Modeling Human Mobility Using Location Based Social Networks

Gene Moo Lee
University of Texas at Austin
1 University Station D5900
Austin TX, 78712
gene@cs.utexas.edu

SUMMARY

Location based social networks (LBSNs) have gained tremendous popularity as people have embraced smart devices at a remarkable rate. This gives us a unique opportunity to understand the human mobility with *check-in* information, which specifies the location of a user at a timestamp. Using a gravity model, this abstract analyzes human traffic matrices, which measure the number of users moving across different locations.

I. INTRODUCTION

We collect the data of two cities (Austin, TX and New York, NY) from two LBSNs (Foursquare and Gowalla). Foursquare data was collected from January 2012 to February 2012. We have 13,718 users, 836 locations, and 51,638 check-ins for Austin and 115,605 users, 5,824 locations, and 481,976 check-ins for New York. Gowalla data was collected from May 2010 to July 2010. We have 6,379 users, 8,022 places, and 355,747 check-ins for Austin and 6,462 users, 8,623 places and 154,701 check-ins for New York.

II. HUMAN TRAFFIC MATRIX

We want to measure the aggregate human movements across different locations. We introduce human traffic matrix (TM) to quantify the number of users between different locations. Each traffic matrix element *T(a,b;t)* is the total number of users moving from a location *a* to another location *b* during a day *t*. We aggregate individual locations into 0.01° x 0.01° grids, which range about 1.1 km x 0.7 km.

III. GRAVITY MODEL

We model the human TM using a gravity model:

$$T(a,b) = \alpha\, C(a)\, C(b) / d(a,b)^\beta$$

where *T(a,b)* is the TM element from location *a* to location *b*, *C(a)* and *C(b)* is the number of check-ins at *a* and *b*, respectively, *d(a,b)* is the distance between *a* and *b*, α is a constant scaling factor, and β is a constant distance exponent.

We use a grid search from 0 to 2 to find the best exponent β that minimizes the estimation error. Given a TM and β, we can form a linear system $\alpha E = T$, where *E* is the vector of estimated TM elements by the gravity model and *T* is the vector of actual TM elements. We find the optimal α and β using the least square fit that minimizes the error.

Figure 1 shows the modeling results comparing estimated values and the actual ones. We see very strong correlation coefficients (CORR) of 0.96~0.99 for both cities and both data sets. The error (NMAE) ranges from 16% to 34%. This indicates that the gravity model accurately fits the human TMs. We observe relatively consistent β values, ranging 0.0~0.3 in Foursquare and 0.4~0.6 in Gowalla. A non-zero β indicates that distance has impact on the human movements. However, the constant α values vary due to diverse properties of the cities.

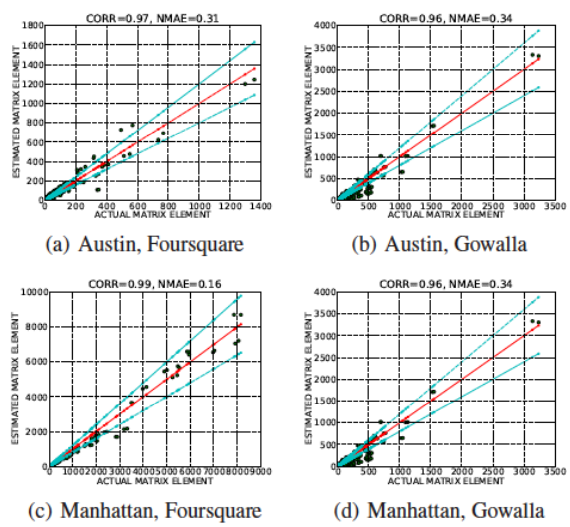

Figure 1: Comparison between the actual TMs and estimated TMs based on gravity model. For references, we have three lines: y = x, y = 0.8x, y = 1.2x.