Ranking the City: The Role of Location-Based Social Networks in Collective Human Mobility Prediction

O. R. Abbasi, A. A. Alesheikh

K. N. Toosi University of Technology, Department of GIS Engineering, Faculty of Geodesy and Geomatics Engineering, Tehran, Iran.

Extended Abstract

Technological advances have led to increasing development of data sources. Since the introduction of social networks, numerous researches on the relationships between users and their behaviors have been conducted. Location-based social networks (LBSN) have provided the possibility of studying the relationships between users and places. Today, due to the wide availability of various spatial data sources, the long-standing field of collective human mobility prediction has been revived and some new models (e.g. population-weighted opportunities, radiation, and rank-based models) have been introduced. There are two major assumptions in modeling the human mobility patterns. Some models (e.g. gravity model) assume that trips are directly related to the distances between origins and destinations. In other words, the more the distance between an origin and a destination, the lower the probability of traveling from the origin to that destination (Zipf 1946). However, some models explain the human mobility using "*opportunities*" concept. These models assume that trips are not directly related to distance, but induced by opportunities provided at destination (Stoufer 1940). Recently, a parameterized model of predicting human mobility in cities, known as rank-based model, was introduced (Noulas et al. 2012). The model predicts the flow from an origin toward a destination using "*rank*" concept. In fact, each destination has a rank, with respect to the origin, that expresses the probability of going from a region to another. However, the question that "*how the rank should be computed?*" is not well answered yet. In this paper, we explore the potential of LBSN data alongside the rank-based model in predicting human mobility patterns in Manhattan, New York City.

LBS 2016
13th Conference on Location-Based Services
Vienna, 14–16 November 2016

Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.

According to the rank-based model, given a set of zones $u \in U$ in a city, the probability of moving from zone $u \in U$ to a zone $v \in U$ is defined as (Noulas et al. 2012)

$$P[u \rightarrow v] \propto \frac{1}{rank_u(v)^{\gamma}} \qquad \text{(Eq. 1)}$$

where $rank_u(v)$ is the rank of zone v relative to zone u and γ is an adjustable parameter. Assuming that the total number of trips generated in each zone T_u is known, trip distribution matrix can be computed as (Yan et al. 2014)

$$T_{uv} = T_u \frac{rank_u(v)^{-\gamma}}{\sum_{k \neq u}^{N} rank_u(v)^{-\gamma}} \qquad \text{(Eq. 2)}$$

where N is the total number of zones in the city.

Rank-based model, in contrast to the well-known gravity model, uses rank-distance rather than spatial distance. Because people and their behaviour, as the most important components in mobility, are neglected in this approach (i.e. using distance alone to rank the zones), the resulted mobility patterns always remain unchanged. In this paper, we consider three methods to compute the ranks in the city using rank-distance and LBSN data. These methods include: 1) computing the rank using rank-distance; 2) considering the rank as the number of venues located in a circle centered at the destination with a radius equal to the distance between the origin and destination ; 3) using a check-in weighted rank concept in the model. Since the rank-based model is parameterized, a repetitive procedure is needed to determine the adjustable parameter. In this paper, the method introduced by Hyman (1969) was employed to minimize the difference between the real average travel distance and modeled average travel distance. Moreover, since the rank-based model presented in (*Eq. 1*) does not guarantee the equality of real and predicted attracted trips, a balancing process called Furness is applied on the matrix. In this paper we considered Manhattan's census tracts as trip zones. The zones and distribution of venues are shown in *Figure 1*.

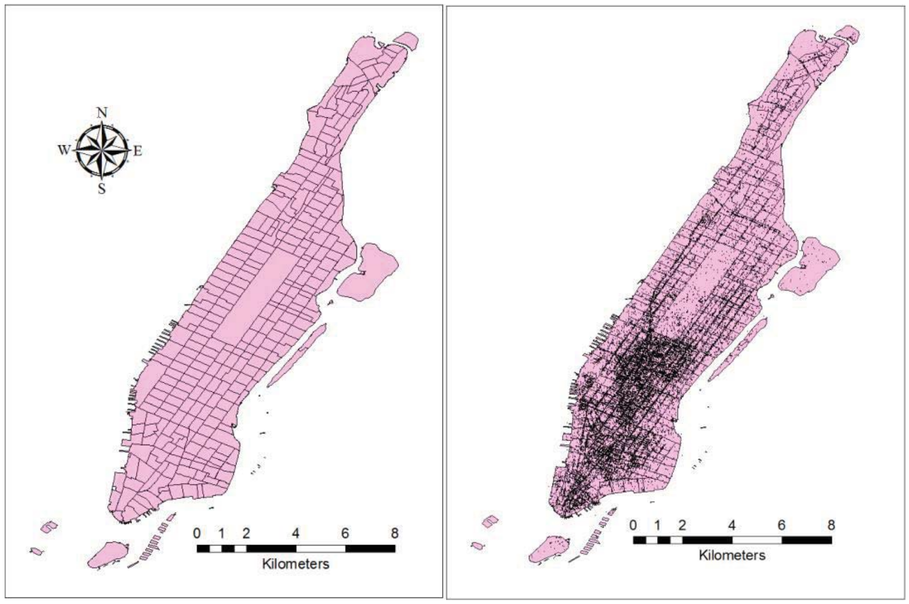

Figure 1. Distribution of venues (right) and census tracts (left) in Manhattan.

In order to evaluate the results obtained from the model and compare the rank concepts, GPS traces of taxi vehicles over Manhattan were used. Following the work of Kang et al. (2015), the Sorensen Similarity Index (SSI) was used as a measure of similarity between real and predicted trip distribution matrices. This index ranges from 0 to 1 where numbers closer to 1 indicate more similarity between two matrices. *Figure 2* presents a comparison among the performances of ranks in the model based on SSI.

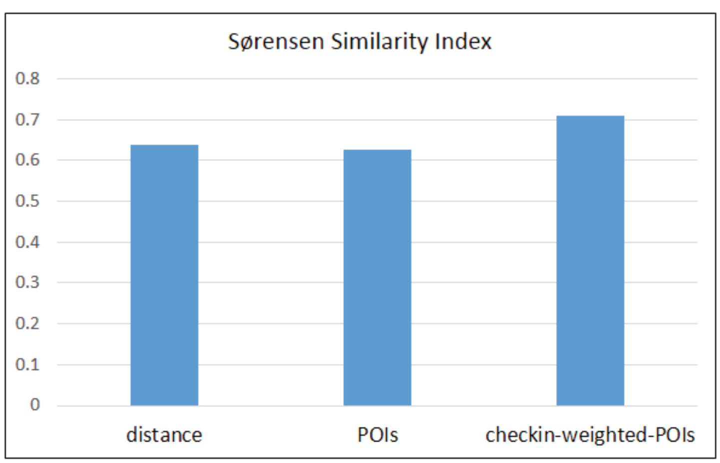

Figure 2. Evaluation of resulted trip distribution matrix based on Sorensen Similarity Index

Figure 2 indicates that using a check-in-weighted rank will result in slightly more similar predictions to reality. In order to have a statistical measure of how close the predicted data are to the real data, we determined R^2 value form regression analysis. The identity line is considered as the ideal case, where all the predicted trips are equal to real trips. *Figure 3* compares the results in terms of R^2.

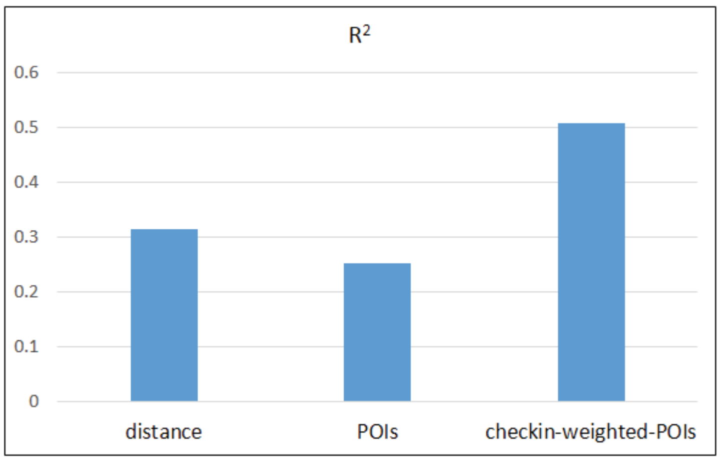

Figure 3. Evaluation of resulted trip distribution matrix based on R-squared

As *Figure 3* illustrates, the value of R^2 for rank-based model along with a check-in-weighted rank concept has been increased. So, check-ins play a significant role in improving the predictability of mobility patterns.

Considering the rank as variables such as distance and number of POIs is somewhat objective in a way that they are not representing real specific conditions of a city. In other words, the concepts of distance and number of POIs are the same for all cities in the world. But, using check-in weights, a real dimension will be added to the model. Surely, the role of a crowded park in human mobility is not the same as a hotel, for example. Thus, applying check-ins occurred at each POI will result in closer predictions to reality.

Moreover, the dynamicity of human mobility could be accounted for. As the check-ins data are dynamic, they can consider the variations in people's interests and behaviors. Using check-ins, any change in land use of POIs is also accountable.

References

Kang, C., Liu, Y., Guo, D., & Qin, K. (2015). A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint. *PloS one, 10*(11), e0143500.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: universal patterns in human urban mobility. *PloS one, 7*(5), e37027.

Stouffer, S. A. (1940). Intervening opportunities: a theory relating mobility and distance. *American sociological review, 5*(6), 845-867.

Yan, X.-Y., Zhao, C., Fan, Y., Di, Z., & Wang, W.-X. (2014). Universal predictability of mobility patterns in cities. *Journal of The Royal Society Interface, 11*(100), 20140834.

Zipf, G. K. (1946). The P 1 P 2/D hypothesis: on the intercity movement of persons. *American sociological review, 11*(6), 677-686.