COLLECTION AND ANALYSIS OF GPS-BASED TRAVEL DATA FOR UNDERSTANDING AND MODELING ACTIVITY-TRAVEL PATTERNS IN TIME AND SPACE

Ram M. Pendyala, Department of Civil and Environmental Engineering, University of South Florida, 4202 East Fowler Avenue, ENB118, Tampa, FL 33620. Ph: (813) 974-1084; Email: pendyala@eng.usf.edu

Introduction

Activity-based microsimulation models aim to provide a rigorous analytical and behavioral framework for forecasting travel demand. Over the past decade, rapid advances have been made in the specification, formulation, and estimation of activity-based model systems. Simpler forms of the activity-based modeling paradigm, commonly referred to as tour-based model systems, have been developed and applied in practice with a view to enhancing the state-of-the-practice in travel demand modeling and transportation planning. Examples of such applications include Portland, San Francisco, New York, and Columbus. Several other agencies, including those of Denver, Atlanta, Dallas-Fort Worth, Sacramento, and the Puget Sound to name a few, are considering or have initiated the development of tour-based model systems for their areas.

Although tour-based model systems constitute a significant enhancement over traditional four-step trip-based models because they account for inter-dependency among trips, trip chaining, and trip scheduling, they are not full-fledged activity based model systems. This is because they do not fully reflect spatial and temporal constraints and opportunities, model interactions among agents, capture time use and allocation behavior, and consider activity participation along the continuous time dimension. In addition, both tour-based and activity-based models developed thus far have done little to account for route choice behavior and the effects of that behavior on activity participation, activity-travel durations, and scheduling patterns.

It is clear that the development of full-fledged activity-based models that fully account for and reflect the aspects noted in the previous paragraph calls for the collection, synthesis, organization, manipulation, integration, visualization, analysis, and modeling of detailed data about people's activity-travel patterns by time of day, mode and route choices, activity-sequencing and trip chaining, spatio-temporal constraints and opportunities, and interactions in time and space. Until recently, collecting such information through traditional travel surveys was virtually impossible. Therefore, activity and tour-based models developed thus far have had to make compromises and simplifying assumptions in model specification, behavioral basis, and structure. However, more recently, GPS-based data collection components/experiments, when combined with activity-travel survey efforts, are offering rich and detailed data about aspects of behavior that hitherto remained a mystery. The time is ripe for geographers and travel demand modeling researchers to work together and explore ways of fully exploiting the GPS data in combination with traditional activity-travel behavior data for developing new generations of activity-based microsimulation models.

Survey Methods: Collection of GPS-Based Activity-Travel Data

In the recent past, considerable experience has been gained in the collection of GPS data in the context of activity-travel surveys. However, there are several aspects of GPS data collection that merit further exploration so that GPS-based data collection experiments can be set up to maximize information acquisition per unit cost. These areas for exploration include, but are not necessarily limited to:

1. *Comparisons of GPS and Traditional Travel Survey Data*: While there have been a few attempts to compare travel patterns from GPS and traditional travel survey samples to understand the extent to which traditional travel survey data sets miss or under-report trips, this appears to be an area that merits further exploration. These comparisons could help in developing correction factors for traditional

travel survey data sets. More importantly, these comparisons will help determine the extent to which short and infrequent trips and trip chaining are under-reported.

2. *Passive vs. Active Data Collection*: GPS instruments are able to collect location and timing information passively on a continuous basis. In other words, the respondent or traveler does not have to enter this information. However, the traveler has to enter or record information such as trip purpose and vehicle occupants. On the other hand, it might be possible to make inferences based on GPS data. The location coordinates of the trip end, when combined with a parcel-level land use map, might provide sufficient information to figure out the most likely trip purpose pursued by the individual. The duration for which the GPS records no locational movement could be used to determine the end point of a trip without having to burden the respondent to manually indicate the end point of a trip. What duration threshold should be used to signify that a trip has been completed? In summary, it merits exploration to perform some comparisons and tests to figure out what data we can reliably gather through passive GPS instruments and what data we need to actively gather from the respondents/travelers.

3. *Resolution of Data Collection*: GPS instruments are able to provide location and timing information on a second-by-second basis. This leads to huge amounts of data/records even for a single trip. Do we really need second-by-second data? Is it sufficient to record data at 30 second intervals, 10 second intervals, 1 minute intervals, or….? Perhaps, it would be possible to construct travel itineraries using records based on different time intervals and compare the itineraries so constructed. These comparisons could shed light on the ideal temporal resolution for GPS data collection. From a spatial perspective, the GPS instruments provide coordinate data (latitude and longitude). This is exactly what we need; the only question is whether we need the latitude and longitude information for every second.

4. *Understanding GPS Data Limitations*: While there is increasing recognition of the benefits of GPS data collection, perhaps it is necessary to make sure that we also understand the limitations of GPS data. In addition to cost considerations, what are potential limitations of GPS data collection efforts in terms of technology reliability, battery life (although this is getting better by quantum leaps), missing data (say, due to travel through urban canyons), and so on? Understanding these limitations will help identify mechanisms and tools for better exploiting and using GPS data in a travel analysis and modeling context.

Analysis and Use of GPS-Based Activity-Travel Data

The application of GPS-based activity-travel data for the development of activity-based microsimulation models of travel demand requires the analysis and use of the data from a variety of new perspectives that have not been traditionally addressed by travel behavior researchers due to data and technology limitations. However, the availability of GPS data together with rapid advances in GIS software technology has made it possible to exploit and apply GPS data in ways not possible in the past. Several aspects merit consideration in the context of analyzing and using GPS data in support of activity-based travel demand modeling. These include, but are not necessarily limited to:

1. *GPS Data Reduction and Synthesis*: GPS data sets are often large and include large amounts of temporal and spatial data on travel patterns. We must develop, and have widely available, tools that can reduce and synthesize the information/records into meaningful data sets and visual displays. Tools are needed to correct for errors, fill missing gaps, decipher beginning and ending points of trips, link trip ends with secondary data such as land use and network data, and so on.

2. *Day-to-Day Variability and Weekly Rhythms in Activity-Travel Participation*: One of the great advantages of using GPS-based methods is that one can potentially collect travel data over several days (and perhaps weeks) because the burden on the respondent is lower. A lot of the spatial and temporal information is recorded automatically by the GPS device. Then, GPS-based data should be used to analyze day-to-day variability and weekly rhythms in activity-travel patterns. The Mobidrive data collected in Germany over a six week period illustrated the potential value in having data over a longer period. We can learn about inter-day interaction, inter-person interaction, task allocation, intra-person variation, and habitual and occasional activity participation. In addition, having data over multiple days and weeks can help inform the time duration that should be covered by our activity-based microsimulation models. While it may be reasonable to have daily activity patterns as an output of these models, the models themselves may need to incorporate inter-day and inter-week interactions (history dependency, anticipatory dependency, etc.) to accurately model a single day's activity-travel patterns.

3. *Space-Time Paths and Prisms*: The biggest benefit that GPS data might be able to provide is in the measurement, representation, description, analysis, and explicit incorporation of space-time prisms and paths into our microsimulation models. What makes a model an activity-based model is the explicit incorporation and recognition of space-time constraints, opportunities, action space, and paths into the model system. While some attempts are being made to do that with currently available travel survey data, these attempts can be advanced by quantum leaps through the use of GPS data that provide *detailed spatio-temporal data* at a very fine resolution in an integrated manner. Essentially, the construction of true 3-D time-space paths and prisms is possible only with GPS data. As activity-based approaches attempt to model activity-travel patterns in *time and space*, there is no doubt that GPS data have much to offer in this regard. The challenge is to develop tools to analyze and model space-time paths and develop activity-based modeling paradigms/structures/frameworks that explicitly utilize these paths and prisms to simulate activity-travel patterns of individuals.

4. *Interactions Among Agents*: GPS data should be used to advance our understanding of interactions among agents. What is the interaction of individuals with the natural and built environment over space and time? By linking GPS data with land use (parcel level) data in a GIS framework, spatial correlations and relationships can be developed. These relationships can, in turn, be incorporated into activity-based microsimulation models in the context of constructing choice sets and modeling destination choice. Similarly, inter-dependencies among activities and trips and between individuals can also be explored in greater detail with GPS data.

5. *Route Choice*: It remains a pity that the route choice modeling community has not yet embraced the use of GPS data for modeling and understanding route choice behavior. This is the best source of route choice data and can provide valuable information on inter-relationships between trip chaining, mode choice, destination choice, and route choice. In addition, one can potentially construct behavioral paradigms on route choice by examining differences between paths chosen and paths that are shorter, faster, safer, and more scenic. Perhaps it is time to bring in the route choice modeling community in the analysis of GPS-based travel data.

6. *Supply (Network) Data*: One area in which GPS-based travel data can be very useful is in developing, enhancing, updating, and correcting network (supply) data. We are all familiar with the rather poor quality network supply data that is currently available and used in our travel demand models. These variables play critical roles in trip distribution, mode choice, and traffic assignment (i.e., at least three of the four steps). However, travel times and speeds are often suspect and inaccurate. GPS-based

travel data provides accurate spatial and temporal information at the finest resolution. In other words, we can get accurate travel times by time of day, potentially build speed-flow relationships using the data, and assess speed and travel time variability by link. This is another way in which the network modeling community can benefit from the use of GPS-based travel data.

Conclusion

The above discussion illustrates the potential use, value, and applicability of GPS data to the development of activity-based microsimulation models of travel demand and understanding relationships that govern activity-travel patterns in time and space. The travel demand modeling community should move forward with an aggressive research agenda that demonstrates the usefulness and value of the data and the ease with which the data can be reduced, synthesized, analyzed, visualized, and utilized in real transportation planning applications. The Commute Atlanta GPS data set and a few others appear to provide the sample size and richness needed to help accomplish this. The research agenda should focus on developing tools, analyzing relationships, studying behavioral patterns in time and space, and informing activity-based model specification.