Project Report 1

CSC 591-601 Social Computing

Craig Kimpel – CKIMPAL
Qiang Zhang - QZHANG15

Description

This project will investigate a hypothesis related to human mobility patterns, to see if it holds at different levels of local abstraction.

Hypothesis

"By measuring the entropy of each individual's trajectory, we find a 93% potential predictability in user mobility across the whole user base. Despite the significant differences in the travel patterns, we find a remarkable lack of variability in predictability, which is largely independent of the distance users cover on a regular basis." (Chaoming Song, 2010).

Entropy Calculations

Random (each location is visited with equal probability):
$S_i^{rand} = \log_2 N_i$ Where i denotes the person and N denotes number of distinct locations they visited.

Temporal Uncorrelated (dependent on frequency of distribution):
$S_i^{unc} = -\sum_{j=1}^{N_i} p_i(j) \log_2 P_i(j)$ Where $P_i(j)$ denotes the probability that user i visited location j.

Actual (dependent on both frequency of distribution and the order):
$S_i = -\Sigma_{T'_i \subset T_i} P(T'_i) \log_2[P(T'_i)]$ Where $P(T'_i)$ is the probability of finding a particular time ordered sequence T'_i in the trajectory T_i. T_i is the set of nodes that were visited in the order that they were visited.

Duration of Tracking data

Per (Chaoming Song, 2010) we will to remove any user data whose tracking data is unknown for >= 80% of the studied time.

Predictability

Per (Chaoming Song, 2010) we will use Fano's inequality (Fano, 1961) to measure the degree that an appropriate predictive algorithm can predict correctly the user's future whereabouts.
$S = H(P^{max}) + (1 - P^{max}) \log_2(N - 1)$
With the binary entropy function $H(P^{max}) = -P^{max} \log_2(P^{max}) - (1 - P^{max}) \log_2(1 - P^{max})$
S is the Entropy calculated above.
N is the number of locations moved between.
P^{max} is the maximum predictability as determined by (Fano, 1961)

Levels of Abstraction

This project will study the following levels of abstraction:

Geographical Coordinates – Coordinates directly from the data steam.

Stay Points – These will be calculated by processing each users geo data with the stay point algorithm from (Xiangye Xiao, 2010).

$$P = \langle p_m, p_{m+1}, \ldots, p_n \rangle \text{, where } \forall m < i \leq n \text{, } Dist(p_m, p_i) \leq \theta_d \text{, } Dist(p_m, p_{n+1}) > \theta_d \text{ and } Int(p_m, p_n) > \theta_d$$

$s.x = \Sigma^{n}_{i=m}\, p_i.\, x \,/ |P|$ average x over the range of points in P

$s.y = \Sigma^{n}_{i=m}\, p_i.\, y \,/ |P|$ average y over the range of points in P

User Defined Places – Labels of places that are defined in the dataset or are derived from a separate dataset that can be tied to the lat/lon coordinates in the dataset.

Point Matching

GPS Lat/Lon measurements are only accurate to within about 10 meters (30 feet) of the recorded GPS measurement. For purposes of matching points a delta of ±10 meters of the point will be used. If any part of the squares around two points overlap, then both points will be considered to be the exact same location.

Dataset

The dataset was collected by Yonsei University.
They deployed their mobility monitoring system, named LifeMap, to collect mobility data over two months in Seoul, Korea. (Yohan Chon, 2011)

The collection application allowed collectors to put labels on places.

The dataset was anonymized, however the MAC addresses with the same anonymous address were collected by the same person.

Method

Test the hypothesis identified above, to determine if they apply to each local abstraction that is identified above.

For stay point calculation a delta d will be chosen after analyzing the data to determine what an appropriate delta d should be.

Bibliography

Chaoming Song, Z. Q.-L. (2010, February 19). Limits of Predictability in Human Mobility. *http://www.sciencemag.org*, pp. 1018-1021.

Fano, R. M. (1961). *Transmission of Information.* New York and London: the MIT Press.

Xiangye Xiao, Y. Z. (2010). Finding similar users using category-based location history. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10)*, 442 - 446.

Yohan Chon, E. T. (2011). Mobility prediction-based smartphone energy optimization for everyday location monitoring. *The 9th ACM Conference on Embedded Networked Sensor Systems (SenSys '11)* (pp. 82-95). Newyork City: ACM.