Understanding Human Movement Semantics: A Point of Interest Based Approach*

Ionut Trestian†, Kévin Huguenin‡, Ling Su†, and Aleksandar Kuzmanovic†
†Northwestern University, Chicago, IL | ‡EPFL, Lausanne, Switzerland
{ionut, akuzma}@northwestern.edu,kevin.huguenin@epfl.ch
lingsu2012@u.northwestern.edu

ABSTRACT

The recent availability of human mobility traces has driven a new wave of research – on human movement – with straightforward applications in wireless/cellular network algorithmic problems. In this paper we revisit the human mobility problem with new assumptions. We believe that human movement is not independent of the surrounding locations, *i.e.* the points of interest that they visit; most of the time people travel with specific goals in mind, visit specific points of interest, and frequently revisit favorite places. Using GPS mobility traces of a large number of users located across two distinct geographical locations we study the correlation between people's trajectories and the differently spread points of interest nearby.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications - Data mining; H.1.2 [**Information Systems**]: User/Machine Systems - Human information processing

General Terms: Measurement, Analysis

Keywords: Mobility, Point of interest, GPS trajectories, Trajectory analysis

1. INTRODUCTION

Human movement has been studied extensively in the past several years. This is mainly due to the rather obvious implications that results might have in fields such as urban planning, disease prevention, mobile advertising, mobile infrastructure placement, *etc.* As such, researchers have tried to understand statistical properties, and also proposed mathematical models that try to capture particular aspects of human movement [3], for example by relating to the movements of banknotes [1] or subatomic particles. But the matter of fact is that, contrary to banknotes and subatomic particles, human mobility is driven by concrete goals such as buying groceries, dining at a restaurant, watching a movie, *etc.* And although it is indeed true that models (*e.g.* Lévy flights) manage to capture important aspects of movement (such as the distribution of trip sizes), they fail to capture the particular *semantics* that gets incorporated into human movement.

Furthermore, as we noted above, it is sensible to believe that most human movement is indeed driven by certain purposes. We can therefore only note that specific and important characteristics of human movement that should get captured in the aforementioned models are linked to two important factors. The first factor consists of *people's daily routines* (such as visiting a coffee shop in the morning, going to work, and/or shopping or going to a restaurant in the evening). A second factor is the actual *location* of the different *points of interest* that humans visit. The combination of these two factors could explain the characteristics that older studies on human movement have observed.

In this paper we aim to answer the following research questions. What are the actual semantics of human movement? What are the locations that people visit, with what frequency, and at what time scales? What is the influence of the spread of these locations on human movement?

2. DATASETS

There are two mobility datasets containing user trajectories that we are using in this paper. In an effort to make our study more generic we use datasets that were collected in different regions of the globe. We further describe them.

The first dataset we use is a GPS trajectory dataset collected by Microsoft Research Asia in their GeoLife project [2]. It contains the data of 155 users that were monitored over a period of over two years. The second dataset we are using is a dataset collected in the United States from the backbone of a mobile provider network. It contains the trajectories of 4,429 users that fielded applications that report their GPS locations back to the mobile provider for different purposes.

Table 1: Point of interest statistics.

Dataset	**Illinois**	**Indiana**	**Michigan**	**China**
Restaurants	19,400	9,231	14,671	116,095
Shops	56,768	25,454	46,199	267,541
Govt. Offices	4,645	2,890	5,185	137,837
Hospitals, Clinics	748	466	963	54,545
Libraries	692	372	585	40,361
Stadiums	71	24	61	17,868
Population [M]	12.91	6.42	9.96	1,331
Size [1000 sq. mi]	57.9	36.4	96.7	3,705

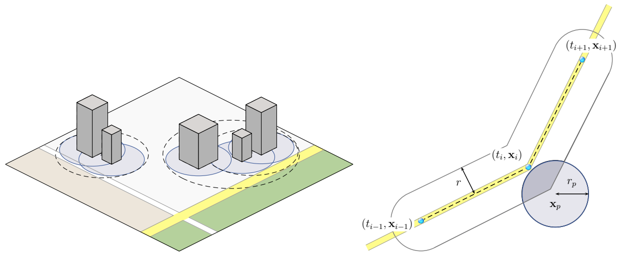
The above datasets are enhanced by extracting datasets containing the GPS coordinates of points of interest from China and the United States. We have the latitude, and longitudes of the locations. Statistics for some of the categories we extracted can be seen in Table 1. We show the points of interest for China and the states of Illinois, Indiana, and Michigan that are located in the United States Midwest (that is the area where most of our users are).

*An extended version is available as a Technical Report [4].

WWW2012, April 16-20, 2012, Lyon, France.
ACM X-XXXXX-XXX-X/XX/XXXX.

3. METHODOLOGY

Our analysis requires a proper formalization of the notion of interactions between users and possibly grouped points of interest. The first stage consists in turning one-dimensional trajectories and positions into two-dimensional areas in such a way that an interaction can be thought of as intersecting areas over a time interval. User positions and points of interest are therefore replaced by discs of radius r and r_p respectively. A linear trajectory thus becomes a rectangle.



(a) Merging businesses. (b) Interaction model.

Figure 1: Methodology of our analysis.

To capture the fact that businesses are often grouped, *e.g.* in shopping malls and the fact that larger points of interest (*e.g.* stadiums) have multiple close-together coordinate entries in our datasets, we use a region-based segmentation approach defining meta points of interest for points of interest of the same type. When two (meta) points of interest are close enough, *i.e.* the intersection between their respective discs is greater than a given threshold, they are merged into a single meta point of interest represented by the *minimum enclosing disc* (see Fig. 1a).

We define a model of interaction inspired from physics: given two successive snapshots $(t_i, \mathbf{x}_i)$ and $(t_{i+1}, \mathbf{x}_{i+1})$ of a user and a point of interest p at position $\mathbf{x}_p$, the *power* of the interaction at time t is defined as a function $p(\cdot)$ of the distance between the user's position $\mathbf{x}(t)$ and the position $\mathbf{x}_p$ of the point of interest. The *energy* E of the interaction during the time interval $[t_i, t_{i+1}]$ is obtained through integration:

$$E((t_i, \mathbf{x}_i), (t_{i+1}, \mathbf{x}_{i+1}), \mathbf{x}_p) = \int_{t_i}^{t_{i+1}} p(d(\mathbf{x}(t), \mathbf{x}_p))\, dt \ . \qquad (1)$$

In our analysis we used the area of intersection as a power function. The energy received by a point of interest is obtained by summing the energy of its interactions with all the segments of all users' trajectories. Such an energy-based approach captures both the spatial and temporal aspects of interactions. In case of an interaction with a meta point of interest, the energy of the interaction is evenly split among the points of interest composing the meta point of interest.

4. EVALUATION

By using the methodology we defined above, we evaluate user interaction with points of interest to answer the question about which point of interest types are more popular. In general, people do not interact with different types of points of interest in the same way. In particular, people visit certain places frequently depending on their function: grocery shopping, dining at a restaurant, watching a game in a stadium, *etc.*

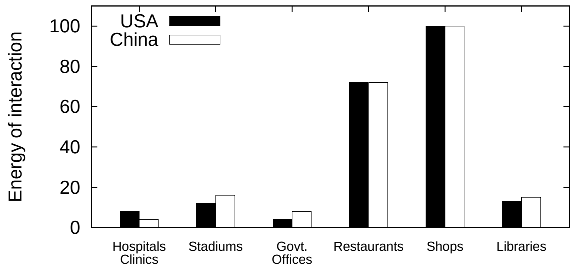


Figure 2: Energy of interaction for users interacting with the specific point of interest types.

Fig. 2 shows the results. The histograms are scaled to the highest energy value observed for each dataset, in both cases shops. We purposely do not show values for companies because values for interactions that users experience with the companies they work at dwarf other values.

A few findings from the figure are as follows. The most important is that the results are surprisingly consistent across the two datasets even though they were collected in different parts of the globe. Further, restaurants and shops lead the energy-of-interaction race in both figures. They hold the most energy out of all the analyzed point of interest types. This is not surprising as they are part of people's daily activities. In addition, although many more users are seen at stadiums for the duration of the trace, users do not go to stadiums frequently. In fact libraries and stadiums compare in terms of energy of interaction most probably because the users that *do* visit libraries do so more frequently than the users that visit stadiums.

Further insights can be found in [4], namely (*i*) time of day effects for point of interest types, and (*ii*) the number of users seen at specific businesses that carry publicly available WiFi access points (*e.g.*, McDonald, Starbucks) that are potential targets for data offloading for mobile users.

5. CONCLUSIONS

In this paper we have proposed a new approach to analyzing the mobility of humans. Noting the lack of expressiveness incorporated in to existing studies, we proposed and performed the first-of-its-kind joint analysis of human mobility correlated with the surrounding environment, *i.e.* the points of interest that they visit. Our key finding is in demonstrating user affinity (with different visiting frequency depending on point of interest type) towards specific points of interest and specific businesses. We believe our results show noteworthy promise for further research in this area, clearing the way for future advances in understanding basic human behavior and impacting problems related to mobile transfer scheduling algorithms, mobile infrastructure placement, mobile social networking, mobile advertising, *etc.*

6. REFERENCES

[1] Brockmann, D., Hufnagel, L., and Geisel, T. The Scaling Laws of Human Travel. *Nature 439* (2006), 462–465.

[2] GeoLife GPS Trajectories. `http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/`.

[3] Gonzalez, M., Hidalgo, C., and Barabasi, A. Understanding Individual Human Mobility Patterns. *Nature 453* (2008), 779–782.

[4] Trestian, I., Huguenin, K., Su, L., and Kuzmanovic, A. Understanding Human Movement Semantics: A Point of Interest Based Approach. Research Report RR-7716, INRIA, Aug. 2011. `http://hal.inria.fr/inria-00616876`.