PROMPTED RECALL TRAVEL SURVEYING WITH GPS

By

Joshua Auld, PhD Candidate
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St.
Chicago, IL 60607
Phone: 312-996-0962
Email: auld@uic.edu

Chad Williams, PhD Candidate
University of Illinois at Chicago

Kouros Mohammadian, Associate Professor
University of Illinois at Chicago

Using GPS technology in the collection of household travel data has been gaining importance as the technology matures. This paper documents recent developments in the field of GPS travel surveying and ways in which GPS has been incorporated into or even replaced traditional household travel survey methods, and details the development of a new internet-based prompted recall survey.

A new household activity survey is presented which uses automated data reduction methods to determine activity and travel locations based on a series of heuristics developed from land-use data and travel characteristics. The algorithms are used in an internet-based prompted recall survey which utilizes advanced learning algorithms to reduce the burden placed on survey respondents. The use of GPS data collection in place of traditional pen-and-paper of telephone assisted surveys allows for the survey to focus on more important and complex travel behavior questions, while automating the collection of traditional travel-pattern questions, such as routes used, locations selected, start and end times, and others, which have traditionally been somewhat difficult for survey respondents to answer accurately

The initial results of a small pilot study are discussed and potential areas of future work are also presented. A small scale initial study involving five individuals showed that the algorithms used can automatically determine location, travel times, and route choice with high accuracy while capturing additional travel behavior details, such as flexibility measures and planning times that are not usually captured in travel surveys. Overall, studies of this type should allow for easier, more accurate data collection, with a greater emphasis on collecting more behavioral data in addition to the usual travel pattern information.

1. INTRODUCTION

As travel demand modeling techniques and methods grow more sophisticated and data intensive there is a growing need for improved methods of data collection. New activity-based models tend to require data on the full activity-travel pattern of individuals and such hard to collect information as planning times and flexibility measures. As data needs have increased, more sophisticated methods of data collection have been developed, represented at first by the shift from travel to activity diaries and continuing on to the development of GPS enabled activity surveying. The use of GPS data collection has many advantages over traditional surveying methods. GPS surveys allow for a more exact representation of spatial and temporal data than respondents can typically provide and have been shown to correct significant trip underreporting errors associated with pen and paper or phone-based activity surveys (Battelle 1997, Wolf et al. 2004). Finally, by reducing the respondent burden through the use of automated activity type, location, timing and travel mode identification routines, GPS-based prompted recall surveys allow a larger number of more complex questions to be asked for a potentially longer duration.

This study attempts to build upon the survey techniques used in the past to determine activity-travel attributes. This work presents the design of a GPS-based prompted recall survey, which is implemented on a web server. The web-based program allows survey participants to upload collected GPS data at their leisure and generates an interactive prompted recall survey based on the uploaded data. Surveys of this type have the ability to capture a higher percentage of trips made by individuals with potentially more accurate timing attributes since the survey does not depend on the recall of the individual. Additionally, GPS-based activity surveys have the additional benefit of allowing full tracking of the routes selected by the individual for travel, information that was previously unattainable in a timely and efficient manner. This paper first describes previous efforts in the field of GPS-based surveying, including using GPS to provide trip rate corrections to activity diary surveys and attempts to completely replace the activity diary. The data reduction routines, including data cleaning and location finding algorithms are then presented. Finally the development of a prompted recall survey which incorporates machine learning algorithms to reduce respondent burden is documented.

2. PREVIOUS WORK IN GPS SURVEYING

The use of GPS data in activity and travel surveying is a relatively new practice, made possible through improvements in the technology itself and the demand for more accurate travel data. The use of GPS data began with a series of demonstration studies designed to prove the ability to use GPS for identifying activity-travel patterns, and has branched out to several more advanced applications in travel surveying. The growth in the use of GPS in household travel surveys has been enabled by the concurrent growth of the GPS technology and its capabilities, especially the increased accuracy gained by the removal of Selective Availability (SA) which added error to the broadcast GPS signal, as well as

developments in GPS receiver and battery technology. An overview of the capabilities of the GPS system for use in transportation can be found in Wolf (2004) and Stopher et al. (2006). Currently, most GPS surveys are conducted to provide trip rate corrections to traditional activity diary surveys. However, work is being done on using GPS to monitor changes in overall travel patterns, develop passive activity-travel diaries, and to generate interactive prompted recall activity-travel surveys. Previous work in these various fields is discussed in the following section.

2.1 Using GPS to Supplement Household Travel Surveys

GPS data collection has been used in transportation surveying for a relatively short amount of time. Initially, GPS data collection was used mostly to provide corrections for trip rates obtained from traditional household travel surveys or to demonstrate the feasibility of doing so. One of the first studies of this type was a proof of concept study which supplemented the Lexington, Kentucky MPO's household travel survey (Battelle 1997, Murakami and Wagner 1999). Successive surveys have tended to improve on some of the methodology, for example, using person-based tracking (Draijer et al. 2000), using a follow up prompted recall survey to determine factors causing trip underreporting (Wolf et al. 2004) or modeling the influences behind trip underreporting (Zmud and Wolf 2003, Forest and Pearson 2005, Bricka and Bhat 2006). Advances such as these have led to a more complete picture of travel behavior through the ability to capture all travel modes, and have led to more appropriate survey design by identifying causes of underreporting.

2.2 Replacing the Traditional Activity Diary with GPS Data Collection

Beyond using the GPS survey data to simply correct the results of a traditional household travel survey, there has been some effort to develop GPS based surveys to completely replace the household travel survey. One of the earliest examples of an attempt to replace the travel survey with passive data collection was conducted on a sample of passively collected GPS traces for 30 participants in Atlanta, Georgia (Wolf 2000, Wolf et al. 2001). Other studies have attempted to build on the process of diary reconstruction, by attempting to automatically identify trip purposes, travel modes or other travel attributes. A long-term passively collected set of GPS traces from Sweden has been used to automatically identify various travel attributes, including trip purposes and estimates of non-vehicle travel (Schönfelder et al. 2002, Axhausen et al. 2004). Similarly, Srinivasan et al. (2006) also developed an automated procedure for determining the basic trip attributes from passive GPS data streams. Finally, McGowen and McNally (2007) also developed an activity purpose model based on land-use and individual/household socio-demographic data. This model used classification and regression trees to predict activity type from the GPS data streams for highly disaggregate activity types. Work has also been done on automatically identifying travel modes, usually based on similar heuristic rules as in Srinivasan et al. (2006) and others.

2.3 Prompted Recall Activity Surveying

An alternative to using either electronic travel diaries with GPS, or using completely passive data collection with post-processing, is to use passive data collection with some type of follow-up survey. This is usually referred to as a prompted-recall survey, since the passively collected GPS data is used to generate a depiction of the trips and activities the individual pursued in order to remind the individual and prompt further responses. A variety of different prompted-recall surveys have been conducted, both vehicle-based and person-based which have used many different prompting strategies. The use of prompted-recall surveying has the advantage of not requiring any respondent participation during the trip, while also being able to capture very detailed information about many aspects of travel and activity participation which can not be automatically deduced. Prompted recall surveys are generally run at the respondent's convenience sometime after the data collection has been undertaken. A proof-of-concept study for prompted-recall surveying was undertaken by Bachu et al. (2001). This work used passively collected vehicle-based GPS data to track a sample of 10 households over a period of 2 or 3 days. Stopher et al. (2002) also performed a small pilot study using prompted recall survey methods with automated/manual trip identification. Much like the previous study the daily travel patterns were displayed on maps, but in addition, the travel patterns were also displayed sequentially in a tabular format, with unknown attributes left blank for the respondents to fill out, including the participants, trip purposes, travel costs, location names, etc. The respondents also validated the identified activities and added any stops that were missed. A similar method was used in the prompted recall portion of the Kansas City GPS study (Wolf et al. 2004). Proposals for other display types for the travel prompts and discussions of the potential strengths and weaknesses of each type were discussed by Doherty et al. (2001) and Lee-Gosselin et al. (2006), and the use of combined spatial and temporal displays was recommended.

A significant development over the initial prompted recall GPS studies was the move to internet-based surveys. As mentioned above, most of the early prompted recall studies involved creating maps or other displays, then mailing back to the respondents for completion, which could involve significant delays and therefore a potential loss of the respondent's ability to recall the travel patterns accurately. Therefore, work by Marca (2002), Stopher and Collins (2005), Lee-Gosselin et al (2006), and Li and Shalaby (2008) have been performed on using prompted recall surveying over the internet. All of these studies are designed to take place over the internet, so that in each case the individual would perform their daily activities and the data would later be transferred to a central server for analysis; either by direct uploading of the data removed from the device after the survey is complete as in the survey by Stopher and Collins (2005), or through continuous wireless communication as in Lee-Gosselin et al. (2006). In both cases, the data is processed to identify the activities and trips from the raw GPS data stream and the recall survey is built using the identified activity-travel episodes. A discussion of data processing techniques for person-based GPS studies can be found in Lee-Gosselin et al. (2006) and in a later section of this paper on the proposed GPS survey.

3. GPS DATA PREPARATION ALGORITHMS

In developing a new GPS-based prompted recall study, a method for reducing the log data into a meaningful form was first needed. The data preparation routines were designed to utilize GPS traces extracted from small portable GPS tracking devices. The data preparation routine uses new algorithms to clean the data, analyzes it to determine activity locations, and validates the results with queries to the user. Since the study tracks users continuously and through all travel modes, several data cleaning and analyzing routines were created to overcome challenges posed by this sort of data. This is especially true when attempting to distinguish walking travel from walking at an activity location. The survey attempts to correct for this through the use of built environment data and travel episode attributes. To reduce the raw GPS data to meaningful activity locations, a three stage process is used by the program which includes data cleaning, location finding and user verification. The first two stages take place with no user intervention as the data is uploaded to the program. The third stage is interactive with the user.

3.1 Initial Data Cleaning

The first step in determining activity locations is to clean up the initial data. This stage involves removing obviously incorrect points, caused by the well-noted urban canyon issues, signal loss, and signal straying. To clean up the data, two error-checking algorithms were developed. The first routine cycles through all the GPS points and evaluates the satellite fix characteristics, such as number of satellites and horizontal dilution of precision, as well as the travel speed to remove obviously incorrect entries. For each point, the distance and time between it and the previous point is calculated. If the speed calculated using these distance and time measures exceeds an upper limit threshold, currently set to 160 km/hr, then the point is eliminated and the next point is evaluated using the last valid point. This routine eliminates a common source of error, when the tracker strays during a travel episode or during a short duration activity.

3.2 Activity Location Aggregation Routine

Another significant challenge faced in using GPS traces to determine activity locations is in aggregating the recorded points to determine the actual activity locations. As opposed to many past GPS tracking studies, which were done only with in-vehicle units or with units that could not receive signals inside building, where locations were assumed at points where the signal was lost, this study tracks users through all travel modes and often captures traces from inside buildings as shown in Figure 1. For this reason, the locations could not be inferred from signal loss alone. A routine was therefore created to identify activity stops from the GPS data stream.

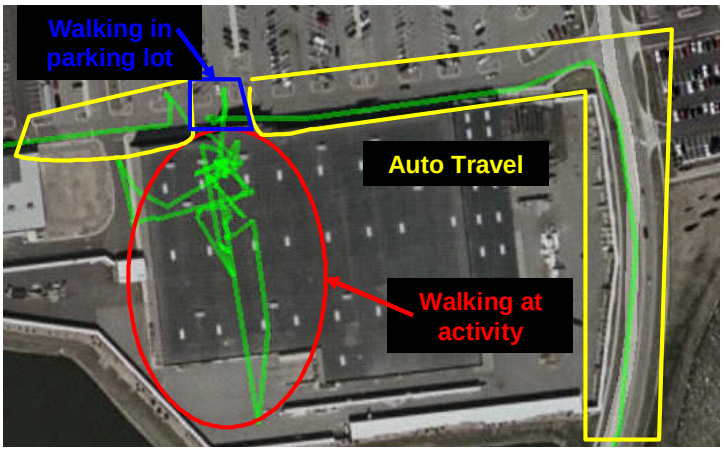

Figure 1. Example of GPS trace around an activity location

The basic clustering algorithm used in the study is fairly straightforward. The algorithm cycles through all of the cleaned GPS points, and when a point is found where the travel speed is lower than a predefined low-speed threshold, it is flagged for further analysis to determine if it is a part of an activity location. The basic location identification procedure sets a current point in the GPS data stream and searches through all subsequent points until the distance between the points exceed a threshold distance. If the individual was within the threshold distance for at least the threshold amount of time then the average of the points is used as the activity location. However, if the distance threshold is exceeded before the time threshold, or if any of the points exceed the low-speed threshold, then no activity is identified and the next point in the data stream is checked. The basic routine works for identifying many locations, but as the trace in Figure 1 shows, walking in the parking lot is indistinguishable from walking to the activity, if the walk mode was used. Therefore, when the walk mode is used, as is often the case in dense urban areas, the routine has a hard time distinguishing between the travel and activity episodes. This is not an issue for most activities in suburban areas where distances between activities tend to be large and the car mode is predominantly used. However, when activity spaces are very large another issue arises; one large activity can look like multiple small activities to the algorithm. An example of this issue is shown if Figure 2. In this figure, one activity shown in the graphic on the left portion of the figure occupies a space roughly the same size as the entire tour of activities shown on the right. In situations such as these, many sub-activities would be identified in the pattern shown on the left, while in actuality it represents one related activity. It is not desirable to question survey respondents about locations within the same activity. For this reasons, improvements were made to the routine to reduce the number of invalid activities.

First, it was observed that differences in urban form can have a great impact on the average size of activity locations. In order to set the distance and time thresholds in a meaningful manner, several rules were developed based on assumptions about activity spaces. The first assumption is that activity spaces are constrained by the block size of the area in which the activity is taking place. Therefore an average street block size measure for the area is used, which is defined as the total street length in the Census Tract divided by the number of intersections. This gives a measure of the average block size in

which activities are taking place. This measure is augmented with measures of the population and employment densities. The population density further constrains the activity space due to the observation that activity locations tend to be smaller in denser environments. The block size and densities are combined into one measure of activity space through a regression equation which models the average Census Block size within the tract, so that a smaller street block size or employment density or a higher population density leads to a smaller distance threshold. The following equation was estimated with an R^2 value of 0.86 to define the average activity space size for the Census Tract:

$$\sqrt{D} = -74.52 + 2.105\left(L_{road}/N_{\text{int}}\right) - 0.03903P + 0.04310E \qquad (1)$$

Where:

D = Average block size in Census Tract (in m^2)
L_{road} = Sum of roadway length in Census Tract (in m)
N_{int} = Number of intersections in Census Tract
P = Population density (in persons per km^2)
E = Employment density (in employees per km^2)

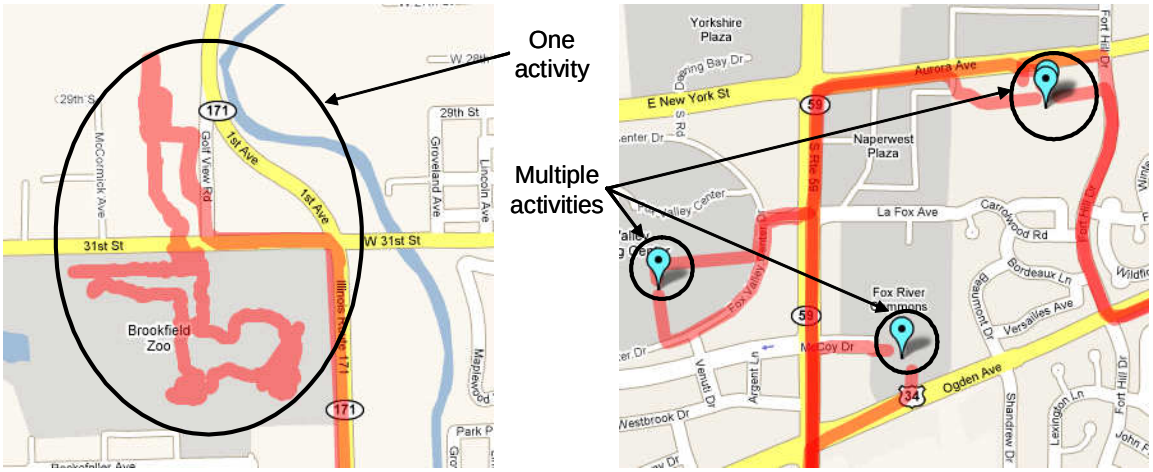

Figure 2. Large vs. small activity space

The final rules used to set the location search thresholds for distance and time involve the travel mode as distinguished by the travel speed. Two modes are defined in the algorithm, slow (less than 16 km/hr) and fast (over 16 km/hr), based on the highest expected likely pedestrian travel speed. Depending on the mode chosen the distance and time thresholds are varied.

After running through the cleaning and location finding algorithms, the results in the form of activity locations and travel episodes are stored in a database on the web server tagged to the individual participant. These results are then used to build a prompted recall activity survey for the participants to complete in order to gather more information on the full activity-travel context of the individual.

3.3 Initial Location Identification Algorithm Performance

To evaluate the current algorithm, a pilot test was run involving 5 individuals using GPS data loggers for an average of 8 days each. The data logger recorded the location, speed, distance and time information every 5 seconds while the device had a satellite fix. The GPS data was downloaded by the survey participants and run through a program which output activity patterns using the processing algorithms described above. The pilot test produced a total of 220 activity observations. For each observation day, the participants were asked to observe each activity location identified by the program and determine if they represented actual activity locations. Afterwards, the participants were asked to enter the number of activities that the program missed. The numbers of valid, invalid and missed activities were then later used to evaluate the performance of the algorithm. During the course of the pilot study only 5 activities were identified as missed by the participants, while 28 of the 220 identified activities were marked as invalid. Comments made by the individuals seemed to indicate that the missed activities were generally due to failure of the device to acquire a signal. The recall of the initial survey test was found by dividing the valid identified activities by the total valid activities, which gave a recall of over 97%. Additionally, with only 28 invalid activities the algorithm had a precision of 87% which appears to be an acceptable number, i.e. not requiring too much processing by the individual to correct the activity-travel pattern. Based on the initial pilot test, the algorithm appears to successfully minimize the number of extraneous activity locations while simultaneously capturing almost all of the actual activities.

4. GPS SURVEY DESIGN

After the development of the activity and travel episode identification algorithm was completed, the routines were incorporated into an internet-based prompted recall survey. As mentioned before, a prompted recall survey combines the ease of use of passive data collection efforts with the detailed data on activity and travel attributes captured from a follow-up survey. The prompted recall survey is especially important for collecting information on attributes which are not automatically identified, such as participants in an activity, planning horizons, schedule flexibility measures, and many of the underlying reasons for decision making. However, much of the work done on automated travel diary creation is useful for reducing the number of questions needed in the survey, so many of these routines are incorporated into the overall prompted recall survey design. The following survey is designed in ASP.NET and JavaScript and utilizes the Google Maps API. Data is uploaded through the internet survey site and stored on a server. The following section describes the various components of the internet based survey design.

4.1 Activity and Travel Verification by Participants

An important component of the survey is the verification of the automatically identified activity and travel locations by the survey participants themselves. Although the current algorithm performs very well in identifying the activity locations, with approximately 97% accuracy and 87% precision in pilot tests, there are still some errors associated with

signal losses due to signal acquisition delays or user error, bad satellite fixes and occasional failures of the location finding algorithm. Therefore it is important to allow the users to both remove activities which did not actually occur and to add activities which where missed for any of the above reasons. Upon uploading of the GPS logger data and completion of the automated data reduction routine, a display like that shown in Figure 3 is generated using the Google Maps API.

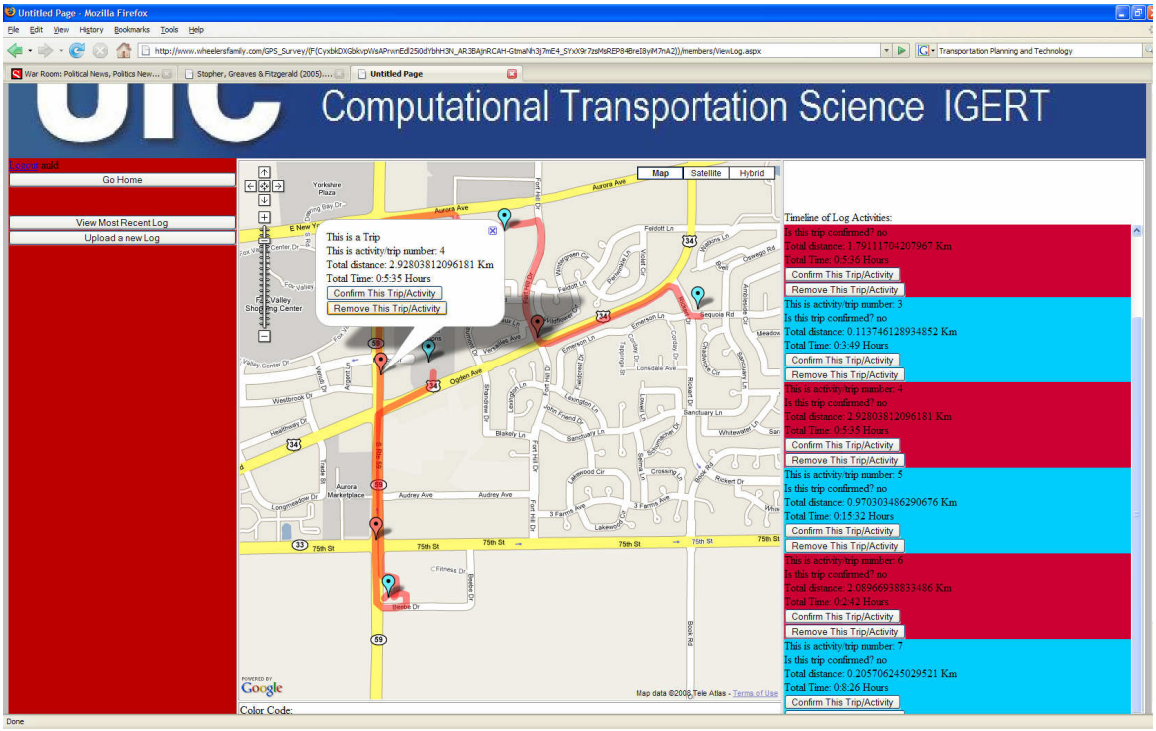

Figure 3. Activity-Travel User Confirmation with Map and Timeline

The activity locations and travel routes stored in the participant's database are drawn on the map and the users are then asked to confirm or remove each episode. This interface presents a familiar display to many users and is generally fairly easy and intuitive to use. The map display allows the user to drag the activity pins to correct errors with the calculated location and also to correct errors associated with the identified start and end times. The map display is also linked to a timeline display. The use of a map linked to a timeline gives the users a more complete spatiotemporal picture of their activity pattern and allows for simpler correction of the schedule, i.e. correcting locations on the map and start and end times on the timeline.

4.3 Surveying Activity and Travel Attributes

After the verification stage is completed, the activity-travel survey is started. The survey consists of a series of questions concerning either attributes of the activity episode, or for travel episodes questions about mode and route choice decisions. The questions are paired with a map display similar to that shown during the confirmation portion of the survey, except that in this stage only the activity or travel episode in question is shown on

the display to jog the individual's memory of that episode. The questions are divided into four basic groups for activities and two for travel episodes. The questions for activity episodes involve either the activity type, individuals participating in the activity, the location of the activity and the timing of the activity. For travel episodes the travel route is displayed in the map window and questions regarding either mode choice or route choice decisions are asked.

One of the major underlying goals behind the study is to capture the underlying process and dynamics of activity patterns. For this reason many of the questions asked relate to decision timing, i.e. when the attribute was planned, underlying reasons for making decisions as in the location selection and mode/route choice questions, or flexibility variables relating to individual participation, timing or location decisions. These values are fundamental to modeling efforts which attempt to describe the actual cognitive processes underlying activity-travel decision making (Doherty et al. 2004). Furthermore, running the survey over long durations allows descriptions of how these processes may change over time or in different contexts.

The first set of questions regarding the activity type and activity participants are shown in Figure 4. The individual first selects the type of activity (or multiple types in the case of multi-purpose stops) from a list of standard activity types. Currently, only the purpose for the out-of-home activities is captured in the survey, while in home activities are simply listed as "At Home". In addition, the individual selects a planning time horizon for this activity, which is when the decision to undertake this activity was made. The individual can choose from a variety of impulsive to pre-planned time horizons as well as "Routine" and "Unknown" options. If the activity type was chosen as "At Home" then the remaining questions about the activity are ignored. For all other activities, the "who with", location and timing questions are also asked

For the "who with" questions, the respondent selects the number of involved persons, their relation to the respondent and the interpersonal flexibility associated with the activity, i.e. whether the participation of others was required or not. Location choice is another important component of the survey. The individuals are asked how many locations are generally available to them for performing this activity, the reason for choosing the selected location as well as the planning horizon for the location decision if it is different from the timing of the participation decision, (e.g. I need to go shopping tomorrow, vs. I need to go shopping tomorrow at Wal-Mart). Finally, some questions about the timing decisions surrounding the activity are asked. The planning horizon for the timing is selected in the same manner as for the location decision, again if it is different than the participation decision planning horizon (e.g. I will go shopping tomorrow vs. I will go shopping tomorrow at noon). Additionally, general flexibilities of the start and end times are selected. So for each attribute of the activity some basic descriptors are collected and then planning horizon and flexibility values are input which should further improve understanding of the underlying activity-travel pattern creation process.

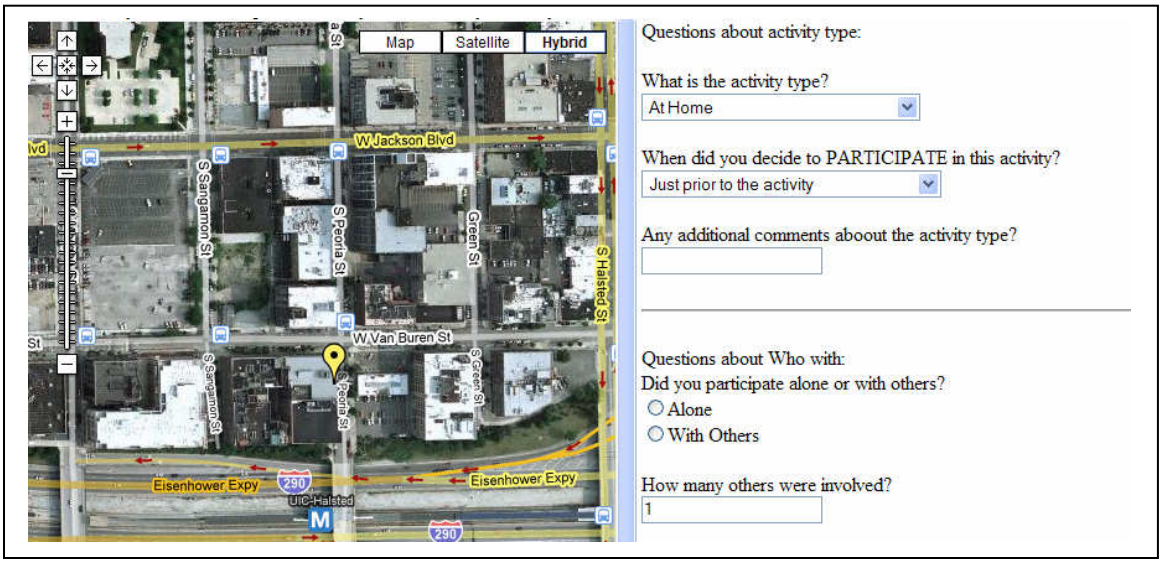

Figure 4. Activity Attribute Questions

The preceding discussion relates to survey questions asked about the attributes of the activities that the respondent engaged in. However, it is also desired to capture some of the decision processes that lead to mode and route choice decisions, which is possible through the use of GPS data collection in a way that has rarely been possible in traditional surveys. Because the exact route selected, travel time and distance traveled is known for each trip, it is relatively straightforward to display this information to the individual on a map and question them about the decisions that lead to the given outcome. In the current survey, for each trip identified in the validation stage, the mode and route choice questions shown in Figure 5 are asked of the respondents. The individuals choose the planning times and underlying reasons for both the mode and route choice decisions. These results, when coupled with activity type, flexibility measures, and other process data, can help to further understand mode and route choice behaviors in the full activity travel context of the individual.

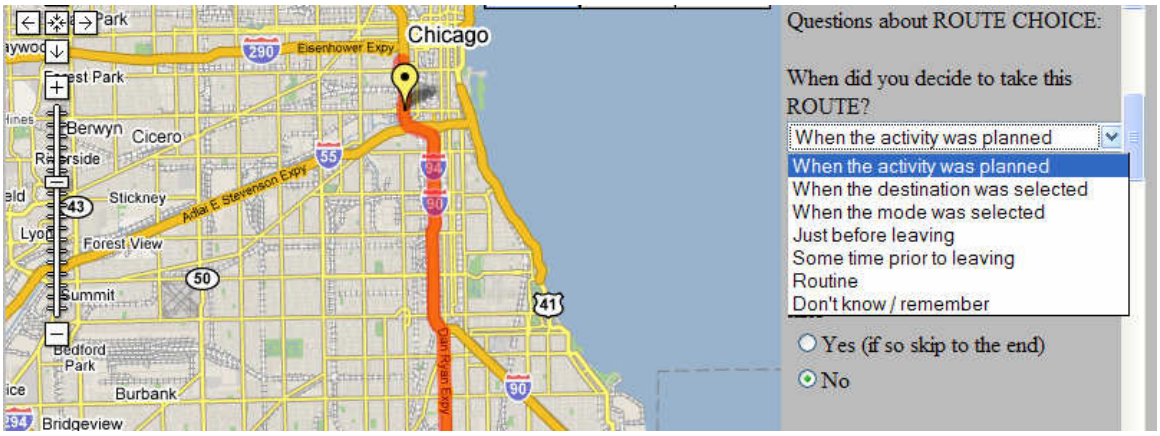

Figure 5. Route choice decision-making questions

This section, combined with the previous data preparation section, has described the processes of creating a fully-automated prompted recall diary from data colleted using GPS data loggers. However, running the survey as described without any further modification would still involve fairly significant respondent burden. In fact, during the pilot study, respondents identified many areas which were felt to be especially burdensome. It was felt that these shortcomings could inhibit the use of the survey for longer term data collection. The next section describes some techniques used to address these shortcomings.

5. REDUCING RESPONDENT BURDEN THROUGH LEARNING

In order to further reduce the burden placed on survey respondents to enable longer duration surveys, the frequency and type of questions asked of participants needs to be significantly reduced. Some routines have been developed to accomplish this, as discussed previously, by automatically detecting some attribute which would negate the need for questioning the individual. Examples of these routines include automated trip purpose detection (Wolf 2000) and mode identification (Tsui and Shalaby 2006). However, these procedures are less applicable for most other attributes which are required from the current survey. It is not obvious, for example, how planning horizons, decision variables and other attributes such as involved persons could be derived from the GPS/GIS data alone. Therefore, a learning approach is needed, which utilizes information already collected in the survey to develop patterns which can predict the various activity-travel attributes. This section discusses some background in machine learning and some ways in which it has been applied in travel pattern prediction as well as propositions for using it to help reduce survey respondent burden.

In data mining and machine learning related work, techniques for learning sequences, referred to as sequential associative mining, have been extensively studied since the original associative mining and later sequential mining techniques were introduced (Agrawal et al. 1993, Agrawal et al. 1995). Identifying patterns in traditional associative mining relies on multiple training sets for its primary constraint support. With associative sequence mining, there is a similar dependency on multiple training sequences.

One of the few examples of using learned patterns to reduce respondent burden within an actual survey occurs within the ANNE survey developed by Marca et al. (2002). In the initial development of the survey, answers to previous activity-location questions were stored and later used for future activity locations to estimate likely activity types based on either the distance and time difference from currently labeled points, or later to develop an activity-type probability distribution over the survey area. This allowed likely responses to be suggested to the user and also was suggested for use in what the authors termed “focused questions”, where users are only asked about activity locations which are not known with high probability. Currently there are no activity-travel surveys which utilize data mining techniques in the survey development that the author’s are aware of.

In GPS-based prompted recall surveys, understanding the context of a traveler and being able to predict their likely next step can be used to help reduce participant burden in the form of data entry requirements. Depending on the goals and participant willingness, there are two different ways these predictive models could be applied: auto population or selective querying.

For question auto population, the predictive model would be applied and questions about activity or travel could be pre-populated based on the user's prior history to be confirmed or changed by the participant. Consider a scenario where a five minute stop on the way to the train station was identified in the GPS data. If the participant's prior activity-travel pattern showed they occasionally stopped in this location for coffee, this information could be used to auto populate the activity type, the end time flexibility, and the likely planning horizon for the activity without the participant needing to enter it manually. For longer term surveys, this type of predictive model could be incorporated to reduce the number of questions asked in a selective querying strategy. Two possible approaches would be high confidence elimination or key event querying. The principle behind high confidence elimination is to eliminate any question where the confidence that the answer is known is over a certain threshold. An alternative to this more suited for longer term surveys would be to only ask about activities or travel that are unusual compared to known patterns. In both of these approaches, while the participant still has a significant burden early on, as the survey progresses their burden is reduced as the application learns their behavior. While learning patterns specific to a participant are valuable, due to the amount of time necessary to observe these trends, augmenting the data with the patterns of others can likely help to reduce the initial learning time.

These learning models can therefore be used to either assist or completely replace the data entry requirements of the respondent. Depending on the length of the survey and the types of attributes required, this can help to significantly reduce the respondent burden, although as mentioned the burden during the initial phase of the survey could still be somewhat large as the algorithms learn the user's likely activity-travel patterns. However, this could further be reduced through the use of a well designed up-front survey of the person, which in addition to capturing socio-demographic information could also be used to identify common locations visited and routines within the respondent's usual activity-travel pattern.

6. CONCLUSIONS AND FUTURE DIRECTIONS FOR RESEARCH

This paper has presented the design of a new web-based prompted-recall activity-travel survey. The survey addresses many issues associated with GPS travel surveying and attempts to overcome much of the difficulty associated with person-based GPS tracking. The survey portion of the work was designed to reduce respondent burden to a minimum level in order to enable longer-term studies, which are essential for capturing the dynamics of activity-travel decision making. Initial results show that the activity location identification algorithms perform well, however, much work remains in evaluating and improving the activity survey portion of the work. The use of this type of survey also

presents new opportunities and avenues for future work in both improving survey design and developing new applications for the collected data.

The most important remaining step in the development of this work is to evaluate the effectiveness of both the algorithm and the survey burden reduction strategies during an actual implementation of the survey. The initial pilot sample for evaluating the data preparation algorithms was very small, incorporating only a total of 197 actual activity episodes. This algorithm needs to be evaluated over a wider range of subjects for longer time periods. The effectiveness of the learning algorithms on reducing the respondent burden also remains to be evaluated. These learning algorithms were incorporated with the specific goal of making survey participation less onerous and hopefully increasing the completion rate for the survey, the retention rate of participants and the duration for which the participants are willing to participate. The effectiveness of the current survey design in achieving these goals, and potential areas of improvement, therefore remain to be investigated.

Another related area for potential research involves the design of the survey questions. More work is needed on identifying the types of questions that can be asked, and which are most effective at eliciting the desired information without being overly complex or confusing to participants. In addition, attempts should be made to determine the possibility of collecting pre-planning data as in the CHASE data collection effort (Doherty et al. 2004) in combination with the activity-travel survey, which would give a more complete picture of the dynamics of the activity scheduling process as suggested by Doherty et al. (2001).

Beyond evaluating the actual survey design, further work is needed in identifying the ways in which data collected from such surveys can be used. One potential is to use the data collected activity location and route choice decisions to investigate the formation of mental or cognitive maps (Golledge and Garling 2004), which could greatly enhance the realism of travel choice models. If the time-frame of the survey is extended long enough, a significant portion of the common places in the persons mental map are likely to be visited. The perceptions about quality, distance, etc. relating to the route or activity locations of the individual can be compared to reality to generate models of individual's perception and mental map formation. Additionally, how the individuals learn and perceive their environment over time can also potentially be observed and the data collected can contribute to modeling of these processes. Knowledge gained during studies of these various processes can then be fed back to improve the overall survey design.

It is clear that much work remains in developing and validating long-term prompted recall surveys using GPS data collection methods. The processing algorithms described in this paper represent important advances in automating the process of activity location identification. In addition, several innovative surveying techniques involving learning algorithms have been presented in order to further reduce respondent burden. Data collection efforts of the type described here should help to further improve knowledge of the dynamics of household activity and travel decisions.

REFERENCES

Agrawal, R.; Imieliński, T. & Swami (1993). A. Mining association rules between sets of items in large databases, *SIGMOD Rec.*, ACM Press, 1993, 22, *207-216*.

Agrawal, R. and Srikant, R. (1995) Mining sequential patterns, *Eleventh International Conference on Data Engineering*, Yu, P. S. & Chen, A. S. P. (ed.), IEEE Computer Society Press, 1995, *3-14*.

Axhausen, K.W., S. Schönfelder, J. Wolf, M. Oliveira and U. Samaga (2004). Eighty Weeks of GPS Traces: Approaches to Enriching Trip Information. *Proceedings of the 83th Annual Meeting of the Transportation Research Board*, January 2004. Washington, D.C.

Battelle Transport Division (1997). *Lexington Area Travel Data Collection Test*, Final report prepared for the FHWA.

Bricka, S. and C.R. Bhat (2006). Comparative Analysis of GPS-Based and Travel Survey-Based Data. *Transportation Research Record: Journal of the Transportation Research Board*, 1972, 9-20.

Doherty, S., Noel, N., Gosselin, M., Sirois, C. & Ueno, M. (2001). Moving Beyond Observed Outcomes – Integrating Global Positioning Systems and Interactive Computer-Based Travel Behavior Surveys. *Transportation Research E-Circular E-C026*, March 2001.

Doherty, S. T., E. Nemeth, M. Roorda and E. J. Miller (2004). Design and Assessment of the Toronto Area Computerized Household Activity Scheduling Survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1894, 140-149.

Draijer, G., N. Kalfs and J. Perdok (2000). Global Positioning System as Data Collection Method for Travel Research. *Transportation Research Record: Journal of the Transportation Research Board*, 1719, 147-153.

Golledge, R.G. and T. Garling (2004). Cognitive Maps and Urban Travel, in: D.A. Hensher, K.J. Button, K.E. Haynes and P.R. Stopher, eds, *Handbook of Transport Geography and Spatial Systems*. Oxford: Elsevier.

Lee-Gosselin, M.E., S.T. Doherty. and D. Papinski (2006). An Internet-based Prompted Recall Diary with Automated GPS Activity-trip Detection: System Design. *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, January 2006. Washington, D.C.

Li, Z.J. and A.S. Shalaby (2008). Web-based GIS System for Prompted Recall of GPS-assisted Personal Travel Surveys: System Development and Experimental Study. *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, January 2008. Washington, D.C.

Marca, J.E. (2002). *The Design and Implementation of an On-Line Travel and Activity Survey*. Center for Activity Systems Analysis. Paper UCI-ITS-AS-WP-02-1. http://repositories.cdlib.org/itsirvine/casa/UCI-ITS-AS-WP-02-1.

Marca, J.E., C.R. Rindt and M.G. McNally (2002). *Collecting Activity Data from GPS Readings*. Center for Activity Systems Analysis. Paper UCI-ITS-AS-WP-02-3. http://repositories.cdlib.org/itsirvine/casa/UCI-ITS-AS-WP-02-3.

McGowen, P. and M. McNally (2007). Evaluating the Potential to Predict Activity Types from GPS and GIS Data. *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, January 2007. Washington, D.C.

Murakami, E. and D.P. Wagner (1999). Can using global positioning system (GPS) improve trip reporting?. *Transportation Research Part C., 7, 149-165.*

Murakami, E., J. Morris and C. Arce (2003). Using Technology to Improve Tranport Survey Quality. *Transport Survey Quality and Information*. Elsevier, Oxford.

Schönfelder, S., K.W. Axhausen, N. Antille and M. Bierlaire (2002). Exploring the potentials of automatically collected GPS data for travel behaviour analysis A Swedish data source, in J. Möltgen and A. Wytzisk, eds. GITechnologien für Verkehr und Logistik, IfGIprints, 13, 155-179.

Srinivasan, S., P. Ghosh, A. Sivakumar, A. Kapur, C.R. Bhat and Stacey Bricka (2006). *Conversion of Volunteer-Collected GPS Diary Data into Travel Time Performance Measures: Final Report*. Center for Transportation Research at The University of Texas at Austin, February 2006.

Stopher, P. P. Bullock and F. Horst (2002). *Exploring the Use of Passve GPS Devices to Measure Travel.* Institute of Transport and Logistics Studies, Paper ITLS-WP-02-06, University of Sydney.

Stopher, P. and A. Collins (2005). Conducting a GPS Prompted Recall Survey over the Internet. *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, January 2005. Washington, D.C.

Stopher, P., C. FitzGerald and J. Zhang (2006) *Advances in GPS Technology for Measuring Travel.* Institute of Transport and Logistics Studies, Paper ITLS-WP-06-15, University of Sydney.

Tsui, S.Y.A. and A.S. Shalaby (2006). An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, 1972, 38-45.

Wolf, J. (2000). *Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data.* Dissertation, Georgia Institute of Technology, School of Civil and Environmental Engineering, Atlanta, Georgia, July 2000.

Wolf, J., R. Guensler and W. Bachman (2001). Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data. *Proceedings of the 80th Annual Meeting of the Transportation Research Board*, January 2001. Washington, D.C.

Wolf, J. (2004). Defining GPS and its Capabilities, in: D.A. Hensher, K.J. Button, K.E. Haynes and P.R. Stopher, eds, *Handbook of Transport Geography and Spatial Systems.* Oxford: Elsevier.

Wolf, J., S. Bricka, T. Ashby and C. Gorugantua (2004). *Advances in the Application of GPS to Household Travel Surveys.* Presented at the Transportation Research Board National Household Transportation Survey Conference, Washington D.C.

Zmud, J. and J. Wolf (2003). Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study. *Proceedings of the 10th International Conference on Travel Behaviour Research*, August 2003, Lucerne, Switzerland.