Analysis of LBS-Derived Data for Social Scientists: Prospects and Limitations

Mei-Po Kwan
Department of Geography, Ohio State University
kwan.8@osu.edu

As the mobile positioning technology used by providers of location-based services (LBS) is still evolving and non-standardized, it is difficult to predict the kind of technology that will dominate the LBS business. But given the need of wireless carriers to meet the federal E911 mandate and for cost effectiveness, it can be assumed that only fairly accurate and cost effective solutions will be adopted by LBS providers in the near future. For social scientists interested in understanding human behavior in space-time and its complex relationship with the urban environment, the possibility of collecting and using data derived from LBS offer new opportunities and pose many challenges at the same time.

Recent studies that use GPS data of individual trips indicate that LBS-derived data may allow the accurate reconstruction of individuals' activity-travel patterns with considerable space-time details. The reconstructed patterns (as revealed preference of travelers) can be used for various purposes, including the study of route planning and route choice behavior, activity scheduling involving complex space-time trade-offs, the daily and weekly temporal rhythms of activities or trips (e.g. cycles and repetitions), and responses to real-time information as in the context of Intelligent Transportation Systems (Kwan, 2000; Zhou and Golledge, 2000). Further, as the precise space-time trajectories of walking trips can be recorded, transport problems of the mobility-impaired population subgroups - such as the vision-impaired or the elderly people - can be analyzed in great space-time details.

LBS-derived data may also enable the development of a whole range of new analytical and modeling methods useful to social scientists. Given the recent development and application of GIS and spatial statistics in many areas of human spatial behavior, it is now possible to analyze and model individual travel behavior in relation to a realistically represented urban environment - in the form of a comprehensive geographic database of the study area that includes information about all land parcels and all street segments (including segment-specific travel speed and turn restrictions). GIS-based geocomputation and 3D geovisualization of activity-travel diary data are examples of techniques useful for the analysis of LBS-derived data.

Several difficulties still remain for future studies that attempt to use LBS-derived data. First, although the network-based or hybrid positioning technology used by most LBS providers can achieve a positional fix faster and easier than conventional GPS-based technology, discontinuities in LBS-derived data cannot be entirely eliminated. This is due to the problem of "loss of fix" that can happen under certain circumstances (e.g. loss of signal inside certain areas of a building). As past studies indicate, dealing with this data discontinuity problem may require considerable time and effort during the data preparation phase of a study. Second, although comprehensive data can be collected if all of the

activities and trips an individual undertakes throughout the survey period are recorded, users of LBS may not keep their devices on throughout the day for various reasons. This will introduce gaps in the LBS-derived data and make it difficult to achieve completeness in LBS-derived data. Third, LBS devices normally do not record what the user is doing at a particular time, the purpose of the activity or trip, the travel mode used, and other simultaneous activities. Methods need to be developed for recording and linking these data with the location data collected by LBS devices. Without these accompanying data that provide information about the nature and context of an activity or trip, the usefulness of the raw LBS-derived data will be limited for social scientists. Fourth, like other high technology services, LBS will tend to be used mostly by people with high income and education level. LBS-derived data may not be very useful for understanding the activity patterns of other population subgroups.

There are difficulties in the analysis of LBS-derived data due to the computational intensity of processing and visualizing large space-time datasets. For instance, the original GPS data file for 100 households collected in the Lexington study contains 794,861 data points of latitude-longitude pairs and time (Murakami and Wagner, 1999). It takes up about 230 megabytes of disk space in the format provided on the data CD. Manipulating files of this size can be taxing for the computer hardware normally available to social and behavioral scientists. Although improvement in computing power in the near future will reduce this problem, much research is still needed to develop more efficient algorithms and data manipulation methods for handling large LBS-derived datasets. For example, instead of including all location points, procedures for reducing and generalizing LBS-derived data can be developed to reduce the problem of storage and computational intensity (Zhou and Golledge, 2000). New geocomputational and geovisualization methods also need to be developed when this LBS-derived data are used with large and complex geographic datasets.

References:

Kwan, Mei-Po. 2000. "Evaluating gender differences in individual accessibility: A study using trip data collected by the Global Positioning System." A Report to the Federal Highway Administration (FHWA), U.S. Department of Transportation, 400 Seventh Streets, S.W., Washington, D.C. 20590.

Murakami, Elaine and David Wagner. 1999. Can using global positioning system (GPS) improve trip reporting? *Transportation Research C*, 7:149-165.

Zhou, Jack and Reginald Golledge. 2000. "An analysis of household travel behavior based on GPS." Paper presented at the Ninth International Conference on Travel Behaviour Research, the International Association of Travel Behavior Research (IATBR), Gold Coast, Australia, July 2000.