Hourly Pedestrian Population Trends Estimation using Location Data from Smartphones Dealing with Temporal and Spatial Sparsity [1]

Presented By: Ziyang Chen

Fudan University

13307130148@fudan.edu.cn

May 18, 2015

[1] powered by pandoc and XeLaTeX

Outline

Abstract

- Describes a pedestrian population trend estimation method using location data of smartphone users.
- Intended to be an alternative to traffic censuses using tally counters.
- Using smartphone users' location data accumulated on Yahoo! Japan.
- Tackles the problem of data shortage when a target area is a small region by using a Gaussian kernel.

Background

- Knowing the number of pedestrians in a time or place can be an essential data source for market research.
- Tridional method: Traffic censuses using tally counters are still commonly used.
  1. Requires many survey crews and much time.
  2. Temporal and spatial limitation.
- Location-based services (LBSs) are widely used by smartphone users
- two problems remain
  1. How to pick out only pedestrians.
  2. How to set the research area and period.

Outline

Data Set

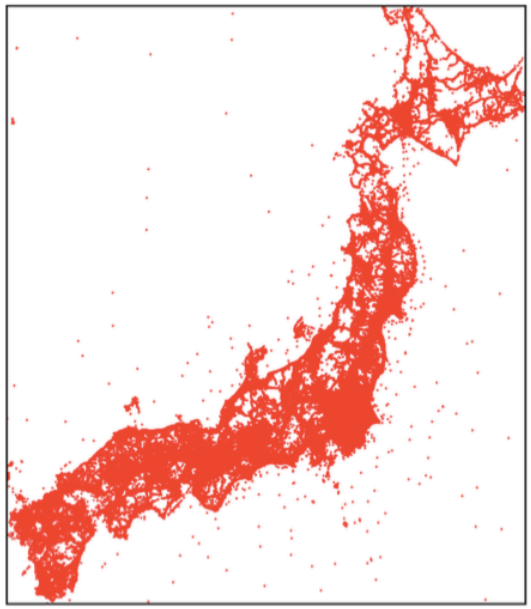
- The location data are provided by smartphone users through services of Yahoo! Japan, contains:
  - Latitude
  - Longitude
  - Horizontal accuracy
  - Time stamp (at a second rate)
  - Anonymized user ID (changes after 24 hours)

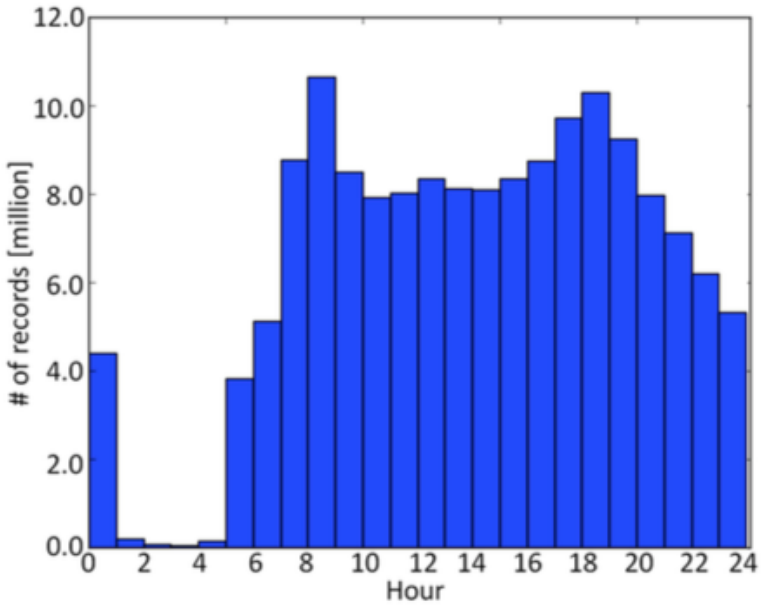
Data Set

Latitude	Longitude	Accuracy	Time stamp	ID
34.8716480098989	135.661309193946	5.00	20130611000000	UID1
38.7213293603050	139.849629867952	30.00	20130611000000	UID2
43.0928880757489	141.371409950081	112.00	20130611000001	UID3
35.5574600559872	139.445981733805	14.88	20130611000000	UID4
35.7329128494678	139.670771645082	65.00	20130611000012	UID5
35.7846891648607	139.899813949837	10.53	20130611000001	UID6
35.6521350751540	140.026954640171	1414.00	20130611000002	UID7
35.6998792435734	139.841407947290	165.00	20130611000002	UID8

Table: Examples of the location data

Data Set



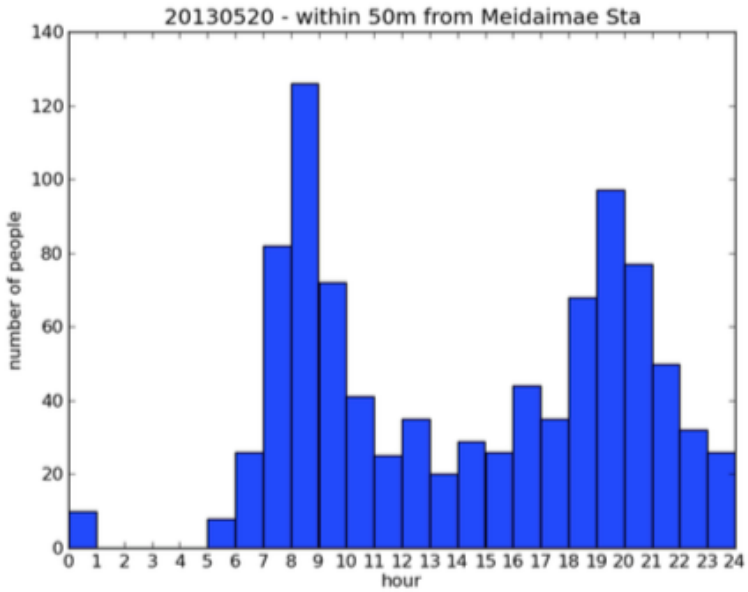


Outline

A simple approach

- Assumes that the number of records in an area is proportional to the people present in the area.
  - Determine target areas(polygonal) and target days
  - Number of records existing in the area hourly is counted and then multiplied by a proper factor.
  - The date has errors more than 300 meters are eliminated.

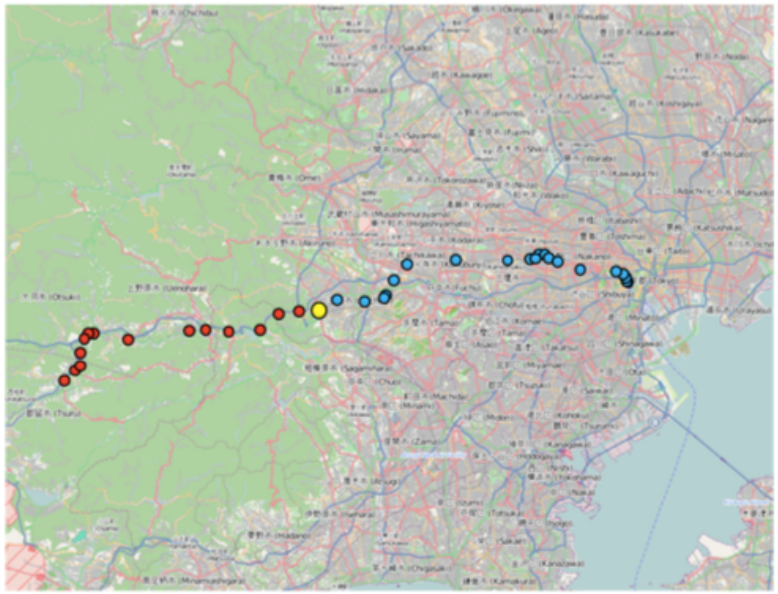
A simple approach



Limitation of the simple approach

- Non-pedestrian records
  - stationary users (e.g. in offices)
  - passing users (e.g. on trains)
- Time continuous estimation
- Sparsity with the smallness of target areas.

Eliminating non-pedestrian data



- Data are extracted user-by-user(by user id).
- Approximate mean velocity is estimated an hour before or after a person visits the target area.
- The records where velocities both before and after are high or near zero are eliminated.

Q&A

Any Questions?

End

Thanks For Attention!