A New Framework for Development of Time Varying OD Matrices Based on Cellular Phone Data

Corresponding Author:
Jingtao Ma, Mygistics Inc.
9755 SW Barnes Rd Suite 550 Portland, OR 97225
Phone +01-503-575-2191, email: jma@mygistics.com

Co-Authors:
Fang Yuan, Delaware Valley Regional Planning Commission (DVRPC)
190 N. Independence Mall West, 8th Floor, Philadelphia, PA 19106-1520
Phone +01-215-592-1800, email: fyuan@dvrpc.org

Chetan Joshi, PTV America, Inc.
9755 SW Barnes Rd Suite 550 Portland, OR 97225
Phone +01-503-297-2556, email: cjoshi@ptvamerica.com

Huan Li, Mygistics Inc.
9755 SW Barnes Rd Suite 550 Portland, OR 97225
Phone +01-503-575-2191, email: hli@mygistics.com

Thomas Bauer, Mygistics, Inc.
9755 SW Barnes Rd Suite 550 Portland, OR 97225
Phone +01-503-575-2191, email: tbauer@mygistics.com

Submission date: November 16, 2011

Word count: 2056

Figures: 4
A New Framework for Development of Time Varying OD Matrices Based on Cellular Phone Data

Abstract:

This paper introduces the methodology and its pilot study in Sacramento, CA that derives the synthetic operational OD matrices from best integrating the mobile patterns revealed by cell phone use activities, socio-economic data and traffic engineering expertise. The high cell phone penetration in the population and their ubiquitous usage in daily and travel activities lend the unique advantage of continuous observation of origin-destination and trip tour information to the greatest geographical coverage from the largest ever sample sizes. This data and the analysis system can find a wide range of use cases in transportation engineering studies, for example, integrated corridor management, regional dynamic traffic assignment, long-distance, inter-regional, external-external travel data, add-on survey tools just to name a few and many more.

Keywords:
AirSage, MobileOD, Traffic analysis zone (TAZ), Demand estimation
Introduction

Estimating time-varying traffic demand has always been a challenge for planners to avoid possible biased decisions made from distorted network traffic flow patterns. This is especially true in operational planning studies; for example, integrated corridor management (ICM) strategies are usually backed by macro-meso-microscopic multi-resolution modeling analysis where time-dependent traffic demand is a vital input. To obtain the time-varying demand, the common practice is to take the origin-destination (OD) tables from the regional travel demand model as the pattern and apply OD estimation techniques with observed traffic counts to come up with the demand for different time increments. Relatively easy to apply, this practice is limited largely by the demand patterns in the travel demand model that is more focused on travelers’ spatial distributions than the time variations of their trips [1]. This is also echoed in a recent presentation [2] that summarized over twenty ICM project practices in the US. According to this summary, key reasons why model estimates deviated from targeted traffic operations in the real world may include: 1) trip tables obtained from outdated travel demand models; 2) coarse temporal resolution not reflecting daily traffic fluctuations; 3) daily traffic variations by nature, not even to mention seasonal variations, perturbations from ball games or special events, weather and other factors. To overcome these limitations, planners are calling for more innovative methods to understand the travelers’ mobility pattern in finer resolutions both spatially and temporally under their interested operational conditions.

The first classes of methods involve setting up data collection systems to sample all or a portion of drivers when they enters and leaves the study cordon or the cross sections. Both automated number plate recognition (ANPR) and bluetooth collection belong to this classes of active probing methods. As the origin-destination obtained from these methods are accurate in general [3], both ANPR and bluetooth technology based methods are usually on a case by case studies for a corridor or a subarea where the number of entrances and exits are limited.

In the context of increasing mobile society, transportation professionals are trying to leverage the “digital footprints” travelers left with their in-vehicle or personal navigation devices (PND) or cell phones. This paper documents the procedures that were applied to utilize the AirSage mobile phone data to estimate the traffic demand for operational planning studies in the Interstate 80 and State Route 65 Interchange area in Roseville, CA.

AirSage Mobile Phone Data

AirSage [4] analyzes wireless network signaling data to determine the location and movement of mobile devices. Working in partnership with national wireless carriers, AirSage data provides a continual network-wide view of all mobile devices on the network. The AirSage platform covers the entire Sprint CDMA network in the US (approximately 35 million mobile devices), and in Q4 of 2011 will add the Verizon network (approximately 90 million devices).

Each time a mobile device interacts with the network, a data record (which AirSage calls a “mobile sighting”) is generated by the network switching equipment. Each mobile sighting record contains a variety of network signaling data elements. Multiple types and sources of mobile sightings (e.g. starting/ending a voice call, cell handoffs, sending or receiving SMS messages, engaging in a data session, etc.) are received by AirSage as a continual raw data stream from the mobile network. The AirSage data platform aggregates the raw data, anonymizes the data stream, and uses advanced multi-lateration techniques to derive an anonymous, time-stamped location for each mobile sighting.
Figure 1: Sample mobile phone (Sprint) trajectories mapped to the transportation network, 600 randomly selected subscribers from over 128 thousand in the entire dataset.

**Methodology**

Trajectories of the mobile devices are analyzed through a series of data mining procedures and identified as vehicle trips. These trips are represented by the trip start/end locations and the departure times.

The applicable traffic analysis zone (TAZ) structure is then imported for the study area. Together with other 2010 US Census Bureau Block Group definitions for adjacent counties, this zoning structure serves the basis to map the identified vehicle trips onto the transportation network. These origin-destination flows identified from mobile phone activities (“MobileOD”) are the basis for generating the seed OD matrix of traffic demand estimation.

Figure 2 shows a visualization of the aggregate OD flows derived from the vehicle trips based on mobile phone traces as “Desire Lines”.
Through analyzing the mobility patterns of an anonymized device over a relatively long period (e.g., one week/month), AirSage is able to determine the home and work locations (at TAZ level) based on certain assumptions (e.g., 7PM-7AM activities usually in home). This information is used to analyze the trip characteristics at the area or regional level for both quality assurance/quality control and demand projection purposes.

Traffic count information serves as the target for demand estimation, the demand estimation method is based on the weighted entropy maximization and its specifics are beyond the scope of discussion of this paper. This data could come from various sources, for example tube counts, turning movement counts at intersections as well as loop detectors. The count data is assembled and analyzed and cleaned for any discrepancies and noise. Once the traffic count data is cleaned, it is used as the basis for application of demand estimation methods, with constraints such as trip length distribution and maximum allowable scaling factors. This process is carried out from each of the time periods for which demand is to be estimated by using the appropriate mobile phone based seed matrices and counts for that time period.
Figure 3: The flow diagram for time varying demand estimation

1. Translate data from mobile phone database to a spatial database based on the coordinates of the sightings.

2. Import traffic analysis zones (TAZ) polygons as a separate layer.

3. Perform a spatial intersect to attach mobile phone trace trip ends to the appropriate origin and destination TAZ.

4. Import, analyze and clean traffic count data for each time period.

5. Apply synthetic demand estimation procedures to the seed matrix.

6. Validate results.
Application Case Study

As an application or use case, the methodology discussed above was applied to a project in Sacramento area in California. The study was conducted in collaboration with a consulting firm active in the local area. In the month of October 2010 (00:00:00 Oct 1st 2010 to 23:59:59 Oct 31st 2010), a total of 256 million (255,828,842) valid sightings were recorded and retrieved for this study. For the purpose of operational OD inference, a mobile device visibility rule was established to keep the data of only those devices that leave sightings at least once during every hour of the study periods (6-10AM & 3-7PM). The filtered dataset had a total of 98 million (98,333,324) sightings, representing the mobile phone activities of over 128 thousand (128,185) individual Sprint users. This was adopted as the study dataset. In total, 3.6 million trips were identified from this study dataset.

The consultant team provided a customized traffic analysis zone (TAZ) structure for Placer County, California where the study area is located. Together with other 2010 US Census Bureau Block Group definitions for adjacent counties (e.g., Sacramento, Yolo), the zoning structure serves the basis to map the identified vehicle trips onto the transportation network.

Traffic counts for this study were derived from three main sources: the PeMS permanent VDS for freeway and ramps for the Caltrans District 3, the turn counts from the City of Roseville websites for major intersections, and the traffic counts shared by the consultant team at a number of important segments.

The PeMS freeway counts and the City of Roseville turn counts were all downloaded and aggregated to hourly increments. The hourly counts were averaged over all Oct 2010 weekday values as the input to the demand estimation. The segment counts from the consultant were also averaged only for the weekdays.

The road network for traffic demand estimation was prepared from NAVTEQ tiles. Only the roads of Functional Class 1-4 were retained for the study. The following excerpt from NAVTEQ GeoDatabase Reference Manual explains the definition of different FCs:

- **Functional Class** = 1 roads allow for high volume, maximum speed traffic movement between and through major metropolitan areas.

- **Functional Class** = 2 roads are used to channel traffic to **Functional Class** = 1 roads for travel between and through cities in the shortest amount of time.

- **Functional Class** = 3 is applied to roads which interconnect **Functional Class** = 2 roads and provide a high volume of traffic movement at a lower level of mobility than **Functional Class** = 2 roads.

- **Functional Class** = 4 is applied to roads which provide for a high volume of traffic movement at moderate speeds between neighborhoods. These roads connect with higher functional class roads to collect and distribute traffic between neighborhoods.

The Navteq network was modified at critical locations to reflect the recent geometric changes in the area. The link types and speed variations were also checked against online mapping services such as Google Maps and Street Views.

To derive and map the MobileOD, the entire Sacramento/Roseville regional network was used. However, a sub-network that is sufficiently large to incorporate the study area was eventually used to help focus only on the traffic of the study area. External zones were created considering both transportation network (e.g., freeways at the network boundary) and the mobile network; this was to better align the Mobile network for mapping the MobileOD and resulting seed matrices.
The demand estimation process was performed on each hour of the study period (6-10AM, 3-7PM). It is well understood that any OD estimation tools, when abused, could lead to “over-fitting” to traffic counts. Therefore, this demand estimation process was performed only once to the seed matrices so that the overall traffic patterns reflected in the MobilOD projection would not be significantly distorted by the process.

The chart for each hour of the resulting traffic assignment analyses from the matrix correction such as the one shown below were used for validation. The assignment analysis depicts link counts (including aggregate from turn counts). Only one chart is shown due to restrictions on maximum number of words.

![Assignment analysis](image)

**Figure 4: Assignment analysis for the hour of 7AM**

**Conclusion**

Based on the results obtained in the application study, it can be deduced that the methodology discussed in this paper holds promise for furthering the developments in the area of time varying demand estimation and will find more and more use as daily travel modeling becomes important due to peak spreading and other factors. The authors have started actively applying the methods to various practices where the demand variations are desired for investigating seasonal, weekday versus weekend and special event e.g. ball game day traffic.

**References:**

