Experiment (1): Extracting Stay points (preprocessing)

Introduction:

The human mobility patterns are considered as spatio-temporal data, which change in both time and space (location).

We are studying the human mobility patterns in daily bases, this work is aimed to be able to infer and predict these daily behaviors in order to advance the location applications and generate new apps which can be customized for every single user using his/her location data. This work can coup with other human patterns like behavior and usage.

Problem definition:

The target of this experiment is to discretize the GPS readings into stay points, which are the regions where the user spent time more than a threshold, and within area of certain geographical diameter, this task as a preprocessing for applying prediction algorithms to predict the users next locations based on the user regular patterns also recommendation engines can be used to predict new places for the user.

Problem description:

Data of the users in the Geolife dataset are logged every 5 to 10 seconds, recording a continuous stream of GPS readings, the data is distributed into files, every file is a trajectory of a day categorized by the user number. Figure (1) shows a sample trajectory.

By applying the below algorithm developed by Yu Zheng et. Al.(1) we can get sequences of discrete regions the user visited per day. Figure (2) shows the results of the trajectory showed in figure (1).

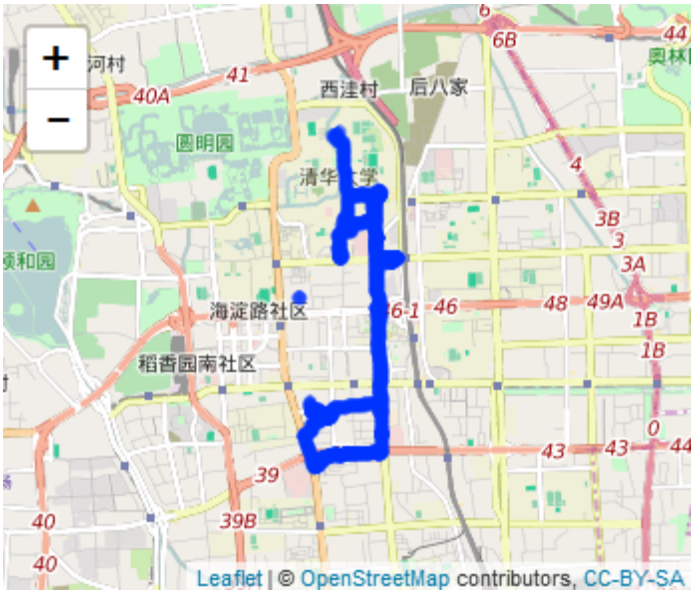

Figure (1) A map displaying the 5th trajectory of user 4 from Geolife dataset before extraction

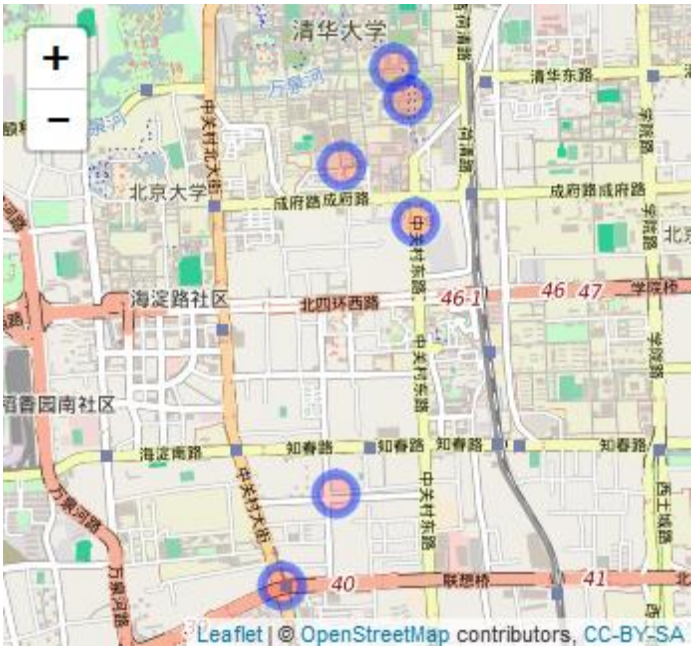

Figure (2) A map displaying Stay points extracted from the 5th trajectory of user 4 from Geolife dataset

Algorithm

Inputs:

- GPS records (longitude, latitude, time, date) from Geolife dataset, which contains 182 users collected in 4 years.
- Time threshold T_{th}
- Space threshold S_{th}

Outputs:

- List of Stay points (SP) with each arrival time T_{arrive} and leave time T_{leave}

Assumption:

- i , j : counters to loop over the GPS points.
- N: number of pints in the trajectory.
- p: The GPS point.

Steps:

While i<N do
J=i+1
While j<N do
Distance=Geodistance (p_i , p_j)
// geographical distance between two points
If (Distance >S_{th})
$\Delta T = p_{j:time} - p_{i:time}$ //time between two points
If ($\Delta T > T_{th}$)
$SP(k)_{lat,long}$= mean $(p_i , p_j)_{lat,long}$
$T(k)_{arrive} = p_{i,time}$
$T(k)_{leave} = p_{j,time}$
i=j
break
j=j+1
end while
end while
return SP

Results:

Applying this algorithm for the 182 users in Geolife dataset using time threshold of 20 minutes and space threshold of 200 meters we extracted 19,042 stay point as in the table (1):

Users	182
Time threshold	20 minute
Space threshold	200 meter
Number of GPS points	24,211,140 point
Number of extracted stay points	19,042 stay point

Table (1)

References:

[1] Y Zheng, L Zhang, Z Ma, X Xie, WY Ma "Recommending friends and locations based on individual location history", ACM Transactions on the Web (TWEB), 2011