Context-Aware Navigation: Improving urban living experience with predictive navigation system

Giusy Di Lorenzo, Santi Phithakkitnukoon, Carlo Ratti
SENSEable City Laboratory
Massachusetts Institute of Technology
{ giusy, santi, ratti}@mit.edu

ABSTRACT

With the goal of making urban living more efficient and pleasant through a better driving experience, we present a predictive navigation system that makes use of city-described data along with the driver's driving history to predict destination in order to provide useful routing recommendation accordingly. In this paper, we describe a predictive destination framework composed of models that work adaptively in both routine and non-routine trip scenario. With existence of temporal sequences in routine trips, Markov model has been adopted for routine model. Non-routine model, on the other hand, is challenged by the lack of such temporal sequences, thus Bayesian approach has been taken with *a priori* derived from relevant individual history as well as city-described data such as land use, population density, and POIs. Experimental results show that our model is able to predict the destinations accurately with the error of 4.29 km.

1. INTRODUCTION

In recent years, a massive increase in the volume of records of when and where people are has been produced with the large deployment of pervasive technologies in the cities. These digital footprints of individual mobility pattern along with the traditional survey data like land use and population density as well as a more up-to-date data that describes the geographical location of interest like POIs (Point of Interest) can be used to enhance urban living experience. We envision a context-aware urban computing system that makes use of these data that describe the cities from different aspects to make urban living more efficient and pleasant. Our initial focus is on the predictive navigation system that mimics the friendly expertise of a driving companion who is familiar with both the driver and the city. Instead of focusing solely on determining routes to a specified waypoint, the system utilizes the analysis of the driver's behavior in order to identify the set of goals the driver would like to achieve. Furthermore, it involves an understanding of the city beyond what can be seen through the windshield, incorporating information such as business and shopping districts, tourist and residential areas, as well as real-time event information and environmental conditions. Current navigation systems are focused on finding waypoints. They are capable of pointing out the shortest route to a destination, integrating traffic information, and identifying points-of-interest. These systems can successfully assist in driving to a fixed desired destination. However, individuals often make trips with the purpose of achieving various goals, such as purchasing gasoline, watching a movie, or participating in a public event, while the physical location of their destination is flexible. Here we take the first and essential steps toward building such system and describe, in this paper, a methodology used to predict the driver's destination using probabilistic approach incorporating the driver's history, land use information, population density, and POIs information.

UBI Challenge Workshop 2010, Copenhagen, Denmark

2. RELATED WORK

The idea of location-based services has motivated a number of researchers to not only attempt to retrieve information about the current location and surroundings but also the mobility and predict future locations. In [7], the author divides a map into 1x1 km^2 cells and predicts the destination cell that may include unlikely destination such as middles of the lakes, ocean, and forrest. In [8], the author proposes to predict the next road segment(s) using a n-th order discrete Markov model, and examines the effect of the number of considered past segments, time of the day, and taking into account other drivers' training data. Froehich and Krumm [5] describe a method to measure to extract regular routes. Their goal is to recognize when the driver is on a regular route and predict that the driver will stay on that route by comparing an ongoing trip to the previously driven routes and finds the closest match. Tanaka et al. [11] propose methods using probabilistic approach to predict driver's destination based on number of trips made between places and time of the day, day of the week, number of passengers, and weather condition.

A common shortcoming of all of these aforementioned models is inability to predict an unknown destination - a destination that has not yet been visited (it does not appear in training data). Krumm [9] defines a destination as a center of road segment and describes a simple algorithm to predict which turning direction will be made by the driver at an intersection based on assumptions that driver will turn at the intersection to reach destination efficiently in terms of time and driver will likely to make a turn to the direction that leads to more choices (efficiently reachable destinations). Krumm and Horvitz [10] solve the shortcoming of [7] by eliminating unlikely destination cells such as middles of the lakes and oceans from the set of possible destinations. They propose a model for predicting destinations that can be either places that driver has previously visited or new destinations. Two main models have been proposed namely open-world model and closed-world model. The final model is the integration of these two models. The shortcoming of this model is the fact that the final model essentially becomes a close-world model that is not feasible for predicting a new destination (after 14 days based on their result). In contrast, our model is feasible to predict a new destination as it adaptively changes a set of likely new destinations with associated probabilities.

3. PREDICTIVE MODEL

Here we focus on the prediction of the destination given a set of initial conditions. These conditions consist of the current position of the car, current time, as well as the past trip history of the driver. In addition, we also incorporating the knowledge of land use and population density into our model to help build a priori knowledge especially for unknown destinations (i.e. places that have never been visited in the past). We speculate that other information might also become useful (e.g. number of passengers, traffic conditions, weather status) and we expect to include near future additions that consider it. Here we propose the baseline methodology and algorithms for this research.

By nature, the problem involves a considerable degree of uncertainty. People seldom move in a totally regular fashion, even in the most common routines of commuting from home to work and vice versa.

For our analysis, we use GPS traces from three sources with a total of 11 drivers. Our first source contains traces of five drivers in San Francisco, USA over the course of three months in 2008. Our second data set includes five drivers from Coimbra, Protugal of a period of two months in 2009. The other source is one-year GPS traces of one driver in Seattle, USA during 2002 and 2003. All drivers carried GPS tracking systems while driving. The tracking is turned on while driving and off while the car is parked. Sampling rate is one second.

We model space using 500x500m^2 square grids. The goal is to predict destination of the driver given a set of prior knowledge. Our predictive model is composed of two sub models corresponding to two main driving scenarios. One refers to when the driver is traveling in a *routine trip*, e.g. home to work, home to supermarket, in which destinations and routes have been visited previously. Hence the individual statistics can be extracted as *a priori* from the history observed in the training data, and used to construct a probabilistic predictive model. On the other hand, when the driver travels in a *non-routine trip* or to a new destination that has never been visited previously, no history can be observed from the training data. Hence *a priori* cannot be obtained directly from individual history but can be induced by relevant individual history e.g. purpose of trip given day of week, as well as community-based statistics e.g., where other people go.

3.1 Routine Model

The *Routine Model* bases its estimation on the driver's driving history adopting Markov model. The probability of cell i being the destination cell is $Pr\{c_i$ is destination cell$\} = P(c_i = dest.|c_j, c_{j-1}, c_{j-2}, , , ,) = P(c_i = dest.|c_j)$, where c_j is the current cell and c_{j-k} is the k^{th} previous cell. The probability transition distribution of the current cell j, P_j, can be computed where the cell i with the maximum probability is then chosen as the most likely destination cell as

$$P(c_i = dest.) \Rightarrow \arg\max_i(P_j) = \arg\max_i(\frac{N_{j,i}}{N_j}), \quad (1)$$

where Nj, i is the number of times the driver has visited cell i after previously visited cell j and N_j is the number of times the driver has visited cell j.

The challenge here is identifying a set of destination candidates. From an infinite number of possible destinations, we can narrow our choices by making use of expected trip time. Instead of including all cells in a set of potential destinations, we can consider only more likely cells. An *expected trip time* can be easily computed from driving history as an average trip time of given time of the day and day of the week. Let $\bar{t}_{o,d}$ denote expected trip time from origin to destination, t_{o,c_j} represent time spent driving from origin to the current cell c_j, and $t_{c_j,d}$ be the driving time from the current cell c_j to the destination. Hereby, $\bar{t}_{o,d} = t_{o,c_j} + t_{c_j,d}$.

Therefore, the set of destination candidates includes only those cells that are $t_{c_j,d}$ time unit away from the current cell. We can also further eliminate cell candidates that the driver has not previously visited. Figure 1 illustrates this candidate search algorithm where blue-highlighted cell indicates the current cell while yellow cells are initial candidates and orange cells are final cell candidates. If the current trip time exceeds the expected trip time $\bar{t}_{o,d}$, then the driver is assumed to be near the destination and hence the adjacent cells will be considered as candidates. Note that $t_{c_j,d}$ is estimated in our algorithm using Google Maps.

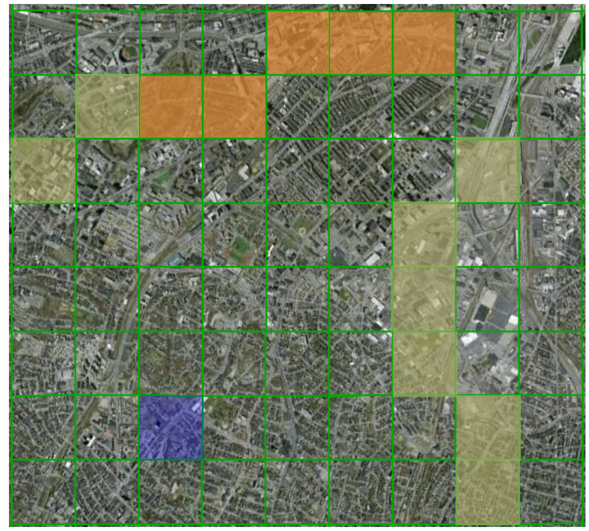

Figure 1: Search for destination cell candidates based on expected trip time. Blue-highlighted cell indicates the current cell while yellow cells are initial candidates and orange cells are final cell candidates.

3.2 Non-routine Model

Clearly, it becomes more challenging in the non-routine scenario as no history of visiting new cells can be observed, thus any statistics that would facilitate building a probability distribution are not obvious. Markov model may not be suitable here as temporal sequences are not sufficiently available. On the other hand, Bayesian model seems to fit well by inferring likelihood of destination seeing relevant data of a given cell. By applying Bayes' rule, we have

$$\begin{aligned} Pr\{c_i \text{ is dest.}\} &= P(c_i = dest. | \Omega = \omega) \\ &= \frac{P(\Omega = \omega | c_i = dest.) P(c_i = dest.)}{\sum_{j=1}^{n} P(\Omega = \omega | c_j = dest.) P(c_j = dest.)} \end{aligned} \tag{2}$$

where random variable Ω represents observed data and n is the total number of destination candidates.

In order to construct a probability distribution, we need to gather relevant prior knowledge. Based on our general observation of non-routine trips, we believe that a non-routine trip is influenced by:

(i) **Purpose of the trip**: Normally there is a purpose of each trip that we make e.g. to eat, to work, to watch movie, to shop, and so on. These purposes of the trip tend to have a temporal pattern e.g. go to a restaurant for lunch every Monday at noon, go to shop at a mall every Sunday afternoon, etc. For a given day of the week and time of the day, trip purpose can be estimated based on the individual history (i.e. most frequent trip purpose of given day and time). This estimated trip purpose can then be used to better guess the headed unknown destination e.g. if the person has history of having lunch around noon on weekday and is currently driving out of his routine on Monday at noon, then most likely he is going to a new restaurant for lunch. To account for this knowledge, we include Land Use and Point of Interest (POI) information in the observed data variable Ω.

We can characterize each cell c_i with a set $poi_{c_i} = \{poi_{c_i}(1), poi_{c_i}(2), \ldots, poi_{c_i}(q) \ldots, poi_{c_i}(Q)\}$ where $poi_{c_i}(q)$ is the number of *POIs* of the category q within cell c_i. The probability of the driver being at the POI category q given c_i, *time* and *day* can be computed as follows:

$$P(q | c_i, Time, Day) = \frac{poi_j(q)}{\sum_{j=1}^{Q} poi_j(j)} \tag{3}$$

An analogous argument can be made for the land use information, once we define the distribution of the land use associated with each cell c_i as $lan_{c_i} = \{lan_{c_i}(1), luc_{c_i}(2), \ldots\}$.

As an example, Fig.2 and Fig.3 show, respectively, the overall distribution drawn from drivers in San Francisco of the different land use and POI categories. A sample here is given for two different time of the day: 10 am and 6pm, to show the different distributions across different different categories and time of the day. As shown here, one can observe that drivers prefer to visit places across different POIs and land use categories. From land use distribution, residential, transportation, and commercial areas are the most visited places in the morning period (10 am) while highway (included in forest), residential, and industrial are more attractive in the evening (6pm). Regarding the POI categories, drivers tend to visit shopping and restaurant areas in the evening more than in the morning.

(ii) **Familiarity**: Familiarity plays an important role in driving. People who are familiar with a particular area, tend to explore new places nearby, whereas people who are not quite familiar with the area tend to go to popular places. With these observations, we construct our model accordingly. We apply *gravity model* [6] – centered around the visited cells so that it accounts for exploring a new place (destination) within vicinity of the familiar places. In the case of traveling to a new and unfamiliar area, land use and population density information can help shape up probability distribution [4].

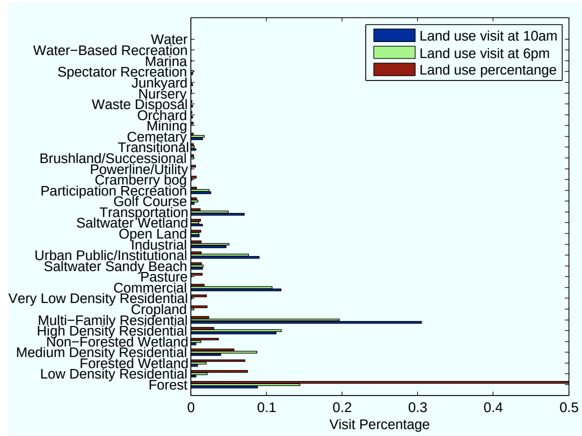

Figure 2: Land use preferences at different time of the day

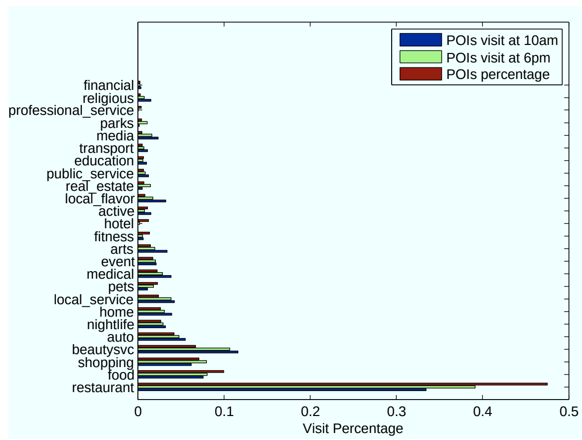

Figure 3: POIs preferences at different time of the day

Consequently, the likelihood function is thus defined as

$$\begin{aligned} L(c_i = dest. | \Omega = \omega) &= P(\Omega = \omega | c_i = dest.) \\ &= \prod_{j=1}^{3} P(\Omega_j | c_i = dest.) \\ &= \prod_{j=1}^{3} f_{\Omega_j | Time, Day}(\Omega_j | c_i = dest.), \end{aligned} \tag{4}$$

where $f_{\Omega_j | Time, Day}(\Omega_j)$ is the probability mass function (pmf) of Ω_j given time of the day and day of the week, and Ω_1, Ω_2, Ω_3, and Ω_4 correspond to population density, land use, gravity model and POIs respectively. The most likely cell destination is thus the one that maximizes the likelihood function, i.e.

$$\text{Dest. cell} = \arg\max_i \prod_{j=1}^{3} f_{\Omega_j|Time,Day}(\Omega_j|c_i = dest.) \quad (5)$$

Likewise, the search for destination candidates for this non-routine model can be done in a similar way with the routine model, except that we use the initial set of candidates without the elimination by visits because none of the initial cell candidates have been visited previously. Nevertheless, a further elimination can be performed using population density information (which provides information about human population per unit area) by discarding any cells with zero human population (these areas are typically middles of the ocean, dessert, etc.).

The main model combines routine and non-routine model with weight factor α as follows:

$$P(c_i \text{ is dest.}) = (1-\alpha)P_R(c_i) + \alpha P_N(c_i), \quad (6)$$

where P_R and P_N are routine and non-routine model.

To be more specific on the sources of land use and population density information that we use for our analysis, here we give details about them. For land use, information is obtained from Web GIS's Land Use and Land Cover (LULC) [3] and MIT Portugal Intelligent Transportation Systems (ITS)'s CityMotion Project [2]. For the POIs, data has been extracted from Yelp[1] at the same cells grid level, and grouped into 22 categories. Population density information is acquired from GeoWeb (public version) – MIT GIS Services [1] as well as MIT Portugal Intelligent Transportation Systems (ITS)'s CityMotion Project [2].

4. EXPERIMENTAL RESULTS

We evaluate our model by examining routine and non-routine model individually, as well as their combination (Eq. (6)). First, we find that Markov approach for our routine model performs better than Bayesian approach previously proposed in [10]. Our Markov model yields an average error of 3.93 km while Bayes has 4.37 km in error. The error is computed as a distance in kilometers (km) between the predicted destinations and the actual destinations. We then examine the error rates for non-routine model using different combination of gravity model, population density, land use, and POIs data. Our model is tested on all of three different sets of GPS traces. Table 1 shows that the complete model with gravity model, population density, land use, and POIs information yields the minimum error. With our combined model (Eq. (6)) with optimal α of 0.9, we are able to predicted accurately with the average error rate of 4.2852 km.

Table 1: Error rates for the combined model comparing different combination of observed data.

Combination	Error (km)
Grav. only	4.7301
Grav. & Pop. Dens.	4.7303
Grav. & Land use	4.3776
Grav. & POIs	4.4939
Grav. & Pop. Dens. & Land use & POIs	4.439

5. CONCLUSIONS

The ultimate goal is to build an innovative navigation system with new interactive capabilities that will assist drivers in identifying and planning activities in real-time and determining optimal routes for commuting between destinations such that urban living experience becomes more efficient and pleasant. A mandatory task is thus to predict the destination of the driver as well as the most likely route that the driver will follow. In this paper, we describe a framework for the predictive navigation system that predicts driver's destination based on driving history and observable data about the cities. Models for routine and non-routine trips have been developed using probabilistic approach. Routine model adopts Markov approach based on individual history while non-routine model is based on Bayesian inference making use of individual history as well as relevant community-based data such as population density, land use, and POIs. The experimental results show that our model is able to predict accurately with an average error of 4.29 km.

As our future direction, we will continue to improve our current model by examining different aspects of the model as well as exploring other relevant observable data to better construct a priori distribution.

[1] $http://www.yelp.com$

Acknowledgment

The authors gratefully acknowledge support by Volkswagen of America Electronic Research Lab, AT&T Foundation, National Science Foundation, MIT Portugal, and all the SENSEable City Laboratory Consortium members.

6. REFERENCES

[1] Geoweb (public version) – mit gis services. In *https://arrowsmith.mit.edu/geoweb/*.
[2] MIT Portugal Intelligent Transportation Systems (ITS)'s CityMotion Project.
[3] Web gis's land use and land cover (lulc). In *http://www.webgis.com/lulcdata.html*.
[4] F. Calabrese, J. Reades, and C. Ratti. Eigenplaces: Segmenting space through digital signatures. *IEEE Pervasive Computing*, 9:78–84, 2009.
[5] J. Froehlich and J. Krumm. Route prediction from trip observations. In *2008 SAE International*, 2008.
[6] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.
[7] J. Krumm. Real time destination prediction based on efficient routes. In *2006 SAE International 2006*, pages 1–6, 2006.
[8] J. Krumm. A markov model for driver turn prediction. In *SAE 2008 World Congress*, 2008.
[9] J. Krumm. Where will they turn: predicting turn proportions at intersections. In *Personal and Ubiquitous Computing*, August 2009.
[10] J. Krumm and E. Horvitz. Predestination: Inferring destinations from partial trajectories. In *UbiComp 2006: Ubiquitous Computing*, pages 243–260, 2006.
[11] K. Tanaka, Y. Kishino, T. Terada, and S. Nishio. A destination prediction method using driving contexts and trajectory for car navigation systems.