Location Tracking using Cell Towers and WiFi

Asitava Sarkar Aryesh Dubey

15 May - 15 July

1 Introduction

Due to the presence of a large number of cell towers and WiFi routers these days along with their ever increasing numbers, they prove to be an effective resource for location tracking. Along with GPS, they can also be an important resource for location tracking with a fair amount of accuracy where their numbers are large. There are several methods for finding the location using cell towers, some common ones like triangulation, trialteration and some others like optimization which is described in this document.

2 Location using Cell Tower

Cell tower data is one of the most easily available data for location tracking and it can be easily detected using a mobile phone. Cell towers are uniquely identified using four codes.

- MCC - Mobile Country Code
- MNC - Mobile Network Code
- LAC - Location Area Code
- CID - Cell ID

For our testing, we fetched the data using 'Netmonitor' app which gives the value of MCC, MNC, LAC, CID and the Received Signal Strength. In order to find the location, we need the location of the cell towers as a reference. So, we need a large publicly available database which has the location (latitude and longitude) corresponding to the above data. So, the OpenCellID database is an extremely good database in this regard. They have a free publicly available API which can be used to find the location of the cell towers.

2.1 Triangulation

The first attempt was to use the methods which are commonly used for location tracking. Triangulation is the process of determining the location of a point by forming triangles to it from known points. Triangulation involves the measurement of angles instead of the measurement of the actual distances between two points and this makes it a less accurate method compared to Trialteration.

2.2 Trialteration

Trialteration is a more accurate method compared to triangulation. This involves calculating the distances between the two points. The points in this case are the coordinates of the location of the cell tower and the coordinates of the point which we want to find. We have the coordinates of the cell towers and the signal strengths available for our calculations. The distances can be calculated from the signal strengths using the following equation-

$$Strength(dB) = -113 + 10\gamma \log(r/R)$$

Here, γ is known as the Path Loss Exponent with its typical value being 4. 'R' denotes the Mean Cell Radius. The Mean Cell Radius is particularly difficult to calculate as its range can vary anywhere between 4-15 km typically depending upon the technology of the communication, the quality of the equipment used in cell towers, presence of obstacles in the path of the signals etc. and is mostly likely to induce inaccuracies in the calculation of location. 'r' denotes the distance of the mobile device from the cell tower. Once the distance is calculated, a set of equations are formed using the distance formula. the number of equations formed are equal to the number of cell tower data available. The formula used for the above was used from this link. However, some approximations were used since the difference in coordinates in this case will be small fractions and the approximated result is shown here:

$$\sqrt{(latitude[i]-x)^2+(longitude[i]-y)^2*cos(latitude[i])^2}*111.3=distance[i]$$

$$distance[i]=R*10^{(strength[i]+113)/40}$$

The values of $latitude[i]$, $longitude[i]$ and $strength[i]$ are known for each cell tower and the set of equations can be solved to get the desired results. However, there are limitations with this method since the value of signal strengths keeps fluctuating a lot due to the presence of obstacles. As a result, a significant error is introduced in the signal strength and as a result in the distance calculated leading to wrong results or sometimes even resulting in non-existence of any solution for the given set of equations. Therefore, there was a need to develop a better method for finding the location using this data.

2.3 Optimization

Optimization does not involve solving any type of equations but rather finding the solution which best fits our requirements. So, we are always guarenteed a solution for every situation which should be the case always. For optimization, we chose to minimize the difference between the distance between the coordinates of the cell tower and our desired location and the distance calculated using the signal strength. The function for optimization was as follows:

$$func[i]=\log(\sqrt{(latitude[i]-x)^2+(longitude[i]-y)^2*cos(latitude[i])^2}*111.3)-\log(R)-(dB+strength[i])/40$$

An additional third variable is also introduced along with dB in some cases to account for the fluctuating signal strengths. Similar functions are created for each set of towers.

2.3.1 When one tower is available

When only one tower is available, the coordinates returned by the OpenCellID API are treated as the location of the user as one tower can only give an area of our probable location and not the precise location.

2.3.2 When two towers are available

the functions available for optimization are:

$$func[1]=\log(\sqrt{(latitude[1]-x)^2+(longitude[1]-y)^2*cos(latitude[1])^2}*111.3)-\log(R)-(dB+strength[1])/40$$

$$func[2]=\log(\sqrt{(latitude[2]-x)^2+(longitude[2]-y)^2*cos(latitude[2])^2}*111.3)-\log(R)-(dB+strength[2])/40$$

The optimisation function is 'fmincon' of MATLAB which is a constrained optimisation algorithm, so the algorithm needs lower bound and upper bound values. Taking the logarithm of the expressions before optimization their difference also helps in reducing the error. The lower bound and upper bound values are set between ± 0.5 of the coordinates of the tower with the highest signal strength. (The tower with the highest signal strength will give the most accurate results, so is given the higher preference)

2.3.3 When three or more towers are available

When three or more tower data is available, it results in the most accurate results since a location can be identified if atleast three reference points are available. The optimization function for each tower is written, however, for using in the algorithm the functions are used in groups of 3 to minimize the error in the data. All the coordinates thus obtained by considering the functions in groups of three are then averaged out to obtain the final location coordinate. The implementation and constraints on the bound are the same as mentioned in the above section.

3 Location using WiFi and Cell Tower

WiFi can greatly improve the accuracy of the above method since the range of WiFi is limited to a few meteres. So, knowing the probable coordinates of the WiFi routers can limit the lower and upper bounds mentioned above to a narrow margin and thereby giving a better output. Since WiFi data is not that readily available in publicly available databases, the code takes care of the cases where WiFi data is available or not. Presence of WiFi can now limit the the range between ± 0.05 of the coordinates of the WiFi router with the highest strength. This along with the cell tower data will give the most accurate results.

4 Code

The MATLAB code can be found here `https://github.com/asitava1998/location` . There are files named app_cell, app_wifi and app with the codes for location finding using cell towers, WiFi and combination of both respectively. There is a test dataset available in the file location_data.xlxs and the code can be run simply by downloading the folder and directly running it in MATLAB. The file location_history.docx has the list of urls containing the code at various stages of our project and it will give an insight into how we arrived at the current code.

5 Conclusion

So, compared to the methods of Traingulation and Trialteration, the optimization methods yield better results for the dataset we tested for and also takes care of the situations where the other two algorithms fail to produce results. The effect of some common problems like fluctuating signal strengths, uncertain value of the Mean cell radius etc. are also minimized in the optimization algorithm. There are quite a few other optimization algorithms available in MATLAB like fminserach, fminunc etc. which are unconstrained optimization algorithms. Almost all the algorithms produce identical results when the data at a given point is sufficiently available, however, when sparse data is available the constrained optimization algorithms like fmincon yield better results.