

Modelling and visualizing macroscopic movements

by

Ananya Chowdhury

Matriculation Number TODO

Piotr Mrówczyński

Matriculation Number 387521

Gabriel Vilén

Matriculation Number TODO

A project documentation submitted to

Technische Universität Berlin
School IV - Electrical Engineering and Computer Science
Department of Telecommunication Systems
Service-centric Networking

June 9, 2017

Supervised by:
Bianca Lüders
Friedhelm Victor
Prof. Dr. Axel Küpper

1 Introduction

1.1 Macroscopic City Movements - problem definition

Here we should describe what problem project is about to solve. In the 1st term presentation I mentioned that we want to answer two questions:

Given a location, where and in what proportion people move to another location - Movement Graph?

Where and how long do people stay at certain location? - but we need to be more precise here.

1.2 Human Mobility within the city

According to research [[HumanMobility1](#)][[HumanMobility2](#)][[HumanMobility3](#)], the average walking speed in the city for man aged 40 is 1.4 m/s (5 km/h), and slowest one is for elderly (with walking impairment) below 1 m/s (3.6 km/h).

1.3 Methodology of obtaining location data

Here we should describe how and from where do we get the data, and emphasise the fact that location updates are based on MOVEMENTS, so each next point is triggered by some mobile phone movement. Point updates are not triggered while being stationary.

TODO: we need here more details about what our Sudanian guy were supposed to research

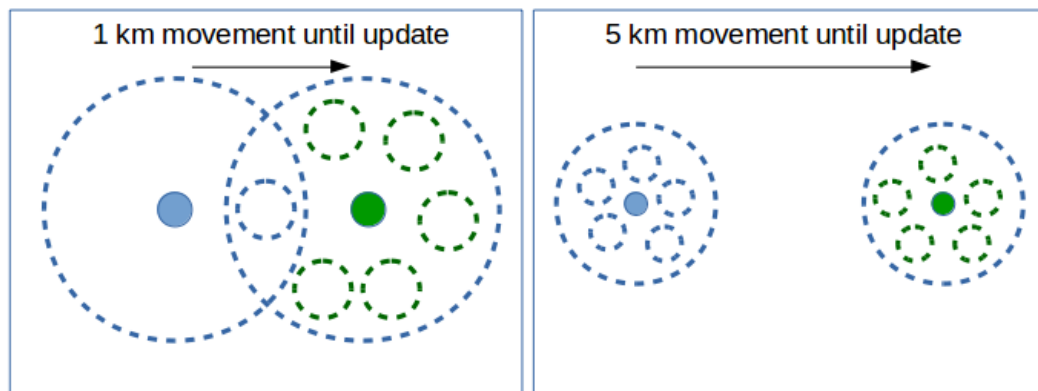


Figure 1.1: Continuous and Discontinuous location update after movement to section lacking area details

1.4 Location data characteristics for Berlin area

1.4.1 General overview

Movement triggered data from mobile devices are spanned across the Germany as shown on Figure 1.2.

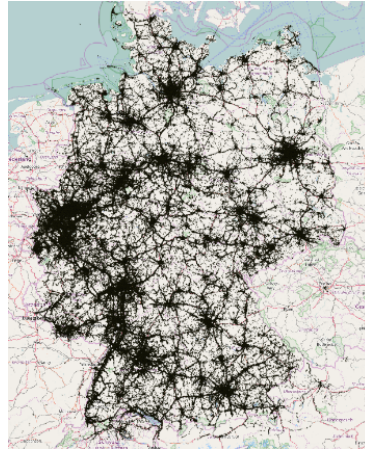


Figure 1.2: Visualisation of data points location in the geographical map of Germany

It points out, that most of the detected movements from one point to another are mostly gathered on the highways and inside the cities.

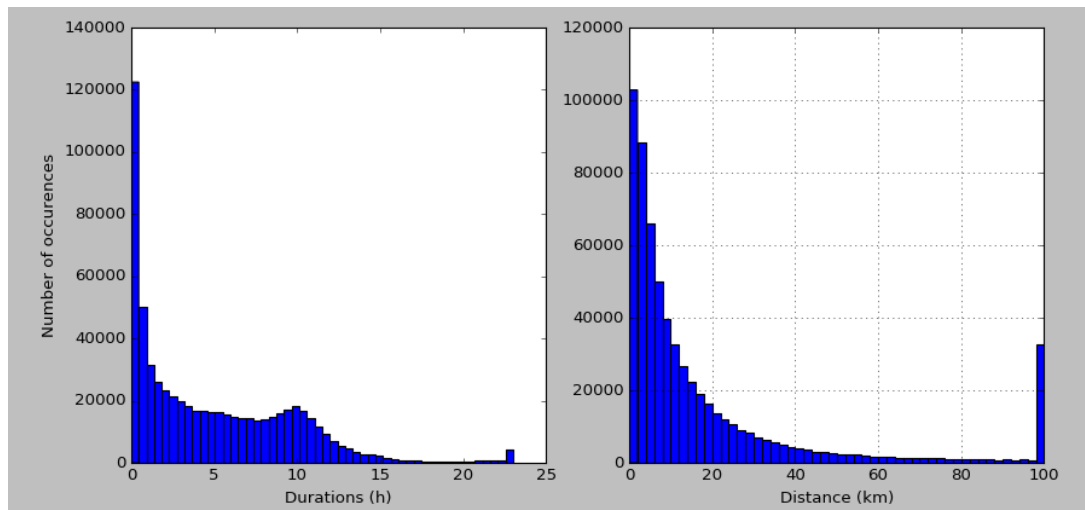


Figure 1.3: Histograms of durations and distances between consecutive points under and cumulatively over 100 km for the area of Germany

The given data batch has been gathered in the interval of 24h starting from 3am in the morning German time. It consists of over 675000 data points, belonging to 47300 unique mobile devices. Figure 1.3 shows, that for unique mobile devices, the durations between consecutive points in

time are most frequent within 30 minutes and most of the data is in the interval 1-10h. Regarding distances, consecutive points are most frequent below 2 km and most data is in the interval 0-10km. Big chunk of distances is also over 100km.



Figure 1.4: Heatmap of data point densities within the area of Berlin

Figure 1.4 shows the heatmap of data points for Berlin area. We can see that most of the movements detected are on major train/tram/metro stations, highways and major living areas.

1.4.2 Data analysis

Figure 1.5 shows that there is significant fraction of points which timestamp or distance did not change, thus duplicated point has been recorded.

TODO: This needs to be investigated

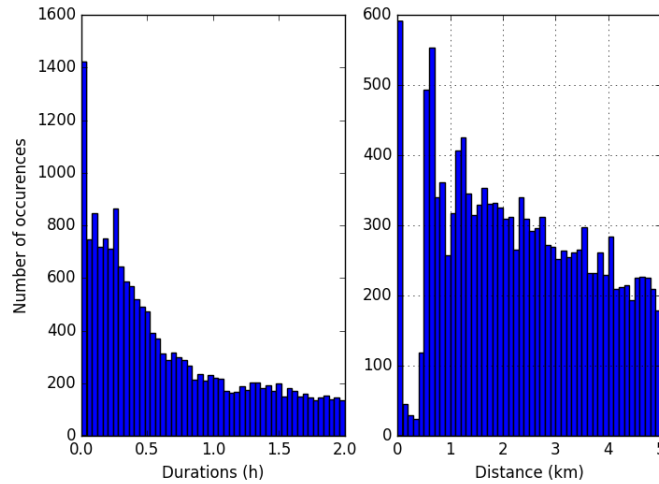


Figure 1.5: Histograms of durations and distances between consecutive points under 5 km and under durations of 2h for the area of Berlin

Furthermore, majority of durations between points is within interval of 30 minutes. For distances between points, there is significant "jump" at 500-1000m and 1100-1500m, which might mean that at these intervals, continuous points been gathered, and distances over that values might be discontinuous (ref. Figure 1.1). It is due to the fact that these distance intervals are most frequent and that might suggest that this is an average "update" distance for the moving mobile devices, as described in section 1.3. Less frequent values might suggest that updates were obtained with lower accuracy (e.g. by presence inside the building, metro line or simply mobile device lag in obtaining its location) and are result of discontinuity between consecutive updates.

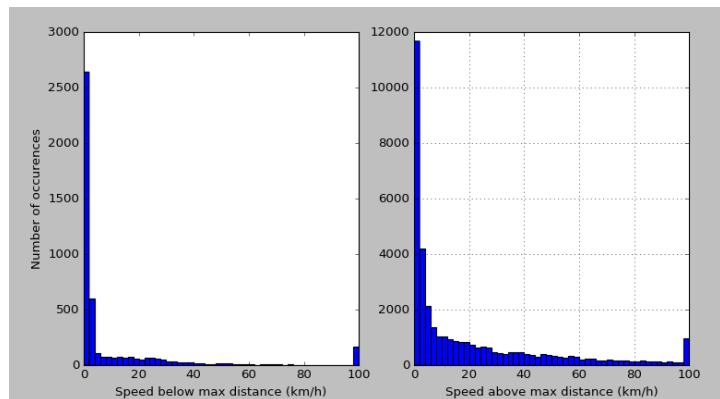


Figure 1.6: Histogram of speed between points with corresponding distance above or below 1.5 km

Figure 1.6 shows that considering the registered distances within range of 1500m, the vast majority of points are between 0-2 km/h, and also significant interval at 2-6 km/h. The rest of the values is sparse distributed in interval 6-100+ km/h. In case of registered point above range of 1500m, we observe that indeed most frequent occurrence is at interval of 0-4 km/h, however most are sparse distributed above 4 km/h, with most points being in interval of 4-20 km/h.

1.5 Approaches to stop detection for localization services

Here we should explain different approaches

1.5.1 Stop Detection analyzing repetitive appearance at location

1.5.2 Stop Detection based on continuous localization

1.5.3 Stop Detection based on mobility index

In the paper [**MobilityIndexGIS**], movement of mobile devices between cells (handovers) is considered. To detect the periods of slow movements or stops at the specific location, parameter called Mobility Index is considered.

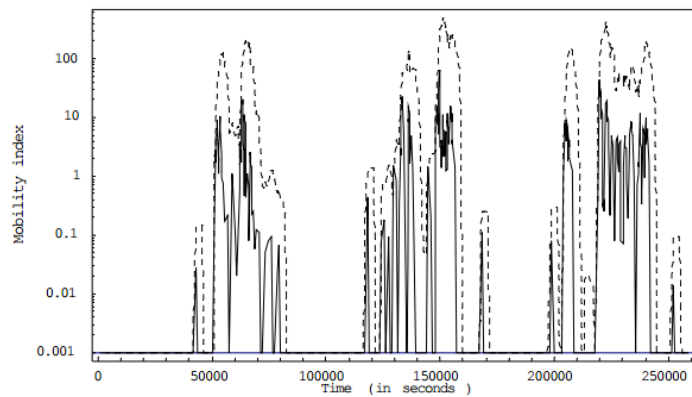


Figure 1.7: Mobility Index. Source: [**MobilityIndexGIS**]

In a populated area, with many GSM cells, the mobile terminal can change from one cell to another within seconds or after several minutes. Given a set of consecutive records, the mobility of a user can be estimated by calculating the Mobility Index over a pre-defined time period (sliding window). The Mobility Index is the sum of the distance between each record and the previous ones, where the distance is the inverse of the time spent on each cell. If the value of MI is below certain threshold, it is assumed that the object stopped or moved slowly.

In the publication, sliding window of 10 minutes and a mobility index threshold value of 6 has been used for certain area. The obtained results were in agreement with actual movements during more than 90% of the time.

2 Concept and Design

We need to describe here what do we want to achieve, what do we mean by modelling and visualizing macroscopic movements etc - thus general architecture that we have 3 components, Stop Detection from telekom data based on movements and updates of "areas" in your mobile phone, Clustering of Stops to find most popular "stops" in the city, Graph Generation to find where do the people move and in what amount

2.1 Stop Detection

Analyzing different approaches for stop detection based on localization data (ref. section 1.5), we decided to base our solution on standard human behavior regarding mobility within the cities (ref. section 1.2) and reusing the concept of Mobility Index (ref. subsection 1.5.3)).

Stop detection is divided into two phases:

Identifying stop candidates according to stop detection algorithm

Reconciliation of stop candidates taking into account mobility indexes of points.

2.1.1 Identifying stop candidates

Stop algorithm, shown on Figure 2.1, is considering 3 parameters - MinWalkSpeed, MinTransportSpeed, ThresholdDistance.

So called ThresholdDistance is distance mentioned in subsection 1.4.2 (which in case of Berlin been in range 800-1500m) as distances between points, where there is significant "jump" in frequency of occurrences, which might mean that at these intervals, continuous points been gathered, and distances over that values might be discontinuous. Furthermore, these distances are relatively small compared to other observed distances, allowing more accurate decision process. We assume, that at these distances, no significant "interruption" in gathering the points has been introduced in the form of buildings, metro or other factors.

Thus, for these higher accuracy points one could assume, that if mobile device have been moving within the minimum walking speed MinWalkSpeed (ref. section 1.2), this point is identified as stop candidate.

Thus, for lower accuracy points one could assume, that if mobile device have been moving within the minimum transport speed MinTransportSpeed (ref. section 1.2), this point is as well identified as stop candidate. In this example, it could mean, that over longer distance, one

could move with transportation with significant speed, stop for longer period and then use fast transportation again, resulting in higher than minimum walking speed, but with speed which is in average lower than other psychological value as minimum transport speed or maximum walking speed.

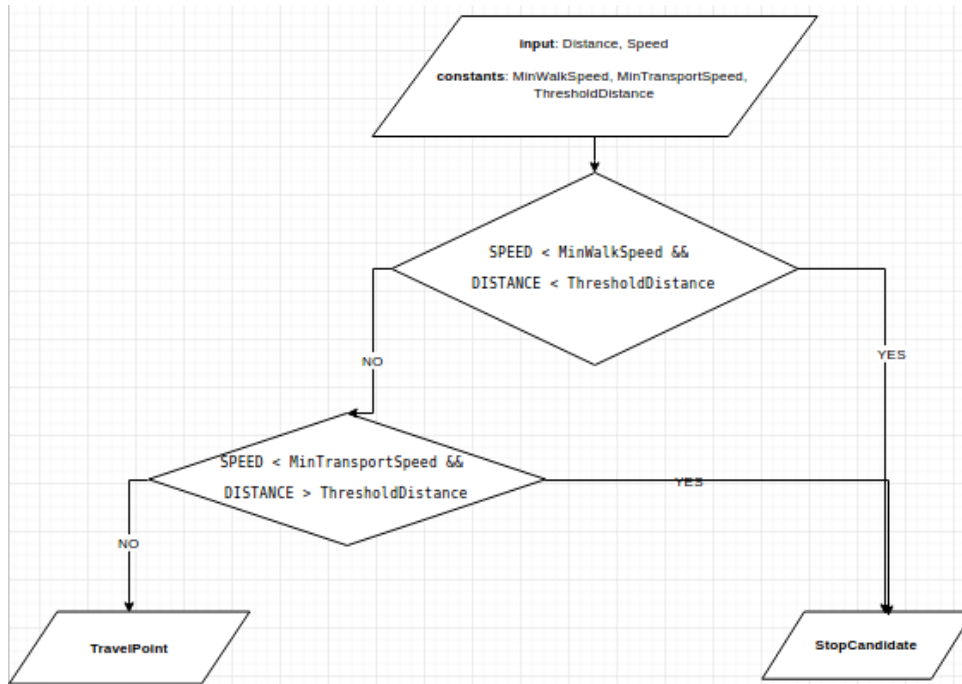


Figure 2.1: First phase of stop detection algorithm

2.1.2 Reconciliation of stop candidates

Figure 2.2 shows that there are many scenarios in which stop candidate has been detected, taking into account movement speed between points, distance and mobility index. In the above example, the mobility index windows was considering past 60 minutes from each of the points.

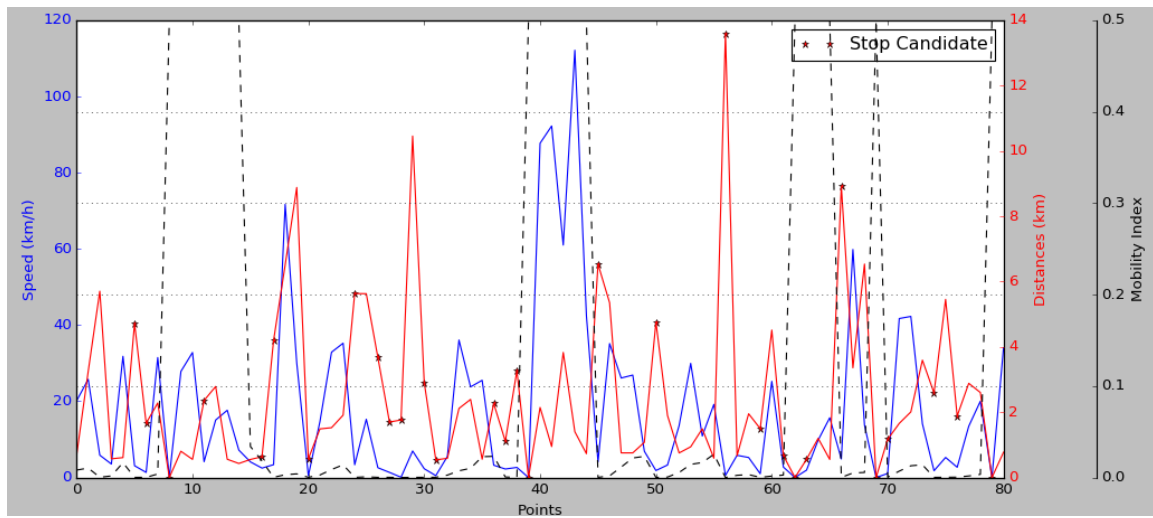


Figure 2.2: Result of identifying stop candidates for single user with their corresponding mobile index, distance and speed between the detected and previous point

2.2 Clustering of Stops

2.3 City Movements Graph

3 Implementation

We need to describe here some details about the implementation, project structure, configuration of the environment, how to use it, important implementation aspects like why did we choose these parameters not the others etc and how did we implement Graphical User Interface.

4 Evaluation

We need to describe here some details about how did we evaluate that our graph is correct e.g. show-casing screens from user interface or QGIS etc. Then we need to show how did we test our code, if it scales, and maybe some performance measurements

