Inferring individual daily activities from mobile phone traces: A Boston example

Mi Diao
National University of Singapore, Singapore

Yi Zhu, Joseph Ferreira Jr and Carlo Ratti
Massachusetts Institute of Technology, USA

Abstract
Understanding individual daily activity patterns is essential for travel demand management and urban planning. This research introduces a new method to infer individuals’ activities from their mobile phone traces. Using Metro Boston as an example, we develop an activity detection model with travel diary surveys to reveal the common laws governing individuals’ activity participation, and apply the modeling results to mobile phone traces to extract the embedded activity information. The proposed approach enables us to spatially and temporally quantify, visualize, and examine urban activity landscapes in a metropolitan area and provides real-time decision support for the city. This study also demonstrates the potential value of combining new “big data” such as mobile phone traces and traditional travel surveys to improve transportation planning and urban planning and management.

Keywords
Individual activities, mobile phone traces, travel survey

Introduction
Characterizing individual activity patterns is crucial for travel demand management and urban planning, as travel demand is derived from the activities that individuals need or wish to participate in. Since the 1980s, travel demand modeling has been evolving from trip-based approach to activity-based approach (Bhat and Koppelman, 1999; Bowman and Ben-Akiva, 2001). Travel diary surveys have been the major data sources for studying individual activity patterns. Such surveys provide detailed information on the activity participation of survey respondents. However, because of the high collection cost, data from surveys are typically limited in sample size and update frequency, which constrains their usage in the types of analyses that require longitudinal observations or large samples.
In the meantime, emerging urban sensing data accompanied by the development of more powerful and affordable computing provide us new opportunities to investigate, understand, and predict urban activities. The past two decades have seen the explosion of pervasive systems such as global positioning systems, mobile phone networks, and WIFI hotspots (Gonzalez et al., 2008; Wang et al., 2009). Such digital infrastructure of cities has become deeply embedded across all facets of our urban systems. In participating in our daily activities, we leave behind footprints from our interactions with the urban environment and its digital infrastructure (Girardin et al., 2008). Digital sensors, as a by-product of their normal operations, are collecting massive amounts of spatiotemporal data related to people and city. As such digital sensors become increasingly portable and ubiquitous, Batty et al. (2012) project that within the next 20 years, most data that we will use to understand cities will come from digital sensors of our transactions and will be available in various forms, with temporal tags as well as geotags in many instances.

Mobile phone traces collected by mobile network operators are a type of such data. It has attracted increasing interest in behavior research. Mobile phone trace data typically contain a series of consecutive spatial locations of phone users, for example, locations of cellular towers that a calling mobile phone is connected to (Gonzalez et al., 2008) or the estimated longitude/latitude locations of users using triangulation (Calabrese et al., 2013). Corresponding time stamps are recorded as well. The time and locations captured by cellular towers when an individual interacts with his mobile phone record a part of his daily space–time trajectory. Many researchers have used mobile phone traces to examine the spatiotemporal human dynamics (e.g. Ahas et al., 2010; Candia et al., 2008; Gonzalez et al., 2008; Ratti, 2005; Song et al., 2010) and shed new light on the mobility patterns of individuals. However, the analytical value and interpretability of the movement information from mobile phone traces are discounted, because there is no explicit activity information attached to mobile phone traces. Whether we can extract the activity information embedded in the spatiotemporal dynamics of mobile phone users has become a key for the application of mobile phone traces in understanding human behaviors and activities and supporting urban planning and management.

A limited number of more recent studies have focused on inferring activity patterns from mobile phone traces. These studies generally approach the problem from two perspectives. The first group of studies focuses on the relationships between land use and human activity and mobility patterns at aggregated levels (Noulas and Mascolo, 2013; Phithakkitnukoon et al., 2010; Soto and Frias-Martinez, 2011). For example, Soto and Frias-Martinez (2011) group locations into clusters based on the temporal distribution of calling activities in the area and investigate the correlations between human calling activity clusters and urban land use types. Phithakkitnukoon et al. (2010) develop an algorithm to identify the most probable activity associated with a specific location based on the spatial distribution of different types of points of interest. These studies are mostly conducted at spatially aggregated levels and the activity inference models are simply based on known land use characteristics of places. The second group of studies draws on locations revealed by mobile phone traces to infer corresponding individual’s activity type. Isaacman et al. (2011) use cellular network data to discern semantically meaningful locations such as home and work. However, the learning and validation of algorithms count on a very small group of volunteers. Jiang et al. (2013) propose to build probabilistic models to infer an individual’s activity type conditional on land use type, time, and daily mobility motifs. Using individual daily mobility motifs for activity inference could apply to frequent mobile phone users whose traces can be easily extracted from mobile phone data but may not be applicable to low-frequency users due to the sparseness of calling activities.
The previous studies mainly rely on mobile phone traces to study the daily activity patterns of mobile phone users and their relationships with urban space. They provide useful insights for urban management, but also suffer from various shortcomings of mobile phone data, such as the sparse temporal frequency and the lack of demographic information. The main purpose of this study is to translate the less than ideal but more readily available raw mobile phone traces to information that is more tangible and useful to planners, decision makers, and researchers. We introduce a new method that combines mobile phone traces and household travel surveys to infer activity information from mobile phone traces to support the longitudinal investigation of individual activities in dense urban areas. Our research can help identify hot spots in a city in terms of the activities being conducted; reveal how urban space is utilized at a particular point of time; and visualize how different neighborhoods evolve along the course of a day, week, or year. A better understanding of urban activity landscape (the spatial distribution of human activities in a city) not only has direct implication for transportation planning, but also for other fields such as urban design and infrastructure management. It can facilitate the development of novel location-based services and targeted policy interventions by providing much richer contexts (the types of interactions between people and urban built environment) to service providers and policy makers. This study also demonstrates the enormous potential of combining emerging “big” data such as mobile phone traces with traditional surveys to gain insights into travel behaviors and urban dynamics.

The rest of the paper is organized as follows: In “Data and methodology” section we describe an analytical framework that applies the associations and dependencies learned from travel surveys to extract the activity information of mobile phone users, using Metro Boston as an example. “Activity detection” section presents the activity detection model calibrated with a travel survey for Metro Boston. “Urban activity landscapes” section interprets the extracted activity information of mobile phone users and visualizes the evolution of urban activity landscape in Metro Boston over time. “Discussions and conclusions” section concludes our study and summarizes its significance.

Data and methodology

In the field of transport research, the mainstream data source used to understand human behavior is household travel surveys. Compared to urban sensing data, surveys can reveal individual activities in greater details with their more complete records of mobility options and choices, richer information on trip purposes (i.e. activities), and demographic information of individuals. But the size and frequency of surveys are often constrained by high collection costs, which prevent surveys from being used for the analyses that are able to gain more insights into the interactions between people, urban environment, and policy interventions, such as longitudinal studies, before and after studies, etc. On the other hand, new urban sensing datasets such as mobile phone traces allow us to study the movement of millions of individuals on a real-time basis, but mobile phone records are sparser and less accurate than surveys; demographic attributes of mobile phone users are not available due to privacy concerns; and the activity information embedded in individuals’ movements is not easy to extract. Therefore, combining urban sensing data and survey data represents a promising way to understand individual daily activities in a metropolitan area. In this study, using Metro Boston as an example, we exploit the richness of survey data to discover the spatial and temporal dependencies and other factors governing individuals’ activity participation in the spatiotemporal space and then apply these laws to large-scale and high-frequency mobile phone traces to extract the embedded activity information.
Data

The main datasets used in this study include the mobile phone trace data, travel survey data, and spatially detailed business location data for Metro Boston.

**Mobile phone trace data from AirSage.** The mobile phone trace dataset is comprised of anonymous location estimations collected by AirSage\(^1\) from about one million mobile phones in East Massachusetts over four months in 2009, which are generated each time a device connects to the cellular network, including calling, messaging, and web browsing. The location estimations not only consist of identifications of the mobile phone towers that the mobile phones are connected to, but also an estimation of the locations of mobile phone users generated through triangulation by means of AirSage’s Wireless Signal Extraction technology. The uncertainty range of mobile phone location data has a mean of 320 m and a median of 220 m. In this study, we employ the approaches presented in Calabrese et al. (2013) to estimate the positions of mobile phone users and identify their home locations at 500 m by 500 m grid cell level, as illustrated in Figure 1. More detailed descriptions on data preparation and processing can be found in Calabrese et al. (2013).

**Position estimate.** In the mobile phone trace data, each location measurement is characterized by a point expressed in latitude and longitude and a time stamp. The location measurements of a mobile phone number are threaded into a sequence according to their time series. To address the potential error in identifying a mobile phone user’s position due to localization errors or users making consecutive network connections in the same area, we identify consecutive points within certain time interval and spatial constraint and collapse these points to a hypothetical position—the centroid of these points (see Figure 1(a)). The trajectory of a mobile phone user over time can then be generated by connecting such hypothetical positions sequentially.

**Home location identification.** The home location estimates of mobile phone users come from Calabrese et al. (2013), which uses the same dataset for an individual mobility pattern study. To detect the home location of a mobile phone user, the study area is divided into 500 m by 500 m grid cells. For each cell we evaluate the number of nights the user connects to the network in the time interval 6 pm–8 am while in that cell in four months, and select as home location the cell with the greatest value (see Figure 1(b)).

Calabrese et al. (2013) assess the accuracy of the home location estimation by computing a measure of repetitiveness, i.e. dividing the number of nights in the home grid cell by the number of nights when the user connects to the network in four months. This measure ranges from 0 to 1. They find that more than 40% of the people have an estimated home location with repetition frequency greater than 0.5, which implies that the mobile phone users have been detected at the estimated home location at least half of the monitored days. They also compare the estimated density of mobile phone users to the population density estimated with the 2010 census data at the census tract level and find that the spatial distribution of mobile phone users’ estimated home locations matches quite well with the population distribution observed in census tracts. These evidences suggest that the home location identification approach has reasonable reliability.

**2010 Massachusetts Travel Survey (MTS) data from CTPS.** The MTS was a large-scale effort that collects information on residents’ daily travel choices, preferences, and behaviors to understand the travel patterns of residents in the Commonwealth of Massachusetts. The 2010 survey was conducted between June 2010 and November 2011 on 15,033 households and 37,023 individuals to identify where and how they traveled on a specific, designated
In order to ensure a sample that was representative of the Massachusetts population, each household was asked a series of detailed questions about their socioeconomic characteristics and transportation options.

InfoUSA business database. The destination choice of individuals’ activity participation is correlated with land use characteristics and the spatial distribution of opportunities. In this study, we use the InfoUSA business database to describe the spatial distribution of

![Diagram](image-url)

**Figure 1.** Illustration of the approaches used for activity position and home location identification for mobile phone trace datasets. (a) User position estimation, (b) home location identification.
activity centers characterized by the number of employees by sector. The InfoUSA database compiled by InfoUSA, Inc. provides employment size by detailed standard industrial classification (SIC), and xy locations of all business establishments in Metro Boston. We include eight major sectors based on SIC business divisions in our analysis, including service (SIC Division I: services), finance (SIC Division H: finance, insurance, and real estate), retail (SIC Division G: retail trade, excluding eating and drinking places), food (SIC Major Group 58: eating and drinking places), government (SIC Division J: public administration), construction (SIC Division C: construction), transportation (SIC Division E: transportation, communications, electric, gas, and sanitary services), and wholesale (SIC Division F: wholesale trade).³

Methodology
Individuals exhibit regular yet rich spatiotemporal dynamics in their social and physical lives. For example, the daily routine of an office worker may consist of working in her office during the morning working hours, having lunch around noon time in a restaurant, going back to office and working in the afternoon, and enjoying some relaxing time in a recreation center after dinner, etc. Meanwhile, individuals show significant heterogeneity in their daily activity patterns, and individuals with similar life style tend to live in neighborhoods of the same kind. Therefore, the activity type that a particular group of individuals engage in at a certain spatiotemporal point is identifiable to some extent based on a series of factors and variables such as location, time, and weather. The relationships between activity engagement and spatial and temporal factors could be revealed by a discrete choice model calibrated with household travel surveys. When calibrating this model, we only use information that is also extractable from mobile phone traces, despite that the survey contains much richer information than mobile phone traces. The reason is that when we infer activities for mobile phone users, we are bounded by the information provided by mobile phone traces. Due to the anonymity of mobile phone traces, it is unlikely to identify the socioeconomic and demographic background of individual mobile phone users. So we use the neighborhood aggregate socioeconomic and demographic attributes as proxies for individual characteristics.

Considering the location error of mobile phone traces (a mean of 320 m and median of 220 m), we select a statewide 500 m by 500 m grid cell layer developed by MassGIS, the state’s Office of Geographic and Environmental Information, as the spatial representation of Metro Boston for analysis. The locations in the MTS survey and mobile phone traces are assigned to corresponding grid cells. The time resolution is set to be 1 h. Within a multinomial logit model (MNL) framework (Ben-Akiva and Lerman, 1985), we estimate the probability that an individual \( n \) will participate in activity type \( i \) in hour \( t \) and grid cell \( g \), denoted by \( P_{ngt}(i) \), with time, location, and individual attributes using the MTS survey data. The model can be specified as

\[
P_{ngt}(i) = \frac{\exp(\alpha_i + \beta_i' H_{nt} + \gamma_i' L_{ng} + \lambda_i' W_{ngt} + \phi_i' P_{nn})}{\sum_{j \in C_n} \exp(\alpha_j + \beta_j' H_{nt} + \gamma_j' L_{ng} + \lambda_j' W_{ngj} + \phi_j' P_{nn})}
\]

where \( C_n \) is the set of activity types available to individual \( n \); \( H \) is a set of temporal characteristics; \( L \) is a set of locational characteristics; \( W \) is a set of other factors such as weather condition, \( P \) is a set of individual specific variables, which is proxied by home neighborhood characteristics; and \( \alpha, \beta, \gamma, \lambda, \) and \( \phi \) are coefficients to be estimated.
The same sets of variables are computed for mobile phone users. We can then apply the calibrated activity detection model to the mobile phone dataset to predict the activity type of mobile phone users at a particular point of time and space. Figure 2 illustrates the methodological framework of this study.

**Activity detection**

In this section, we calibrate an activity detection model with the MTS data to reveal the common laws governing individuals’ activity participation. We group the 25 activity types in the MTS into seven activity categories, as shown in Table 1. Our primary interest is the nonhome activities. Therefore, the objective of the activity detection model is to predict the activity category $i$ (one of the six nonhome categories) of an individual $n$ at time $t$ and grid cell $g$ outside his/her home grid cell.

Four groups of variables are included in the activity detection model:

1. **Locational attributes of the grid cell in which the activity occurs**

   The built-environment characteristics and availability of opportunities of places can influence individuals’ activity participation. In this study, we consider locational characteristics along four dimensions: density, land use profile, distances to CBD and home, and transit provision. In computing the density measure, the population count at the 500 m by 500 m grid cell level is provided by MassGIS using an algorithm that can allocate census population count to residential land uses and then to grid cells. We utilize the employment by sector data from the InfoUSA business database to characterize land use profile and opportunity availability at the grid cell level. The distances to CBD and home variables are both Euclidean distances computed with GIS tools. The transit provision variables measure the numbers of subway stations and bus stops in the grid cell, respectively. It should be noted that this grid-cell-based approach could lead to some potential biases. For example, the activity profile in one grid cell could be influenced by
the location attributes of neighboring grid cells. Capturing the spillover effects could improve the prediction power of our model.

(2) Temporal attribute

Individuals show regular temporal patterns in activity participation. To capture this effect, we include dummy variables for day of the week and time of the day in our model. In this study, we only consider weekdays due to the fact that the MTS survey only covers weekdays. The weekend pattern is typically different from weekday pattern because most working activities do not occur in the weekends. One day is divided into eight time segments to capture the intraday variations in activity participation: early morning (3–6 am), morning-peak hour (6–9 am), morning-work (9 am–12 pm), noon (12–2 pm), afternoon-work (2–5 pm), afternoon-peak hour (5–8 pm), night (8 pm–12 am), and midnight (12–3 am).

(3) Weather of the day

Weather conditions vary across time and locations and could influence individuals’ activity choices. Due to the lack of fine-grained weather information, we include two

---

Table 1. Mapping of activity types into activity categories.

| ID | Activity category     | Activity type                                                                 |
|----|----------------------|-------------------------------------------------------------------------------|
| 1  | Home                 | Working at home (for pay)                                                   |
|    |                      | All other home activities                                                    |
| 2  | Work                 | Work/Job                                                                      |
|    |                      | All other activities at work                                                 |
|    |                      | Volunteer work/Activities                                                    |
|    |                      | Attending class                                                              |
|    |                      | All other School Activities                                                  |
| 3  | Eat out              | Eat meal outside of home                                                     |
| 4  | Personal business    | Drop off passenger from car                                                  |
|    |                      | Pick up passenger from car                                                   |
|    |                      | Service private vehicle (gas, oil lube, etc.)                                |
|    |                      | Household errands (bank, dry cleaning, etc.)                                 |
|    |                      | Personal business (visit government office, attorney, accountant)            |
|    |                      | Health care (doctor, dentist)                                                |
| 5  | Social and recreational | Civic/religious activities                                                |
|    |                      | Outdoor recreation/entertainment                                             |
|    |                      | Indoor recreation/entertainment                                              |
|    |                      | Visit friends/relatives                                                      |
| 6  | Shopping             | Routine shopping                                                             |
|    |                      | Shopping for major purchases or specialty items                              |
| 7  | Other                | Changed type of transportation                                               |
|    |                      | While traveling                                                              |
|    |                      | Work business related                                                        |
|    |                      | Loop trip                                                                    |
|    |                      | Other                                                                         |
aggregate variables in the model to control for the weather effect: average precipitation and temperature in the metro area over the day.

(4) Socioeconomic attributes of the home grid cell

Individuals with heterogeneous socioeconomic characteristics tend to display different daily activity patterns. One shortcoming of mobile phone trace data is the absence of individual characteristics. In this study, we use aggregate socioeconomic characteristics at the home grid cell to control for this factor, including percent of owner-occupied housing, percent of population under 5, median household income, percent of white population, and distance of home to CBD. Each grid cell is assigned the characteristics of the block group that contains the grid cell’s centroid. This approach can improve the differentiation of mobile phone users but still has some potential for ecological fallacy issues.

The descriptive statistics of variables in the model are summarized in Table 2.

Table 3 presents the results of the activity detection model with the “working” activity as the base case. The MNL compares the odds that an individual chosen at random will engage in any of the five other activities (rather than be working) if that individual resides in a neighborhood characterized by the “home location” attributes and is observed at a time and place categorized by the “activity location,” temporal, and weather attributes listed in Table 3. In the model calibration, Friday and early morning (3–6 am) are used as the base cases for the day-of-the-week and time-of-the-day variables, respectively.

The estimation results suggest that spatial and temporal characteristics play an important role in activity participation. For example, in the shopping activity model, the coefficients of the retail employment and time-of-the-day dummy variables for afternoon and night hours are all positive and significant, while the day-of-the-week dummy variables (Monday to Thursday) have negative and significant coefficients. These results suggest that the probability that an individual conducts shopping activity is higher in a grid cell with high retail presence late Friday. In the meantime, the probability that an observed individual is
Table 3. Estimation results for predicting time/place activity likelihood (relative to “working” activity).

|                      | Eating out | Personal business | Social and recreational | Shopping | Other |
|----------------------|------------|-------------------|-------------------------|----------|-------|
|                      | Estimate   | t-value           | Estimate                | t-value  | Estimate |
| Intercept            | -3.606     | -10.570 **        | -2.243                  | -11.036 ** | -1.090 |
|                      | -8.501 **  | 2.667             | -8.589 **               | -1.277   | -8.580 **|
| **Activity location attributes** |           |                   |                         |          |        |
| Population density (k/sq km) | 0.031     | 4.299 **          | 0.062                   | 13.266 ** | 0.069 |
| Employment—administrative (k) | -0.354    | -3.239 **         | -0.192                  | -3.660 ** | -0.258 |
| Employment—retail (k) | 0.523      | 5.946 **          | 0.106                   | 1.328     | -0.523 |
| Employment—wholesale (k) | -1.540     | -4.864 **         | -1.781                  | -6.796 ** | -2.306 |
| Employment—eating (k) | 1.020      | 7.906 **          | 0.212                   | 1.859     | 1.124 |
| Distance to CBD (km)  | 0.018      | 6.849 **          | 0.011                   | 5.845 **  | 0.014 |
| Distance to home (km) | -0.024     | -1.1468 **        | -0.054                  | -34.269 **| -0.015 |
| Number of subway stations | 0.054   | 0.755 **          | -0.062                  | -1.098    | 0.052 |
| Number of bus stops  | 0.062      | 7.781 **          | 0.028                   | 5.558 **  | 0.015 |
| **Home location attributes** |           |                   |                         |          |        |
| Distance of home to CBD (km) | -0.008     | -3.265 **         | 0.000                   | -0.059    | -0.012 |
| Percent of owner-occupied housing | -0.006   | -8.444 **         | 0.001                   | 1.561     | 0.006 |
| Percent of population under 5 | -0.009     | 2.016 *          | -0.005                  | -1.654    | 0.001 |
| Median household income (k$) | -0.003    | -2.718 **         | 0.004                   | 7.777 **  | 0.001 |
| Percent of white population | 0.008  | 7.790 **          | 0.000                   | -0.019    | 0.002 |
| **Weather** |           |                   |                         |          |        |
| Precipitation        | 0.045      | 6.853 **          | -0.108                  | -2.631 ** | -0.179 |
| Average temperature  | 0.008      | 6.853 **          | 0.003                   | 4.195 **  | 0.012 |

(continued)
Table 3. Continued.

|                      | Eating out | Personal business | Social and recreational | Shopping | Other |
|----------------------|------------|-------------------|-------------------------|----------|-------|
|                      | Estimate   | t-value           | Estimate | t-value | Estimate | t-value | Estimate | t-value | Estimate | t-value |
| **Day of week**      |            |                   |            |         |          |         |          |         |          |         |
| Monday               | -0.607     | **-9.643**        | -0.124    | **-3.234** | -0.401   | **-10.648** | -0.308   | **-5.993** | 0.027    | 0.659   |
| Tuesday              | -0.587     | **-9.609**        | -0.089    | *-2.336*  | -0.350   | **-9.519**   | -0.374   | **-7.211** | -0.011   | -0.264 |
| Wednesday            | -0.498     | **-8.169**        | -0.018    | -0.476   | -0.286   | **-7.742**   | -0.211   | **-4.173** | 0.115    | 2.866   |
| Thursday             | -0.411     | **-6.894**        | 0.032     | 0.849    | -0.219   | **-5.948**   | -0.277   | **-5.323** | 0.142    | 3.507   |
| **Time of day**      |            |                   |            |         |          |         |          |         |          |         |
| Morning-peak hour (6–9 am) | -0.086 | -0.270            | 0.670     | 3.584   | **-1.890** | **-18.297** | **-0.414** | **-1.398** | -0.454   | -3.630   |
| Morning-work (9 am–12 pm) | -0.486 | -1.544            | 0.363     | 1.949   | **-1.826** | **-18.413** | 0.532    | 1.840    | -1.043   | -8.391   |
| Noon (12–2 pm)       | 0.475      | 1.520             | 0.252     | 1.348   | **-1.757** | **-17.440** | 0.777    | 2.685    | **-1.062** | **-8.436** |
| Afternoon-work (2–5 pm) | -0.038 | -0.123            | 0.736     | 3.953   | **-1.363** | **-13.814** | 0.991    | 3.433    | **-0.763** | **-6.142** |
| Afternoon-peak hour  | 2.026      | **6.501**         | **1.378** | **7.349** | **0.259** | **2.617** | **1.854** | **6.401** | 0.298    | 2.371   |
| (5–8 pm)             |            |                   |            |         |          |         |          |         |          |         |
| Night (8 pm–12 am)   | 2.782      | **8.837**         | **1.415** | **7.252** | **1.164** | **11.025** | **1.810** | **6.099** | 1.008    | 7.622   |
| Midnight (12–3 am)   | -0.192     | -0.369            | 0.023     | 0.081   | 0.693    | **4.919** | **-2.262** | **-2.166** | 1.755    | **11.089** |

**Significant at the 0.01 level.
*Significant at the 0.05 level.
eating out increases in areas with many restaurants as suggested by the positive and significant coefficient of the eating-related employment variable in the eating-out activity model. The favorable time for eating out is when individuals finish their afternoon work, because both the afternoon peak hour and night dummy variables have positive and significant coefficients. The likelihood of eating out increases from Monday through Friday because the four day-of-the-week variables (Monday to Thursday) have negative and significant coefficients, and their magnitudes decrease from Monday to Thursday (−0.607, −0.587, −0.498, and −0.411, respectively). The individual characteristics as represented by neighborhood socioeconomic attributes influence activity participation as well. Individuals living in wealthier neighborhoods are more likely to participate in personal business and social and recreational activities all else being equal, as suggested by the positive and significant coefficients of the median household income variable in corresponding activity models, while less likely to eat out, as suggested by the negative and significant coefficient in the eating-out activity model. The model also suggests that an individual’s choice of activity participation is influenced by weather conditions. A rainy day will reduce the possibilities of personal business and social and recreational activities. By contrast, precipitation appears to be not significantly correlated with the possibilities of eating-out and shopping. Meanwhile, personal business and social and recreational activities increase with the average temperature of the day.

In term of goodness of fit, the likelihood ratio test of the activity detection model is highly significant (p value = 0.000), which rejects the null hypothesis that all the parameters other than the alternative-specific constant are zero and suggest that individuals’ activity participation can be predicated with spatial, temporal, and socioeconomic information. The McFadden R squared value is 0.103. It should be noted that McFadden R squared is different from the R squared in ordinary regression analysis and should not be judged by the standard for a “good fit” in ordinary regression analysis. Normally, a value of 0.2–0.4 for McFadden R squared represents an excellent fit (McFadden, 1978).

The estimated coefficients of variables cannot be directly compared due to their different units. To further understand the magnitude of various factors’ impacts on activity choice, we compute the aggregate elasticity of the probability of engaging an activity with respect to changes in these factors, following Ben-Akiva and Lerman (1985). In the context of this study, the aggregate elasticity measures the effect of an incremental change in a continuous variable on the expected share of individuals choosing activity type $i$, which is computed as a weighted average of individual level elasticities using the choice probabilities as weight (Ben-Akiva and Lerman, 1985). The aggregate elasticities of continuous variables in the model are presented in Table 4 (variables with insignificant coefficients are excluded). We find that the socioeconomic characteristics of the home grid cell play important roles in individuals’ activity participation. For example, 1% increase in median household income in the home grid cell is associated with 0.219% decrease in the share of individuals participating in eating-out activities and 0.191% increase in personal business activities, all else being equal. The elasticity values also indicate that employment by sector has higher impact on the share of related activity type compared to shares of other activity types. For example, 1% increase of retail facilities (as proxied by retail employment) in a grid cell will lead to 0.141% increase in the share of individuals who visit the location to participate in shopping activities, while its impact on the shares of other activity types are significantly lower (below 0.05). In terms of the weather effect, average temperature has a substantial impact on individual activity participation, higher than most location characteristics, while the impact of precipitation is minimal. One percent increase in average temperature will lead
to 0.222% increase in the share of eating-out activities, 0.325% increase in the share of social and recreational activities, and 0.194% increase in share of shopping activities.

**Urban activity landscapes**

With the activity detection model calibrated in the previous section, we apply the modeling results to the observed mobile phone traces to extract the hidden activity information and visualize the evolution of urban activity landscape over time. Among the one million mobile phone users in East Massachusetts in the AirSage data, we select individuals for whom we have estimated their positions in 60% of the hours during the four months in 2009. Using this filter, we obtain 20,858 frequent mobile phone users and analyze their mobile phone traces during five weekdays (14–18 September 2009) as a demonstration. It should be noted that frequent mobile phone users are not a random sample of the population as we might expect, but the method that we propose can be easily applied to all mobile phone users in the metro area. According to Calabrese et al. (2013), mobile phone users are good representatives of the population in Metro Boston. Meanwhile, we can expect the representativeness of mobile phone users could be further improved in the future as the penetration rate of smartphones keeps increasing. By combining the calibrated model parameters using the MTS survey and variables generated from the mobile phone traces, the activity detection model can help extract the embedded activity information from the mobile phone traces, thus providing stronger support to more responsive urban planning and management than the raw mobile phone trace data. For example, Figure 3(a) shows the aggregate number of frequent mobile phone users by time during the five days, which can
be easily computed with the raw mobile phone trace data. Figure 3(b) plots the predicted activity participation of the same group of users for the same time period. By decomposing the mobile phone users into subgroups classified by different types of activity endearment, the evolutions of different activities over time in the city are illustrated vividly.

The mobile phone traces combined with the activity detection model not only allow us to illustrate the temporal dynamics in the metro area, but also to visualize the evolution of the spatial distributions of activities within a city—the urban activity landscape—over time. In this demonstration, we only analyze the activities of 20,858 frequent mobile phone users during the study time period. To overcome the sparseness of data, we combine the five days into a hypothetical day to display the daily activity patterns. Figure 5(a) illustrates the spatial distribution of frequent mobile phone users at different points of time during the hypothetical day. These figures are readily producible with the raw mobile phone trace data and provide useful information in managing our cities. Then using working and social and recreational activities as examples, Figure 5(b) and (c) plots the spatial distributions of activities by the frequent mobile phone users in Metro Boston by time in the hypothetical day based on the model prediction. We can observe that after midnight the city falls asleep. There are very few individuals working during this time period but some social and recreational activities still going on in the urban center. As time passes, the city gradually

Figure 3. Predicted activity participation by hour (14–18 September 2009). (a) Counts of frequent mobile phone users in all activities, (b) predicted count of frequent mobile phone users in working activities, (c) predicted counts of frequent mobile phone users in social and recreational activities.
wakes up as increasing numbers of individuals start working, especially in the urban core and regional subcenters such as Lawrence and Lowell. At 10 am, the working activities reach a stable state, with hot spots clustered around the CBD area. This spatial pattern continues until 4 pm. From 7 to 10 pm, as more and more workers switch from working to social and recreational activities, the urban center still plays a pivotal role in the activity space. The evolution of urban

**Figure 4.** The evolution of urban activity landscape over time (color version). (a) Counts of frequent mobile phone users in all activities, (b) predicted count of frequent mobile phone users in working activities, (c) predicted counts of frequent mobile phone users in social and recreational activities.

Diao et al. 15
activity landscape vividly illustrates the spatial structure of Metro Boston—a monocentric city with several subcenters. Although these figures only showcase several snapshots in a hypothetical day, the methodology presented in this paper has the potential to provide real-time support for urban management with the support of powerful computing.

Beyond revealing the urban activity landscape in the metro area, the predicted daily activity patterns of individuals can further help depict the activity profiles of individual places. Characterizing place is not only crucial for urban design but also important for location choice models in transportation and urban modeling. Traditionally, researchers rely on static land use data to differentiate locations. With the mobile phone trace data and the activity detection model, we can characterize how urban space is utilized by individuals during the day. Figure 5 shows the amount and type of activities by the frequent mobile phone users at some typical urban nodes in Metro Boston during the hypothetical day with activity polar graphs, juxtaposed against employment statistics by sector at the location aggregated from the InfoUSA data. To illustrate how activity patterns are correlated with land uses, different types of places are selected, including Government Center, Downtown Crossing, Financial District, and Chinatown/Boynton Street.

In Figure 5, the activity profiles of four locations in downtown Boston are represented by shaded circles. The radius of each circle represents the maximum number of frequent mobile phone users who are observed during any one hour within any 500 m by 500 m grid cell in Metro Boston. The circles are divided radially into 24 segments, each representing an hour. Mobile phone users are differentiated by colors according to their predicted activity types for the time in which they are observed within each grid cell. Using the InfoUSA data, we compute the number of employees by sector at the grid cell level in Metro Boston. The maximum numbers of employees by sector that are observed in any grid cell are represented by the length of a gray column below each circle. The yellow columns represent the actual employment by sector within the charted grid cell. By comparing the lengths of the gray and
yellow columns, we can identify the relative importance throughout Boston of each grid cell’s contribution to the corresponding employment sector. Note that each of the four locations of interest in Figure 5 has Metro Boston’s highest observed grid cell employment for one of its eight employment sectors.
As job centers with one dominating sector and limited supporting facilities, Government Center (a cluster of administrative jobs) and Financial District (a cluster of financial jobs) share similar activity profiles characterized only by working individuals and dramatic temporal variations in the usage of urban space during and outside daytime working hours. In contrast, Downtown Crossing is a mixed-use, high-density downtown shopping area, which supports a much richer activity profile than single-use job centers. Compared with the other three locations, the Chinatown/Boylston Street area displays less intensive yet more balanced space usage across the day and a higher proportion of eating and recreational activities. These figures illustrate how the interpreted mobile phone data can be used to characterize locations in a manner that can facilitate understanding, aid in location choice modeling, and improve urban management.

Discussions and conclusions

The increasingly portable and ubiquitous urban sensors, accompanied by the fast advancing ICT technology provide great opportunities for urban planning and management by enabling better understanding of the patterns and mechanisms of human mobility, activities, and their relationship with the urban environment. In the meantime, methods to fully exploit the depth, breadth, and richness of “big data” generated by urban sensors such as mobile phone traces are still limited, which constraints the usefulness of big data in real-world planning practices.

In this paper, we propose a new method that can infer the embedded individual activity information from mobile phone traces, combining the advantages of both emerging big data and traditional travel surveys while addressing some of the limitations of each dataset in the meantime. We first calibrate a MNL using travel surveys to reveal the common laws governing individuals’ activity participation. This model incorporates factors that are commonly provided by mobile phone traces such as spatiotemporal characteristics, home location, and weather conditions to make the laws transferrable to mobile phone traces. Then, activities behind the mobile phone traces that the mobile phone users are engaged in are inferred. This approach helps decompose the aggregate location estimates of mobile phone users into sublayers based on their activity types, thus providing richer context for urban management. The application of the proposed analytic framework to Metro Boston produces reasonable results and patterns, which demonstrates the practicality of this approach. The urban activity landscape maps generated for Metro Boston provide a vivid illustration of the spatiotemporal dynamics in the city. The activity profile of individual locations allows a more detailed characterization of urban space compared to traditional land use-based approaches.

Because the method can be applied to anonymized mobile phone trace datasets generated from administrative mobile phone records, changes in the spatiotemporal activity patterns within a metropolitan area can be detected quickly without waiting years for the next travel survey. The seasonal effects of human activities can also be identified as we observe the longitudinal behavior of mobile phone users. With the support of powerful computing, the approach proposed in this study has the potential to provide near real-time support of urban management and could have direct applications in multiple fields. For example, by knowing “how many people are doing what and where” on a quasi-real-time basis, we can achieve more efficient use of city infrastructure and more responsive urban planning and management. In the meantime, understanding how different locations are utilized over time can contribute to better design of urban space.
This study does have several noteworthy limitations, including: (1) the potential nonrandomness of mobile phone users, yielding some mismatch between mobile phone users and the general population; (2) the limited explanatory power of the activity detection model due to the complex nature of individual activity participation choice; (3) the potential activity bias due to the fact that mobile phone trace data are event driven (i.e. location measurements are available only when the device makes network connections, and such events could be correlated with activity participation); and (4) potential measurement errors because the socioeconomic backgrounds of users are estimated from the neighborhood attributes of estimated home locations, which could introduce the possibility of ecological fallacy biases.

The framework proposed in this paper to infer an individual’s activities from mobile phone traces enables the spatial and temporal quantification, visualization, and examination of urban activity landscape in a metropolitan area in a manner that is repeatable at relatively low cost and can provide timely decision support for urban management efforts. The method also demonstrates the enormous potential of combining large-scale urban sensing data with survey data to improve urban planning and management. Instead of being substituted by emerging big data, traditional survey data are actually helpful for making better sense of big data such as mobile phone traces. This study suggests the need for a new model of hybrid data collection and analysis to better understand city dynamics. In the meantime, new technologies enable us to collect, combine, and query huge amount of data from different sources about individuals and the world. Just as big data lays out many promises, it brings many challenges when it comes to privacy. We must carefully assess the privacy risk while exploring the enormous opportunities offered by new data.

**Conflict of interest**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**Notes**

1. http://www.airsage.com/.
2. http://www.massdot.state.ma.us/planning/Main/MapsDataandReports/Reports/TravelSurvey.aspx.
3. https://www.osha.gov/pls/imis/sic_manual.html.
4. The numbers for the time period between 11 pm and midnight are interpolated by the mean value of the hours before (10–11 pm) and after (0–1 am) due to errors in the original data. The same procedure is also applied in generating Figure 6.
5. For example, SkyHook, Inc. has used its location-based service traces to generate an anonymized “SpotRank” index of people presence in 100 m by 100 m grid cells for major metropolitan areas that could be routinely used for the purposes suggested in this paper.
References

Ahas R, Aasa A, Silm S, et al. (2010) Daily rhythms of suburban commuters’ movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies* 18(1): 45–54.

Batty M, Axhausen KW, Giannotti F, et al. (2012) Smart cities of the future. *European Physical Journal Special Topics* 214: 481–518.

Ben-Akiva ME and Lerman S (1985) *Discrete Choice Analysis: Theory and Applications to Travel Demand*. Cambridge, MA: The MIT Press.

Bhat CR and Koppelman FS (1999) Activity-based modeling for travel demand. In: Hall RW (ed.) *Handbook of Transportation Science*.ème Kluwer Academic Publisher.

Bowman JL and Ben-Akiva ME (2001) Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A* 35: 1–28.

Calabrese F, Diao M, Di Lorenzo G, et al. (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies* 26: 301–313.

Candia J, Gonzalez MC, Wang P, et al. (2008) Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* 41(22): 1–16.

Girardin F, Blat J, Nova N, et al. (2008) Leveraging urban digital footprints with social navigation and seamless design. Working paper. SENSEable City Lab, MIT. Available at: http://senseable.mit.edu/papers/pdf/2008_girardin_blat_dalfiore_ratti_CHI_umr_workshop.pdf.

Gonzalez MC, Hidalgo CA and Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196): 779–782.

Isaacman S, Becker R, Cáceres R, et al. (2011) Identifying important places in people’s lives from cellular network data. In: *Pervasive Computing*. Berlin Heidelberg: Springer, pp.133–151.

Jiang S, Fiore GA, Yang Y, et al. (2013) A review of urban computing for mobile phone traces: Current methods, challenges and opportunities. In: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, Chicago, IL, USA.

McFadden D (1978) Quantitative methods for analyzing travel behaviour on individuals: Some recent developments. In: Hensher D and Stopher P (eds) *Behavioural Travel Modelling*. Croom Helm, pp. 279–318.

Noulas A and Mascolo C (2013) Exploiting foursquare and cellular data to infer user activity in urban environments. In: *2013 IEEE 14th International Conference of Mobile Data Management (MDM)*, Vol. 1, pp.167–176.

Phithakkitnukoon S, Horanont T, Di Lorenzo G, et al. (2010) Activity-Aware Map: Identifying human daily activity pattern using mobile phone data. *Human Behavior Understanding* 6219(3): 14–25.

Ratti C (2005) Mobile landscape – graz in real time. In: *Proceedings of 3rd symposium on LBS and TeleCartography*. Vienna University of Technology. pp.28–30.

Song C, Qu Z, Blumm N, et al. (2010) Limits of predictability in human mobility. *Science* 327(5968): 1018–1021.

Soto V and Frias-Martinez E (2011) Robust land use characterization of urban landscapes using cell phone data. In: *Proceedings of the 1st Workshop on Pervasive Urban Applications, in conjunction with 9th Int. Conf. Pervasive Computing*.

Wang P, Gonzalez M, Hidalgo C, et al. (2009) Understanding the spreading patterns of mobile phone viruses. *Science* 324(5930): 1071–1076.

Mi Diao

Yi Zhu

Joseph Ferreira Jr

Carlo Ratti