Measuring cell plan quality

Distributed Systems Seminar 2015

Joosep Rõõmusaare

May 3, 2015

Institute of Computer Science, University of Tartu, Estonia

Abstract. Cell plans are used for attempting to position mobiles in a cellular network to their geographical location. Cellular location and each cell effective area are visualized with cell plans. These cell plans can be created with multiple methods. This paper examines which method would be best to measure qualities of different cell plans.

Keywords: cell plan, cellular network, visualizing, mobile positioning

Introduction

Researchers, who are trying to understand human-mobility patterns, need collected data about crowd movements. Collecting that data is costly, resulting in small sample sizes or infrequent data collection. Data collecting with cellular telephone networks would be cheaper and more frequent as billions of people keep a phone near them most of the time. Mobiles when connected to the cellular network are making events when they are used for SMS, calling, or Internet. Mobile operators are collecting these events. But that data has a significant limitation: collected location point is the cell tower location.[1]

When trying to visualize cellular networks on the map, each cell has effective area coverage. Visualized map of each cell effective area coverage is called cell plan. These cell plans are used for positioning mobiles to their geographical location. So a cell plan must be as accurate as possible. Accurate means that if person was standing in the cell effective area, then that person would be connected most likely to that cell. There are currently used different methods for generating cell plans, but there isn't any measurements for which method would give more accurate cell plan. This report creates methodology for measuring quality of different cell plans.

1 Description of the problem

1.1 Mobile network

Mobile networks are created from base transceiver stations (network sites), each of which having one or more cells. Stations and cells are located so it would give best possible radio coverage over all the network. Each mobile will connect to a cellular network by some specific cell. Each cell is identified with the cell global identity (CGI) and when a mobile phone creates events to the connected cell, then the mobile phones current cell CGI is recorded with the event.[2]

When trying to get some mobile phone location from the events log we get the time and the CGI, from the CGI we can get the cell and it's location. Cell effective coverage areas are needed, because mobiles aren't mostly next to the site location, but somewhere in the direction of the cell antenna azimuth from the site location.

1.2 Visualizing cell plan

Cell plan is used for visualizing mobile location and it is created from cells effective geographical area. That means when a user is in that cell area, that user will likely connect to that cell.

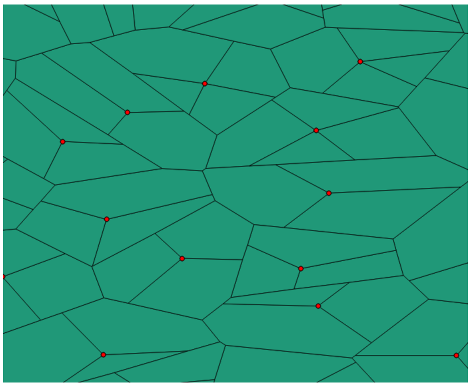

Figure 1: Voronoi cellplan, red dots are sites, green polygons each cell effective area.

Currently three different approaches are used for generating cell plans: Voronoi diagram, best service data or using receive signal strength indicator (RSSI) data.

- Cell plan with Voronoi diagram needs minimum data from the mobile operators. Each network cell effective area coverage is one Voronoi diagram cell. For Voronoi diagram sites locations and each cells antenna azimuth is used. Each cell is moved 70 meters from the site in the direction of it's azimuth. That new location will be the Voronoi diagram entry point. That creates sectors from site for each cell in correct direction.

- Mobile network best service data needs more specific data from mobile operators. With best service data each cell has coverage where it's RSSI is greater than any other cell, and that coverage is used as effective area for a given cell.

- RSSI datas difference from best server data is that the effective area for each cell is given where RSSI is stronger than a given decibel, which means it doesn't compare different cells. Because of that there may be effective area overlaps or areas, which aren't part of any effective cell area.

1.3 Cell plan quality measurements

Each method described in the last section can use various parameters, which would result in different cell plans. For understanding which cell plan would give the best outcome, there's a need for some quality measurements.

There are used three ideas for measuring quality in this paper:

- traffic flow prediction[3]. Trying to predict traffic flows using various cell plans. Finding which cell plan gives best results.

- "feederswap" algorithm[4]. Using "feederswap" algorithm for testing different cell plans. Comparing cell plan and people movement trajectories.

- using collected GPS points and CGI data to calculate different probability density functions (PDF) likelihood. [5].

2 Results

2.1 Traffic flow algorithm

Traffic flow algorithm was already done for another paper and it was tested how the results would vary if input cell plans have different quality.

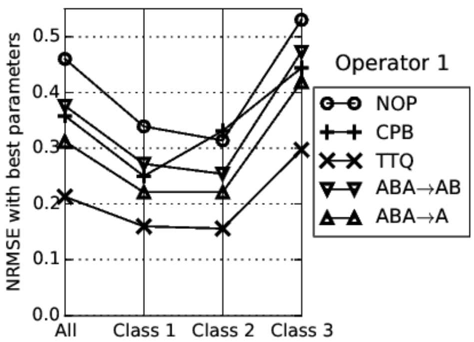

Figure 2: Traffic flow results with correct cell plan

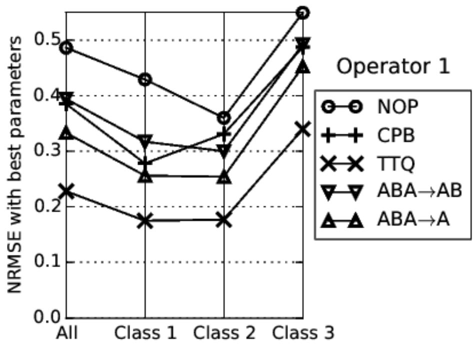

Figure 3: Traffic flow results with stretched cell plan

See Figure 2 and Figure 3 above. Operator means different ways how mobile events are taken into account. And classes are different road types. NRMSE means normalized mean square error. As it can be seen, with correct cell plan, NRMSE results were better than with cell plan, which was stretched out in South-North direction. This gives reason to believe that traffic flow algorithm can be used for measuring qualities of different cell plans.

2.2 "Feederswap" algorithm

When "feederswap" algorithm was tested with the same two different cell plans as traffic flow algorithm, it gave same results. Even if one cell plan should have been better for algorithm input. So in current "feederswap" algorithm state, it can't be very good measurement for cell plan qualities.

2.3 GPS and CGI likelihood

2.3.1 Collecting GPS data

OpenCellId. First place for GPS data collection was OpenCellId[6]. Crowdsourcing site, where everyone can upload their location and with which cell they were connected at that moment. From there we got 88305 measurements, but because most of the measurements were years old, we had to match CGI's from OpenCellId data with cell plan CGI's and got only 62437 measurements from which we could calculate likelihood.

There were also some cases where user GPS location was nowhere near the cell location. That means in time that cell CGI may have changed or moved to another location.

In Figure 4 there's an example of GPS measurements and to which cell they were connected. Different colors mean different cells. In this example different people were moving along the highway and where the lines end together, there is cell polygon. This view was created to get the idea how measurement point GPS may differ from connected cell location.

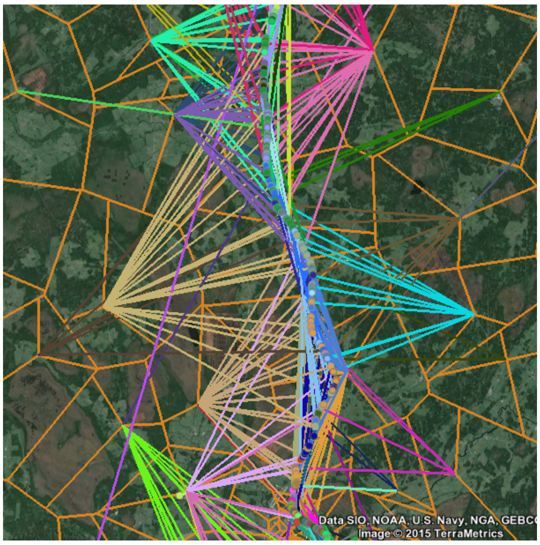

Figure 4: Example of GPS measurements and their connected cells

Collected GPS data. Another way used for collecting GPS and CGI data was using our own, family, or colleague mobiles to collect more fresh data. There were used two different datasets, but one of them still needs little preparing. Other dataset had GPS and CGI data from three persons (Table 1).

Events	User
84658	user1
23950	user2
5569	user3

Table 1: Collected GPS dataset

2.3.2 Calculating likelihoods

PDF likelihood values are used as quality measurements. Likelihood of PDF should be different with various cell plans when other input data and parameters are same. Higher likelihood value indicates that the PDF is more precise estimate to the real probability distribution that generated the measured data. Therefore better cell plan would have a higher likelihood value.

Resulting values were a logarithm of calculated likelihood L, $log\,L = log\,P(measurements|model) = log\prod_i P(x_i|C_i) = \sum_i log\,P(c_i|C_i)$, where m is mobile positioning events connected to the cell $C_i, i \in \{1, ..., m\}$, and these geographical locations of measurements $x_i, i \in \{1, ..., m\}$.

The measured data may contain outliers that deviate significantly from typical behavior. The likelihood calculation article applied outlier elimination procedure to the data: $[m \cdot q]$ data items with lowest $P(x_i|C_i)$will be excluded from the dataset. q is relatively small value and percentage of the typical distribution in the mixture is at least $1 - q$.[5]

Two different datasets were used: collected GPS and downloaded from OpenCellId.

Four different cell plans were used for testing this:

- Voronoi cell plan from January 2014. where cells were stretched little in North-South direction.
- Voronoi cell plan from January 2014, where North-South stretching was fixed.
- Voronoi cell plan from March 2015. That would be the most correct cell plan, because most measurements

were from March or later date. OpenCellId measurements were distributed over multiple years.

- cell plan created from RSSI data from March 2015. Also most consistent with the collected data.

Cell plan	Number of cells
IRSSI	9105
Voronoi	9066
Voronoi 2014.01	9087

Table 2: Cell plans

Cell plan	Collected	OpenCID
IRSSI	219	32091
Voronoi	219	32142
Voronoi 2014.01	184	35206

Table 3: Used measurements

In table 2 there is described each cell plan's cell count. For January 2014 Voronoi cell plan's number of cells is same for the correct and with stretched cells cell plans.

In table 3, there are counts how many measurements were kept for calculations after filtering. Firstly measurements which CGI value wasn't part of the cell plan were eliminated. Secondly only first measurements were taken from the sequence of measurements that happened when user was still connected to the same cell.

From PDF likelihood calculator few examples are shown in table 4. Name contains four parts:

1. boolean value for showing if bayesian was used in calculations.
2. blur ratio used for blurring each cell.
3. buffer ratio used for generated buffer around each cell.
4. cell plan identifier.

name	q0.05	q0.1	q0.25
0_1_-0.2_rssi		-191	-15
1_1_0_rssi	-263	-165	-33
0_0.5_0.2_rssi		-161	-37
1_0.5_0.5_rssi	-257	-167	-42
1_2_-0.2_rssi	-248	-170	-45
0_2_-0.2_rssi	-269	-212	-102
0_0.5_-0.2_voronoi	-315	-255	-122
1_0.5_-0.1_voronoi	-317	-257	-127

Table 4: cumulative likelihood

Values q0.05, q0.1 and q0.25 are cumulative logarithmic likelihood values if 0.05, 0.1 or 0.25 quantiles are described as outliers and were dropped from the data. Table 4 is generated with different likelihood generating parameters. There is a need to know which parameters would be best for trying to compare cell plans.

PDF quantile likelihood values are described in figure 5.

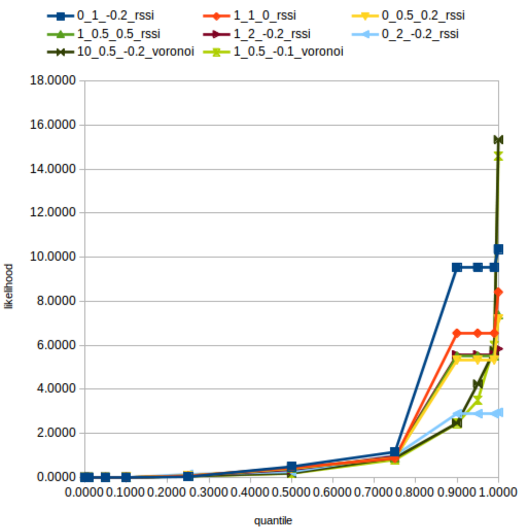

Figure 5: Likelihood quantile

It is seen that a lot of GPS measurements give a small likelihood for PDF and only few are measurements where PDF likelihood is higher.

There may be two different explanations:

1. there are small cells with only a few GPS measurements.

2. most measurements are done near the cell border and because of that likelihood values are small. Figure 4 supports this explanation because most of the measurements in there are far from the cell itself.

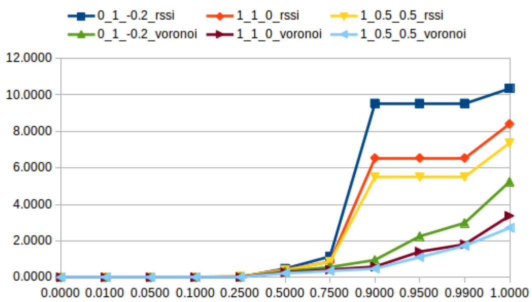

Figure 6: IRSSI and Voronoi cell plan comparison

In Figure 6 three different likelihood calculation configurations are used for comparing Voronoi and RSSI cell plans. All RSSI cell plan results have better likelihood values than Voronoi ones. From that we can conclude that likelihood could be used for comparing quality of cell plans.

3 Conclusions

Currently each cell plan algorithm was tested lightly and not much work was done with the quality measurements. Most tested algorithm was the PDF likelihood algorithm. It was also most configurable. And it has been concluded that PDF likelihood can be used for comparing cell plan qualities.

There's still need for further work to be done. Best method to work with should be PDF likelihood. More concrete configurations should be selected for that method. The algorithm should be tested with different cell plan configurations. Currently only one version of RSSI and Voronoi cell plans was used. More GPS measurements with different collectors should be used. Current three person data has a too small sample size.

PDF likelihood calculation itself could be made better. There's two possible research directions:

1. trying to take into account GIS layers (buildings, road networks, road surface). Hypothesis would be that GIS layers give better prior and likelihood would be higher.

2. separating points to moving and standing measurements. Hypothesis would be that stationary and moving action cell areas are different.

References

[1] R. Becker, R. Caceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky, "Human mobility characterization from cellular network data," 1 2013.

[2] M. Fillekes, "Reconstructing trajectories from sparse call detail records," Master's thesis, University of Zurich, 6 2014.

[3] T. Vajakas, J. Vajakas, and R. Lillemets, “Trajectory reconstruction from mobile positioning data using cell-to-cell travel time information.” 2 2015.

[4] T. Vajakas and J. Vajakas, “Feederswap algorithm.” 2015.

[5] T. Vajakas and J. Vajakas, “Estimation of spatial density from mobile positioning data using bayesian statistics.” 2015.

[6] “Opencellid.” http://opencellid.org/.