. Kukkonen. A Grid-Based Algorithm for On-Device GSM Positioning. In *Proc. ACM UbiComp*, 2010.

[4] Y. Cho et al. Improved Wi-Fi AP position estimation using regression based approach. In *Proc. Int'l Conf. on Indoor Positioning and Indoor Navigation (IPIN)*, 2012.

[5] M. Ji et al. A Novel Wi-Fi AP Localization Method Using Monte Carlo Path-loss Model Fitting Simulation. In *Proc. IEEE PIMRC*, 2013.

[6] Z. Li et al. Identifying Value in Crowdsourced Wireless Signal Measurements. In *Proc. ACM World Wide Web Conference (WWW)*, 2017.

[7] M. Fida and M. K. Marina. Uncovering Mobile Infrastructure in Developing Countries with Crowdsourced Measurements. In *The Tenth International Conference on Information and Communication Technologies and Development (ICTDX*, 2019.