Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 47 (2020) 283–290

Transportation Research Procedia

www.elsevier.com/locate/procedia

22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18-20 September 2019, Barcelona, Spain

Inferring Urban Mobility and Habits from User Location History

Guido Cantelmo[a], Piergiorgio Vitello[b],Bogdan Toader[b],

Antoniou[a], Francesco Viti[b]

[a]Techincal University of Munich, Arcisstrasse 21, 80333 München, Germany
[b]University of Luxembourg, Avenue de l'Université 2, 4365 Esch-sur-Alzette, Luxembourg

Abstract

Retrieving exhaustive information about individual mobility patterns is an essential step in order to implement effective mobility solutions. Despite their popularity, digital travel surveys still require a significant amount of inputs from the respondent. Consequently, they require great efforts from both respondents and analysts, and are limited to a relatively short period of time – between a few weeks and a year. Driven by these motivations, the approach proposed in this paper uses mobile phone location history to automatically detect activity location without any interaction with the respondent. The proposed methodology uses raw location data together with a special indexing technique to calculate the probability of performing a certain activity in a certain location. It uses a heuristic rule to improve this estimation by considering the value of information over time. Finally, GIS data about the number of facilities located in a certain area is downloaded in real-time to further improve the overall estimation. Results of this exploratory study support the idea that the proposed approach can reconstruct complex mobility patterns while minimizing the number of active inputs from the respondent.

Keywords: Passive data collection, travel survey, machine learning, activities, GIS data;

1. Introduction

Traditional travel data collection methods require significant contributions from survey respondents, which usually have to manually record all trip sequences and activities information. As a result, travel surveys usually run for a limited amount of time (no more than a few weeks) and, due to the self-reporting process, include several under-reported or incomplete information (such as wrong departure times or missing activities) (Zhou and Golledge 2007). Such restrictions make traditional surveying methods inappropriate to capture variations in travel behaviour that do not occur within a short time period (Abdulazim et al. 2013).

A number of studies show how digital travel surveys or travel diaries (via e.g. mobile phones) reduce survey burdens while increasing data accuracy. In other words, we can collect data that are more accurate for a longer time (Abdulazim

10.1016/j.trpro.2020.03.100

et al. 2013; Cottrill et al. 2013). Moreover, after an initial phase of training, the system learns user preferences further reducing respondent's efforts (Cottrill et al. 2013). Hence, such survey methods can overcome the main problems of manually collecting travel diaries, mainly their high costs, their accuracy and the limited collection times.

Although device-based logging has been widely adopted worldwide, it still relies on some nomadic devices that users have to carry all the time (Zhao et al. 2015). Moreover, smartphones are powered from batteries with a limited capacity, meaning that power consumption is a limitation. Even though many applications claim to have a low power consumption, battery duration, reliability, and lifetime depends on the manufacturers. Moreover, the higher the number of sensors, the higher the power consumption.

In order to avoid these limitations, this study aims at using a mix of internet data and, specifically, web-based applications (such as Google Map or OpenStreetMap), in order to reconstruct user behaviour without the need of running a travel survey. Large corporations, such as Google, already collect information at a user level with a high level of precision. In Europe, by law (Art 15 - Right of Access by the Data Subject. General Data Protection Regulation, GDPR), users have always access to their own information, meaning that every single individual has always access to his own data. Hence, instead of relying on wearable devices or external applications, we propose an automatic classification technique that exploits this existing information to infer activity locations.

In earlier works (Toader et al. 2018), the authors already proposed a data analytic tool to extract activity location using only spatial location data, indexing techniques, and a set of aggregate statistics about activity scheduling and preferences. This paper extends this approach by introducing temporal and geographical information into the system. First, a new heuristic rule is introduced to explicitly account for the fact that information is time dependent. Then, Geographic Information System (GIS) data (such as the location of services) are also considered into the model to better estimate secondary activities (such as eat, sport or shopping). Instead of using only a historic database from past locations, we directly estimate online (and in real-time) service locations through the web-service OpenStreetMap (OSM). As these data are freely available worldwide, under the condition of having the location history, this approach becomes applicable everywhere.

2. Literature Review

Concerning the data collection, a number of studies exploited GPS-based surveys to collect digital information in an automated or semi-automated way (Bricka et al. 2012). Their usage has many advantages such as reduction in respondents' burden, higher data accuracy, detailed trip route and the ability to extract additional information such as vehicle speeds (Chen et al. 2016). Often, this type of survey has been considered as a complementary solution to household travel surveys (Bricka et al. 2012). Although these methods have clear advantages, equipping the respondents with these devices is not only expensive but also generates additional problems, such as insufficient signal (e.g. within a building) or forgetting to carry the device (Zhao et al. 2015).

Data fusion of GPS traces and GIS data have also been proposed by different authors (Bohte and Maat 2009). The main idea is that GIS information can be combined with some heuristic rule about activity scheduling and duration in order to infer activity location (in the case of services), and mode of transport (in the case of transport facilities). However as mentioned in the introduction, this GPS logging has the main limitation of carrying an extra device. Also, they require prior information about home and – usually – work location (Bohte and Maat 2009). Additionally, land-use (location of residences, workplaces and other activities) changes continuously over time, meaning that the GIS database needs also to be constantly updated. This is the main limitation when the main interest is to collect activity information over a long period. To avoid this limitation, other authors proposed smartphone-based applications that do not require an additional device (Bohte and Maat 2009).

While almost any method successfully identifies home and work location, recent studies show that last generation surveying methods show better accuracy and higher resolution in representing leisure activities (Nahmias-Biran et al. 2018). While different systems have been proposed, such as rMove (Greene et al. 2016) and FMS (Zhao et al. 2015), all of them require a certain degree of user reporting. Other works focus on inferring travel information. These can be broadly classified into two main groups: heuristic and learning-based approaches (Abdulazim et al. 2013). The main difference is that heuristic models rely on simple rules related to recurrent user behaviour (such as activity scheduling or duration) to learn trip characteristics (Tsui and Shalaby 2006). These group of models can be considered model-driven, as it combines traces with land-use (or GIS) data in order to exploit some existing knowledge about the transportation system (Tsui and Shalaby 2006).

The main limitation is that these approaches are usually not general as they depend on a specific region or transport system (Abdulazim et al. 2013). For this reason, learning-based models have also been developed to derive these rules from some data through machine learning or data mining techniques (Byon, Abdulhai, and Shalaby 2009). However, in this case, the main problem is that different machine learning techniques will classify data in a different way causing different errors, which can be difficult to identify and fix.

3. Methodology

The proposed methodology consists of three interconnected components: (a) Clustering and Pattern Extraction (b) Heuristic process to identify home/work locations, and (c) a Bayesian updating process to profile activities performed in all remaining locations.

The first phase uses unsupervised machine learning to cluster frequently visited locations and provide estimations of activities performed in that area. Then, the Heuristic rule and the Bayesian process leverage temporal and geographical information to improve estimation results. In this section, we will discuss each of these sub-components.

3.1. Clustering and Pattern Extraction

As the *Clustering and Pattern Extraction* phase is not an original contribution of this paper, in this section, we will simply describe its main characteristics. We refer the interested reader to (Toader et al. 2018; Toader et al. 2017) for more details.

Given the location history for each user, the weekly activity pattern for each location visited is generated by clustering all the visit records from the beginning of the data collection period. However, we are aware that only certain activities can be identified without any validation from the respondent. Thus, we need to identify which activities can be directly estimated by the data. To do so, we adopted a hierarchical clustering approach. Starting from an available travel survey, activities have been aggregated based on the number of visits within a certain reference period. Then, the hierarchical clustering showed the similarity between different groups of activities. This approach has been adopted on two different databases, a travel survey collected in the region of Ghent (Belgium), and a second database at the University of Luxembourg in 2015. In both cases, results show that data can identify 4-5 groups of activities that strongly differ among each other (Toader et al. 2018). For more than 4-5 groups, additional feedback from the user was required.

Selected activities are then grouped generating tables of observations. For each group of activities, data are classified based on day and time of the week, creating a table that shows the permanence of each user in each location, for each day of the week (Fig. 1). Specifically, each cell of this matrix represents the number of times that a certain user U has been observed in a certain location L, given a certain time and day of the week.

This matrix can be obtained directly from the travel survey (Fig. 1a). In this case, the matrix is the aggregation of all users and locations within the database – thus representing their average behaviour. Then, this matrix is calculated for each user and each location. The weekly activity pattern for each location visited is generated by clustering all the visit records from the beginning of the data collection period. The clustering technique is based on indexing techniques using graphs (such as ND-tree). For more details on the clustering procedure, we refer the interested reader to (Toader et al. 2017). Figure 1 shows the table obtained from the travel surveys (1a) and the one obtained for a generic user (1b). In Fig. 1b, the activity is unknown. However, we verified with the user the real activity performed at the location. Figure 1 shows two points. First, the aggregated activities look extremely different, meaning that a smart algorithm could estimate each of them. Second, while home and work can be easily estimated, this is not evident for the activity "*Sport*". This suggests that we need to include additional information and, specifically, we need to understand if there are sport facilities in that area or not. This is the mission of the Bayesian Updating rule.

3.2. Home-Work location: Heuristic rule

Many works show that combining GIS and GPS data can provide a good estimation of the activity performed at a certain location (Li and Shalaby 2008). However, in these works, "*Home*" and "*Work*" locations are known. By removing this information from the system, the model is aware that only the remaining locations need to be classified. As the proposed model does not require any information from the user, we need to identify "*Home*" and "*Work*" location before estimating the remaining activity locations.

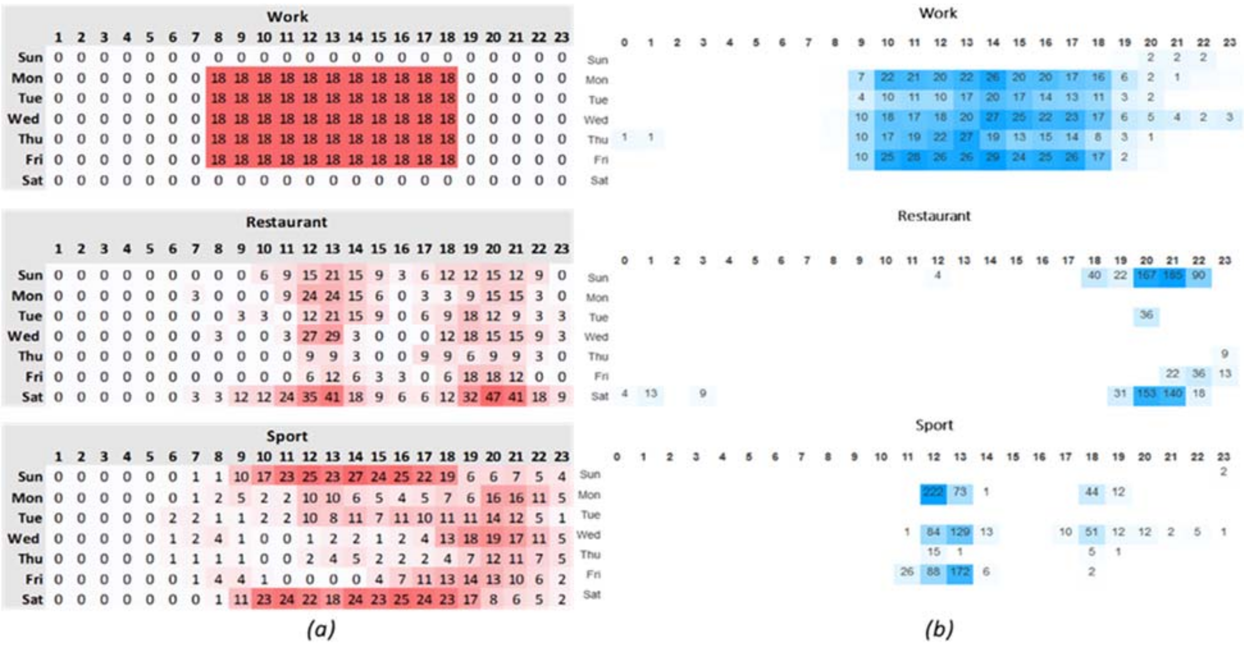

Fig. 1: (a) Matrix of aggregate activities from the data; (b) Matrices for a user;

By looking at Fig. 1, it seems reasonable to assume that the *Clustering and Pattern Extraction* phase should be able to identify these locations. However, this phase just provides a first and coarse estimation of the probability to perform an activity in a certain location. Specifically, the *Clustering* calculates the similarity between data and observations. However, human behaviour is not explicitly considered. For instance, it does not take into account that some time intervals carry more information (for instance, it is very likely that users will be home between 3-4AM on a working day). While Digital Travel Surveys have a validation phase, we need to introduce some behavioural model into the process to consider those phenomena that are not modelled within the clustering. Hence, we introduce a weight that modifies the clustering result by stressing that information is time dependent. These weights derive from *Utility Maximization Theory*. Following the theory that utility changes over time and that this phenomenon can be represented through a continuous function (Ettema et al. 2007), the weight will maximize the information related to those time intervals that carry more information (i.e. local minimum/optimum of the utility function).

In general, the proposed heuristic rule builds upon utility theory. By defining W_a^t the weight of activity *a* at time t, and assuming that each activity has a utility funcition $U(t)$, as discussed in (Ettema et al. 2007), the heuristic rule proposed in this paper can be written as:

1) If $U(t)$ is close to a local minimum or, then $W_a^t > 1$
2) If $U(t)$ is not a local minimum, then $W_a^t = 1$
3) If $U(t)$ is next to the beginning/end of the activity, then $W_a^t \leq 1$

Rule (1) is a *must-be* (since $U(t)$ is a local minimum the information is maximum), rule (3) takes into account the flexibility of the demand (it is more likely to have larger errors next to the beginning/end of the activity), and rule (2) takes into account all other scenarios.

3.3. Bayesian updating rule

Given the heuristic rule discussed in section 3.2, this section introduces a probabilistic approach based on GIS data to identify secondary activities. Specifically, we exploited the read-only "Overpass" API to retrieve online data through OpenStreetMap*. In this paper, we focused on three secondary activities: shopping, sport and food as, according to the hierarchical clustering, data about these activities show significant differences. We adopted the following process to extract how many services are located in a certain location:

1) Provide a location (x,y), where x and y are the coordinates of the point.
2) Identify edges of the surrounding area [(x+r,y+r),(x-r,y-r)], where *r* is the radius of the area we want to consider.
3) For each element *e* within the area [(x+r,y+r),(x-r,y-r)]:
    a. Draw *activity_type* from tag {*amenity, shop, leisure, sport*}
    b. Assign *activity_type* to one of the categories {*shopping, sport, food*}
4) Count the number of locations n_a for each *activity_type a*

For more details about Overpass and OpenStreetMap, we refer to their official documentation. We then use a Bayesian updating rule to combine this information with the probability previously calculated. Specifically, the indexing techniques discussed in sub-sections (3.2) provides the posterior probability $P^a(U|L)$ of user *U* doing an activity *a* given a location *L*, meaning that we can write the Bayesian updating rule as:

$$P^a(L\,|\,U) = \frac{P^a(U\,|\,L)P^a(L)}{P^a(U)} \tag{1}$$

- $P^a(U)$ the probability of user *U* doing activity *a;*
- $P^a(L)$ the probability of performing activity *a* in location *L;*
- $P^a(L|U)$ the posterior probability of performing activity *a* in location *L,* given user *U* and his location history.

In order to calculate $P^a(L|U)$, we define $P^a(U)$ and $P^a(L)$ as:

$$P^a(U) = \left(Number\_Activities\right)^{-1} \tag{2}$$

$$P^a(L) = \frac{e^{\frac{n_a+\varepsilon}{\theta}}}{\sum_a e^{\frac{n_a+\varepsilon}{\theta}}} \tag{3}$$

Equation (2) implies that the prior probability $P^a(U)$ follows a uniform distribution (in this case we consider 5 activities, thus $P^a(U) = 0.2$). To calculate $P^a(L)$, we leverage the GIS data and the number of activity locations n_a. In essence, if there are two restaurants and no sport centers, the probability for activity "food" will be higher. The error term ε in equation (3) takes into account that, even though OpenStreetMap has a rich database, our information could be incomplete (i.e. unreported locations). Moreover, "Home" and "Work" locations are not in the database, so the probability of performing work related (out-of-office) activities is penalized.

To overcome this issue, we define "shadow locations" those locations that are not described within our database. Then, we can introduce the number of shadow locations - n_s- which represents the trust we have in our database. A high value of $n_s$means that the available database poorly represents activities in that location. We can now calculate the error term as:

$$\varepsilon = n_s\left(Number\_Activities\right)^{-1} \tag{4}$$

---

* https://wiki.openstreetmap.org/wiki/Overpass

Equation (4) assumes that shadow locations are uniformly distributed with respect to the different activities we are considering. This also means that for $\varepsilon \to \infty$ equation (3) becomes also a uniform distribution (which is expected, since we do not have information about activity distribution in location L).

4. DisplayText cannot span more than one line!

In order to test the profiling, we used data from five individuals willing to share their data from the University of Luxembourg. Data is collected from the history of locations recorded by Google Map from respondents' smartphones. The time for each user varies from one to eight years. The tool can be accessed online (https://mobilab.lu/profiler-demo/) and tests can be done using the demo data provided or by loading any Google Map dataset, following the instructions provided (data will be locally stored – no sharing of sensitive data).

Results of the clustering phase correctly identified the activity type for activities "*Home*" and "*Work*", which was expected since the behaviour for these activities is very repetitive and close to the ideal one. Another interesting feature is the ability to capture the activity location changing over time. These results have been reported in (Toader et al. 2018). The same considerations do not hold for leisure activities. The reason is that the reference data (in red – Fig.1a) are calculated as the average behaviour for a reference population. For instance, it is unlikely that a user goes to the restaurant or to the gym every day. In this sense, the clustering technique still provides a reasonable result, but it will always underestimate the probability of performing leisure activities. For this reason, it is extremely important to consider both the proposed *Heuristic rule* and the *Bayesian updating*.

Table 1: Results of the profiling method on three users

	User 1				User 2				User 3			
	With Heuristic		No Heuristic		With Heuristic		No Heuristic		With Heuristic		No Heuristic	
	P_{work}	P_{home}	P_{work}	P_{home}	P_{work}	P_{home}	P_{work}	P_{home}	P_{work}	P_{home}	P_{work}	P_{home}
WP1	90.59	50.69	89.13	67.8	88.51	48.77	86.85	66.91	78.90	50.93	77.53	68.13
WP2	88.05	49.3	86.67	67.05	90.00	49.62	88.36	67.32	91.26	51.17	90.03	68.07
Leisure	70.67	38.2	73.97	61.4	67.6	77.03	73.83	80.94	75.03	51.09	79.07	68.2
Home	65.46	92.65	71.87	93.57	65.11	92.61	71.75	92.92	72,30	87.25	77.27	90.38

The heuristic improves the estimation for activities "*Home*" and "*Work*", which also means reducing the probability for these activities in all other cases. Similarly, the *Bayesian updating* introduces a prior probability that compensates the underestimation related to the data aggregation phase. Table I shows the improvement related to the *Heuristic rule* proposed in Sub-Section 3.2.

Specifically, we present the results for three users who accepted to collaborate during the validation phase. For each user, we calculated the probability of performing activity home (P_{home}) and work (P_{work}) in four different cases. All users have at least two work locations (WP1 and WP2). In one case this is related to office relocation while in the other two cases the reason is that users have multiple working locations. In this case, we can observe that, when the proposed heuristic is adopted, for all users the value of P_{work} increases while P_{home} decreases. A "*Leisure*" location has also been analysed. In this case, as expected, both P_{work} and P_{home} decreases when the proposed rule is implemented.

Finally, location "*Home*" has also been analysed. In this case, we would expect P_{work} to decrease and P_{home} to increase. However, as shown in Table 1, the estimated probability for location home is slightly lower when the heuristic rule is applied. However, we do not consider this as an issue, as in all cases the estimated probability is close to 90%, thus the model properly identifies the activity performed in that location. If it is true that, in this case, we reduce the probability, while looking at all analysed location, we clearly see that results are more consistent when the heuristic rule is implemented. Finally, we boost the prediction for leisure activities. To verify the result, we studied the activity "sport".

TABLE 2: Experiment Results

Results without GIS data:					
	$P^{Work}(U\|L)$	$P^{Home}(U\|L)$	$P^{eat}(U\|L)$	$P^{sport}(U\|L)$	$P^{Shop}(U\|L)$
	50.12 %	*0 %*	*23.2 %*	*41.12 %*	*0.13 %*

Results with GIS data:					
$\boldsymbol{n_s = 0}$	$P^{Work}(L\|U)$	$P^{Home}(L\|U)$	$P^{eat}(L\|U)$	$P^{sport}(L\|U)$	$P^{Shop}(L\|U)$
$\theta = 0.3$	*18.44 %*	*0 %*	*45.19 %*	*80.1 %*	*0.4 %*
$\theta = 0.6$	*32.96 %*	*0 %*	*35.11 %*	*62.23 %*	*0.8 %*
$\theta = 1$	*39.79 %*	*0 %*	*30.36 %*	*53.82 %*	*1.03 %*
$\boldsymbol{n_s = 1}$	$P^{Work}(L\|U)$	$P^{Home}(L\|U)$	$P^{eat}(L\|U)$	$P^{sport}(L\|U)$	$P^{Shop}(L\|U)$
$\theta = 0.3$	*27.61 %*	*0 %*	*38.82 %*	*68.81 %*	*0.7 %*
$\theta = 0.6$	*38.63 %*	*0 %*	*31.17 %*	*55.25 %*	*1 %*
$\theta = 1$	*43.27 %*	*0 %*	*27.95 %*	*49.54 %*	*1.12 %*
$\boldsymbol{n_s = 10}$	$P^{Work}(L\|U)$	$P^{Home}(L\|U)$	$P^{eat}(L\|U)$	$P^{sport}(L\|U)$	$P^{Shop}(L\|U)$
$\theta = 0.3$	*44.42 %*	*0 %*	*27.15 %*	*48.12 %*	*1.1 %*
$\theta = 0.6$	*47.3 %*	*0 %*	*25.15 %*	*44.58 %*	*1.2 %*
$\theta = 1$	*48.43 %*	*0 %*	*24.36 %*	*43.19 %*	*1.25 %*

The profiling phase showed us that, during lunchtime, one of the respondents was often visiting a location close to the University Campus. Based on OpenStreetMap, the area has both a sport centre and a restaurant, thus both options have to be considered. The respondent confirmed that he performs sport activity in that location three times a week during the lunch break. Table 2 shows the probability we estimated through the Bayesian updating process, for different values of n_s and θ. Table 2 shows that the probability obtained through Equation 1 is more reliable than the one derived only from the location history. For n_s equal to 0 or 1, it always identifies the right solution (i.e. sport is the most likely option in the given location), while for large values of n_s and θ the model collapses to the original estimation. It is also interesting to highlight that, for average parameter values (n_s=1 and $\theta = 0.6$), the Bayesian Approach returns similar values of $P^{Work}(L|U)$ and $P^{eat}(L|U)$, which is realistic considering that (i) the user is recurrently visiting that location during the lunch break and (ii) probability $P^{Work}(L|U)$ for that user to perform working activities at location L is greater than 50%.

5. Conclusions

The main purpose of this study was to explore the possibility to classify activities performed infrequently visited locations without any user report or additional information. Differently, from most of the state-of-the-art approaches, which present frameworks to collect travel information, the proposed framework leverages existing location history to infer activity location, meaning that there is no need for collecting new data. Our integrated framework uses three interconnected components to learn user behavior: (i) a clustering technique to identify the most likely activity performed in a location; (ii) a heuristic rule to explicitly account for user behaviour while estimating the "*Home*" and "*Work*" locations; (iii) GIS-based Bayesian Updating to include land use and properly estimate leisure activities.

Results reveal the tool's capabilities to automatically compute the probability that activity performed in each location is either at home, work, restaurant, daily shopping, sport, using only the location data. Moreover, this paper brings the following scientific and practical contributions:

- The proposed framework enables the process and analysis of travel data over a long time period (several years);
- For a limited number of activities, activity location can be identified without user validation;
- The methods identify dynamics such as activity relocation, which is essential information difficult to retrieve with traditional or digital travel surveys.

Future research will explore the possibility to include from external sources additional information regarding the visited locations (such as opening hours) and using pattern recognition techniques to match matrices and improve the clustering approach. Finally, the authors stress that working with a larger database is not feasible at the moment. With the increasing concern about privacy, users are becoming more aware of their rights. On one hand, this has a huge potential, as users can freely access their information and decide to share with the community. On the other hand, many users systematically delete their data, which is a clear limit for methodologies such as the one proposed in this paper. Given these assumptions, the authors aim at testing this methodology with anonymize data and validating this work with a test case on a larger number of users in the future.

Acknowledgements

This research has been partially sponsored by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462, the European Union (European Regional Development Fund) through FEDER (European Regional Development Fund) project MERLIN (R-AGR-3313-10-C) and by the Luxemburgish FNR (Fonds National de la Recherche) through an AFR grant for the PLAYMOBeL project (9220491).

References

Abdulazim, Tamer, Hossam Abdelgawad, Khandker M. Nurul Habib, and Baher Abdulhai. 2013. "Using Smartphones and Sensor Technologies to Automate Collection of Travel Data." *Transportation Research Record: Journal of the Transportation Research Board* 2383 (1): 44–52. https://doi.org/10.3141/2383-06.

Bohte, Wendy, and Kees Maat. 2009. "Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands." *Transportation Research Part C: Emerging Technologies* 17 (3): 285–97. https://doi.org/10.1016/j.trc.2008.11.004.

Bricka, Stacey G., Sudeshna Sen, Rajesh Paleti, and Chandra R. Bhat. 2012. "An Analysis of the Factors Influencing Differences in Survey-Reported and GPS-Recorded Trips." *Transportation Research Part C: Emerging Technologies* 21 (1): 67–88. https://doi.org/10.1016/j.trc.2011.09.005.

Byon, Young-Ji, Baher Abdulhai, and Amer Shalaby. 2009. "Real-Time Transportation Mode Detection via Tracking Global Positioning System Mobile Devices." *Journal of Intelligent Transportation Systems* 13 (4): 161–70. https://doi.org/10.1080/15472450903287781.

Chen, Cynthia, Jingtao Ma, Yusak Susilo, Yu Liu, and Menglin Wang. 2016. "The Promises of Big Data and Small Data for Travel Behavior (Aka Human Mobility) Analysis." *Transportation Research Part C: Emerging Technologies* 68 (July): 285–99. https://doi.org/10.1016/j.trc.2016.04.005.

Cottrill, Caitlin, Francisco Pereira, Fang Zhao, Iněs Dias, Hock Lim, Moshe Ben-Akiva, and P. Zegras. 2013. "Future Mobility Survey." *Transportation Research Record: Journal of the Transportation Research Board* 2354 (Oct.): 59–67. https://doi.org/10.3141/2354-07.

Ettema, Dick, Fabian Bastin, John Polak, and Olu Ashiru. 2007. "Modelling the Joint Choice of Activity Timing and Duration." *Transportation Research Part A: Policy and Practice*, Selected Papers on Applications of Discrete Choice Models Presented at the European Regional Science Conference, Amsterdam, August 2005, 41 (9): 827–41. https://doi.org/10.1016/j.tra.2007.03.001.

Greene, Elizabeth, Leah Flake, Kevin Hathaway, and Michael Geilich. 2016. "A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana." In . https://trid.trb.org/view/1394353.

Li, Zhigang Jason, and Amer S. Shalaby. 2008. "Web-Based GIS System for Prompted Recall of GPS-Assisted Personal Travel Surveys: System Development and Experimental Study." In . https://trid.trb.org/view.aspx?id=848806.

Nahmias-Biran, Bat-hen, Yafei Han, Shlomo Bekhor, Fang Zhao, Christopher Zegras, and Moshe Ben-Akiva. 2018. "Enriching Activity-Based Models Using Smartphone-Based Travel Surveys." *Transportation Research Record: Journal of the Transportation Research Board* 2672 (42). https://trid.trb.org/view/1497046.

Toader, Bogdan., Cantelmo, Guido, Popescu, Mioara, and Viti Francesco. 2018. "Using Passive Data Collection Methods to Learn Complex Mobility Patterns: An Exploratory Analysis." In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 993–98. https://doi.org/10.1109/ITSC.2018.8569935.

Toader, Bogdan, Moawad, Assaad, Fouquet, Francois, Hartmann, Thomas, Popescu, Mioara, and Viti Francesco. 2017. "A New Modelling Framework over Temporal Graphs for Collaborative Mobility Recommendation Systems." In 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 1–6. https://doi.org/10.1109/ITSC.2017.8317664..

Tsui, Sheung Yuen Amy, and Amer S. Shalaby. 2006. "Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1972. https://trid.trb.org/view.aspx?id=776863.

Zhao, Fang, Francisco Câmara Pereira, Rudi Ball, Youngsung Kim, Yafei Han, Christopher Zegras, and Moshe Ben-Akiva. 2015. "Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2494. https://trid.trb.org/view/1338572.

Zhou, Jianyu (Jack), and Reginald Golledge. 2007. "Real-Time Tracking of Activity Scheduling/Schedule Execution within a Unified Data Collection Framework." *Transportation Research Part A: Policy and Practice*, Bridging Research and Practice: A Synthesis of Best Practices in Travel Demand Modeling, 41 (5): 444–63. https://doi.org/10.1016/j.tra.2006.09.012.