Inferring Missing Cell Tower Locations using Call Handoffs

Daniel Björkegren
Brown University

When analyzing cellular network usage, data on cell tower locations may not be complete. This document outlines a simple procedure to estimate the geographical locations of unknown towers based on call handoffs with known towers.

Problem: location records are incomplete

We have a set of cellular towers whose GPS coordinates are known, K, and a set whose coordinates are unknown and to be estimated, U. We also have usage data from the mobile phone operator which references all towers. Specifically, let us consider using anonymized call detail records (CDRs) which list the tower used at the beginning and end of a transaction for both the sending and receiving phone.

Approach: use call data to determine relationship between towers

One way to infer the locations of the missing towers is to take advantage of calls that were handed off from one tower to another during a call. This can happen if a person moves during a call, or if a tower is overloaded. If many calls of short duration are handed off from tower X to tower Y, this suggests that X and Y are near each other. Thus, one straightforward way to infer the missing tower locations is to perform a weighted sum of the coordinates of known towers, where weights are derived from the nearness of the towers.

More formally, we predict the coordinates of an unknown tower $\mathbf{x_u}=(x_u^{long}, x_u^{lat})$ by computing a weighted average of the coordinates of the set of known towers K, with weights given by a metric w_{ku} of the relationship between k and u:

$$\hat{\mathbf{x}} = \sum_{k \in K} w_{ku} \mathbf{x_k}$$

although more complex metrics can be used, a simple metric based on the number of handoffs between towers k and u works quite well:

$$w_{ku} = \frac{N_{ku}^{handoffs}}{\sum_{j \in K} N_{ju}^{handoffs}}$$

The metric could also be made more sophisticated, for example using only short calls that were handed off (when the duration of a call is short, it is less likely that a call was handed off due to travel). In this first pass, I found that this information was less useful than the raw number of handoffs, presumably because even short calls can be handed off for reasons other than travel (such as load balancing).

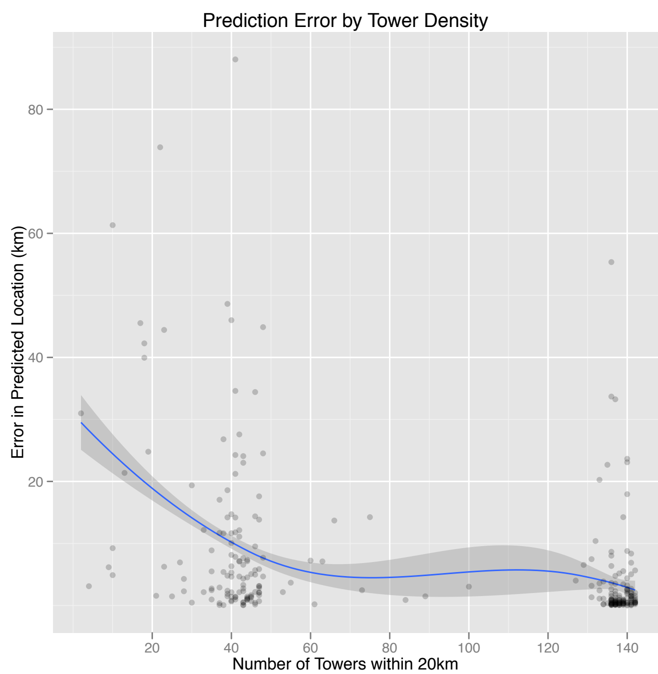
This work is supported by the Stanford Institute for Economic Policy Research Shultz Graduate Student Fellowship in Economic Policy.

If the call data does not include handoffs, the metric can be adapted. If only one tower is referenced per transaction, one could consider separate calls close in time. Specifically, the metric can be set to the number of times the same individual places a call from tower k, and another call from tower u within Δ minutes of each other.

Test

The procedure was tested on mobile phone data from Rwanda. I gauge its performance by predicting the locations of known towers, with leave-one-out sampling as if their location were unknown. I then compare the predicted location to the actual location.

The mean error in predicted location is 7.0 km, with a standard deviation of 12.3 km. Performance is heterogeneous: as shown in the following figure, the method is more precise in areas with higher tower density, such as cities.



This leave-one-out procedure can be used to estimate precision as a function of the characteristics of the missing tower.[1]

Limitations

The method can only locate missing towers within the convex hull of known towers, which is problematic if missing towers are on the periphery.

[1] Note also that if tower locations are missing systematically rather than randomly (e.g., towers in a sensitive part of the country are missing), then estimation of the error using leave-one-out estimation of known towers may be misleading.