An Optimization Framework for Travel Pattern Interpretation of Cellular Data

Sarit Freund and Hillel Bar-Gera
Department of Industrial Engineering Ben-Gurion University of the Negev, P.O.Box 653,
Beer-Sheva 84105,
(saritsh@bgu.ac.il), (bargera@bgu.ac.il)

[Received date; Accepted date] – to be inserted later

ABSTRACT

Collection of travel data by traditional survey methods is costly, thus limiting the amount of data being collected, as well as their frequency and coverage. Recent technologies offer new types of data collection options. In particular, cellular systems generate substantial amounts of data, including records regarding the connection between handsets (phones) and base stations (antenna). These records, collected by cellular service providers for various internal purposes, may provide an excellent source of information regarding travel, with several critical advantages relative to traditional travel surveys: low cost, large sample, long duration, and high response rate. Cellular data has some limitations too, particularly with respect to accuracy and traveler identity; therefore, such data cannot provide a complete replacement for traditional surveys, but it can complement and enhance them.

This paper explores methods for identifying travel patterns from cellular data. A primary challenge in this research is to provide an interpretation of the raw data that distinguishes between activity durations and travel durations. A novel framework is proposed for this purpose, based on a grading scheme for candidate interpretations of the raw data. A genetic algorithm is used to find interpretations with high grades, which are considered as the most reasonable ones. The proposed method is tested on a dataset of records covering 9454 cell-phone users over a period of one week. Preliminary evaluation of the resulting interpretations is presented.

1. INTRODUCTION

For many years the main method for collecting data on travel patterns has been household travel surveys. This method, as typically conducted, suffers from several limitations including: high cost, limited sample size, low frequency and coverage.

Another problem is the low and continuously decreasing response rate which may cause data bias [1, 2].

In recent years, researchers proposed various alternative methods to collect personal travel data, relying on technological advancement in the utilization of handheld computers, GPS devices and cellular phones [3, 4, 5, 6]. Wolf [7] examined the use of GPS data and found it was possible to identify trips which are often forgotten or dropped from travel diaries, like short trips or trips that occur as part of a trip chain.

Similar to GPS devices, cellular systems may provide objective data on locations and times. Cellular phones are very widely used in some countries, by the vast majority of the population; thus generating much larger quantities of data compared to GPS surveys, while avoiding the need to provide dedicated equipment.

Despite the mentioned advantages, using cellular data has some limitations too, related in part to the technology and in part to privacy concerns and regulations. Location information recorded by most cellular systems is relatively inaccurate. Recorded data in some cases indicates the cell, and in other cases only base-station location is recorded. Some cellular systems can actively determine handset position to better accuracy, but often there is an additional communication cost associated with these higher accuracy observations.

Privacy concerns differ between countries and cellular companies. The main case we consider is that access is enabled to anonymous handset location data. Since handsets are anonymous, there is no data on participants socio-demographic or other characteristics which may be of interest. It is also not possible to contact handset owners for complementary information, such as trip purpose, or to verify cellular data interpretation. Under more relaxed conditions, data on handset owners may be available; however, the handset user is not necessarily its owner. It should be recognized that there are cases with more restricting privacy conditions, where access to any cellular location data may be completely prohibited; clearly, in such cases the approach discussed here is not relevant.

In Estonia, Ahas *et al*. [8] used cellular data to map travel of suburban populations. The study focused on two main issues: what are the locations of services suburban population frequently visited downtown? And when do these visits take place (weekdays, weekends). Data was collected both passively and actively. Active location collection was performed by querying the cellular handset in real time and providing its geographical location. Generated offline, passive location collection occurred by examining a log report. To allow active collection, the users had to approve their participation, due to privacy concerns. The experiment included 277 people, actively sampled every 15 minutes over 8 days. The cellular data was compared with interview results. 61% of the cellular records in urban areas were within 1000 meters of the actual location, while in rural areas 53% of the records were within 3000 meters of the actual location.

The usefulness of cellular data in travel demand models was illustrated by Friedrich *et al*. [9], describing an analysis based on Floating Phone Data (FPD). The paper focuses on FPD usage in three applications: route choice analysis, generation of OD matrices and monitoring service quality in networks. FPD is divided to two types: 1)