11th International Conference on User Modeling (UM 2007), June 25-29, 2007, Corfu, Greece.

Driver Destination Models

John Krumm and Eric Horvitz

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 U.S.A.
{jckrumm, horvitz}@microsoft.com

Abstract. Predictive models of destinations represent an opportunity in the context of the increasing availability and sophistication of in-car driving aids. We present analyses of drivers' destinations based on GPS data recorded from 180 volunteer subjects. We focus on the probability of observing drivers visit previously unobserved destinations given time of day and day of week, and the rate of decline of observing such new destinations with time. For the latter, we discover a statistically significant difference based on gender.

Keywords: driving, mobility, destinations, cars, automobiles, navigation

1 Introduction

Computing is increasingly coming to the aid of drivers, with improved in-car navigation systems, advanced routing services, and comprehensive point-of-interest databases, some with intermittent and ongoing network access. With this sophistication comes the opportunity for developing better models of driver preferences and behavior, to both improve services and decrease unnecessary driving distractions. An understanding of drivers' destinations is one promising direction for improving in-car services. Knowledge of a driver's destination can be used to give anticipatory alerts about traffic and recommendations for re-routing, reminders about location-based tasks, relevant advertising, and useful suggestions for parking, restaurants, and other points of interest. Destination modeling can be especially useful in methods for destination prediction[1, 2]. In [1], we present probabilistic models that predict the destination of drivers as trips progress, based on observational data. In this paper, we review statistics of destinations that support the probabilistic modeling efforts. The analysis is based on logs of trips of 180 drivers. We first show how destinations vary with time of day. Then, we review our research on the scope of destinations for users, exploring how the likelihood of seeing new destinations visited decreases with the observational period.

Our studies are based on GPS data logged from volunteer drivers participating in the Microsoft Multiperson Location Survey (MSMLS), an ongoing study of driving behavior we initiated in early 2004. We recruited employees from our institution and their adult family members by offering participants a 1 in 100 chance of winning a

US$ 200 MP3 player. Each subject was asked to complete a demographic questionnaire. Based on the demographic questionnaire, the average age of our 180 subjects was 36.6 years. 36% of them had non-adult children, 72% were male, and 25% were single.

We ask participants in the MSMLS study to keep the GPS device on the dashboard of their car for two weeks. We modified the GPS devices to allow them to be used without intervention over the observational period; the modified devices turn on when receiving power and retain logs between trips. The devices are set to record time-stamped latitude and longitude coordinates only when the car is moving, reducing the chance of exceeding the GPS's 10,000-point memory over the two-week period. The adaptive recording mode gave points whose median separation distance was 64.4 meters and whose median separation time was 6 seconds. We segmented the GPS data into discrete trips by splitting the sequence at points separated by more than five minutes and eliminating trips under 10 points or one kilometer long. The final point in each trip segment is the trip's destination, which gives us a list of latitude and longitude points, one for each of the 8319 resulting trips.

2 Destinations over Time

We sought to understand how often drivers take trips to different locations and how often they go to places that they have not been observed to visit before, as a function of the length of the observational period. Such data provides prior probability information that can be used within predictive models of destination, which may also consider such factors as time of day and the trajectory of a trip in progress.

In our work on predicting destinations, we explicitly model the likelihood that a driver will visit a destination that they have not been observed to visit over the course of the observational period [1]. The probability of a user visiting a location that has not been observed before is critical in *open-world modeling* of destinations, which admits previously unseen destinations into location prediction, also described in [1]. Prior research on destination prediction has assumed a closed world, limiting the scope of destination prediction to those locations that have been previously visited by drivers. This work includes work by Marmasse and Schmandt[3], Ashbrook and Starner[4], Hariharan and Toyama[5], Liao *et al*.[6], , and Gogate *et al*.[7]. All of these studies only consider as candidate destinations those locations that have been extracted from GPS histories, *i.e.* places that subjects have actually visited.

We begin by examining when drivers make trips. Using the time-stamped trips from our MSMLS data, we computed the mean number of trips per week over all our driving subjects for each hour of the day. For each of these trips, we also computed whether or not the destination had been observed before in the study. For a given subject, we defined a location as a *new* destination when no previously observed destination visited by the user had been within a 200 meter radius of that location.

The results are displayed at the top of Figure 1. The dark curve shows the average number of trips made per week in a given hour of the day over the course of the survey. We see that the number of trips peaks around 5 p.m. – 6 p.m., when the average driver made about 2.66 trips per week. There is another peak around 8 a.m. –

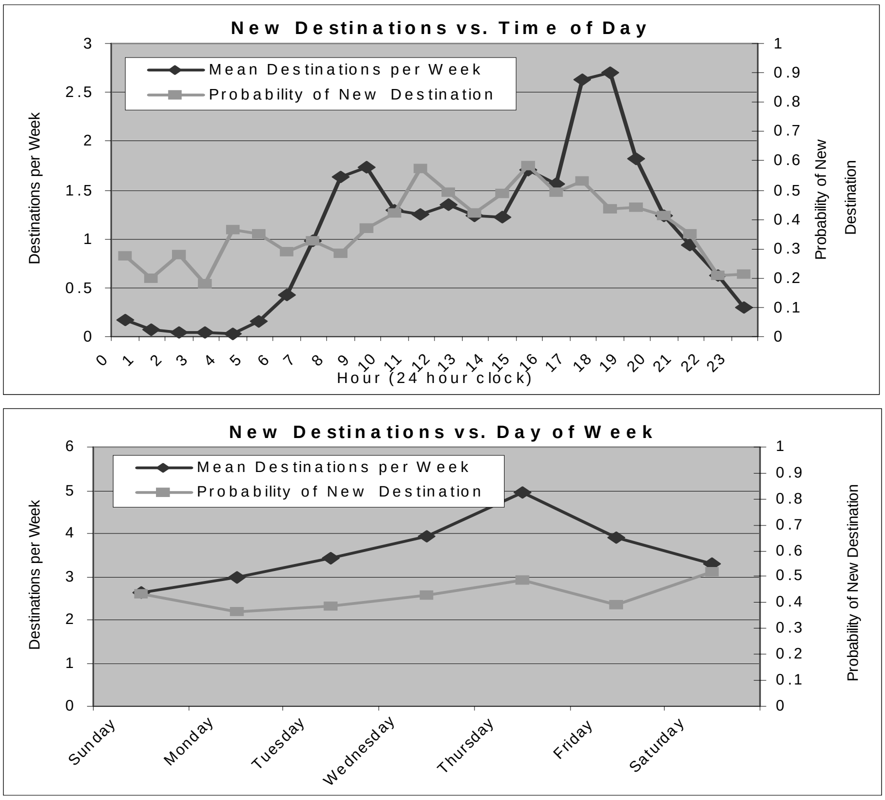

Figure 1: The number of mean trips in a week and the probability of visiting a previously unobserved destination over the time of day (top graph) and the day of the week (bottom).

9 a.m., with about 1.69 trips per week. Both peaks may reflect commuting to and from work. As expected, very few trips are made late at night.

This data can serve as a prior probability distribution in probabilistic models of driving destination and activity. The lighter curve in the graph at the top of Figure 1 shows the proportion of new destinations, providing the probabilities that destinations reached at different times of the day have been previously visited. The probability of visiting a new destination peaks at 11 a.m. and again at 3 p.m. The data shows that, over the course of the observational period of the study, new destinations are most likely visited in the middle of the day. Such data could help a navigation system automatically determine whether or not to offer driving assistance and information about the predicted destination.

We look at the same data conditioned on the day of the week in the bottom of Figure 1. Thursday is the peak day for driving trips, after a steady rise in trips starting with Sunday. The plot also shows the probability of visiting a new destination as a function of the day of the week. Saturday slightly beats Thursday as the most probable day to visit a new destination on a given trip.

3 Falloff in Observing New Destinations

As we mentioned, most destination prediction algorithms limit their predictions about future destinations to previously observed destinations [2]. Clearly, the rate of seeing previously unobserved destinations is highest at the outset of the observation period. We would expect to see the rate of observing such new destinations decrease with ongoing observation, given drivers' habitual patterns of visiting locations, based on recurrent activities. We now focus on our studies of the change in rate of seeing new locations with observation period for all participants as well as breakouts for people with different demographical attributes.

The black squares in Figure 2 show how the number of new destinations decreases significantly with the number of days into the MSMLS survey. The average number of new destinations over all subjects on the first day is 3.6, dropping to 1.6 on the second day and 1.34 on the third day. To avoid edge effects, we ignore the first day of the study. The drop in new destinations is well modeled by an exponential decay of the form $d(t) = ae^{-bt}$, where t is the number of days into the survey and $d(t)$ is the number of destinations on day t that have not been visited on any previous day. We performed a least squares fit of this equation to the measured average of new destinations and found $d(t) = 2.142e^{-0.134t}$.

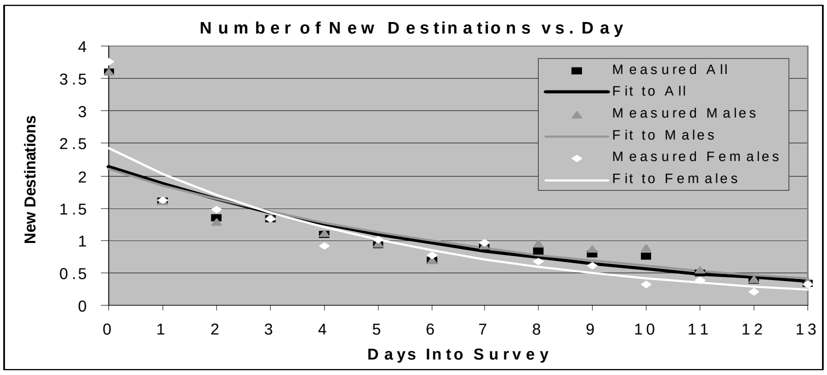

Figure 2: Number of new destinations visited over duration of observation period for all participants and for groups conditioned on gender.

We now explore how new destinations decrease over time based on drivers' demographics. For each subject, we computed parameters a and b of the exponential model describing decay in seeing new destinations using least squares. We examined splits along gender, single versus partnered, children at home, and the existence of extended family within a 50 mile radius. For each category, we removed all b exponents beyond the category's 3-sigma points to eliminate the effect of outliers. We also removed subjects whose least squares fit was degenerate due to a lack of data, *e.g.* only one day of available data. We performed a *t*-test on the b values to compare the decay rate for each pair of categories. Of the four splits, we found a statistically significant difference in decay rates only along gender lines. For men, the mean decay rate is $b = 0.0695$, and for women $b = 0.1960$, implying that women's rate of visiting new destinations falls off faster than men's ($t(35) = 2.03$, $p < 0.026$). The aggregate fits for men and women are shown in Figure 2.

The analysis shows that, in aggregate, drivers quickly approach a steady state where visiting new destinations is relatively rare. After 14 days of observation, for example, drivers are visiting an average of 0.33 new destinations per day. Since the median number of trips per day for all our subjects was 3.53, this implies that after two weeks, the probability of a driver going to a new destination per trip is approximately $0.33/3.53 \approx 0.09$.

4 Conclusions

We reviewed the MSMLS dataset and presented several analyses of the data in support of ongoing efforts to construct probabilistic models of destination. We showed how the number of trips and visits to previously unobserved destinations varies with the time of day and day of week. We described how quickly drivers tend to reach a steady state in visiting new destinations over an observation horizon. Finally, we explored the influence of demographic attributes on the rate of visiting previously unobserved destinations and identified an influence of gender.

References

1. Krumm, J. and E. Horvitz. Predestination: Inferring Destinations from Partial Trajectories. in Eighth International Conference on Ubiquitous Computing (UbiComp 2006). 2006. Orange County, CA.
2. Kostov, V., et al. Travel Destination Prediction Using Frequent Crossing Pattern from Driving History. in 8th International IEEE Conference on Intelligent Transportation Systems. 2005. Vienna, Austria.
3. Marmasse, N. and C. Schmandt, A User-Centered Location Model. Personal and Ubiquitous Computing, 2002(6): p. 318-321.
4. Ashbrook, D. and T. Starner, Using GPS To Learn Significant Locations and Predict Movement Across Multiple Users. Personal and Ubiquitous Computing, 2003. **7**(5): p. 275-286.
5. Hariharan, R. and K. Toyama. Project Lachesis: Parsing and Modeling Location Histories. in Geographic Information Science: Third International Conference, GIScience 2004. 2004. Adelphi, MD, USA: Springer-Verlag GmbH.
6. Liao, L., D. Fox, and H. Kautz. Learning and Inferring Transportation Routines. in Proceedings of the 19th National Conference on Artificial Intelligence (AAAI 2004). 2004. San Jose, CA, USA.
7. Gogate, V., et al. Modeling Transportation Routines using Hybrid Dynamic Mixed Networks. in Uncertainty in Artificial Intelligence (UAI 2005). 2005.