

Original Paper

Inferring Destinations and Activity Types of Older Adults From GPS Data: Algorithm Development and Validation

Sayeh Bayat^{1,2}, BAsC; Gary Naglie^{2,3,4,5,6}, MD, FRCPC, FGSA; Mark J Rapoport^{7,8}, MD, FRCPC; Elaine Stasiulis^{5,9}, MA; Belkacem Chikhaoui¹⁰, PhD; Alex Mihailidis^{1,2,11}, PhD

¹Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

²KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

³Department of Medicine, Baycrest Health Sciences, Toronto, ON, Canada

⁴Department of Medicine, University of Toronto, Toronto, ON, Canada

⁵Rotman Research Institute, Baycrest Health Sciences, Toronto, ON, Canada

⁶Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

⁷Department of Psychiatry, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

⁸Department of Psychiatry, University of Toronto, Toronto, ON, Canada

⁹Institute of Medical Science, University of Toronto, Toronto, ON, Canada

¹⁰Laboratoire en Informatique Cognitive et Environnements de Formation Research Institute, Department of Science and Technology, TELUQ University, Montreal, QC, Canada

¹¹Department of Occupational Therapy and Occupational Science, University of Toronto, Toronto, ON, Canada

Corresponding Author:

Sayeh Bayat, BAsC

Institute of Biomaterials and Biomedical Engineering

University of Toronto

550 University Avenue

Toronto, ON

Canada

Phone: 1 416 597 3422 ext 7345

Email: sayeh.bayat@mail.utoronto.ca

Abstract

Background: Outdoor mobility is an important aspect of older adults' functional status. GPS has been used to create indicators reflecting the spatiotemporal dimensions of outdoor mobility for applications in health and aging. However, outdoor mobility is a multidimensional construct. There is, as of yet, no classification algorithm that groups and characterizes older adults' outdoor mobility based on its semantic aspects (ie, mobility intentions and motivations) by integrating geographic and domain knowledge.

Objective: This study assesses the feasibility of using GPS to determine semantic dimensions of older adults' outdoor mobility, including destinations and activity types.

Methods: A total of 5 healthy individuals, aged 65 years or older, carried a GPS device when traveling outside their homes for 4 weeks. The participants were also given a travel diary to record details of all excursions from their homes, including date, time, and destination information. We first designed and implemented an algorithm to extract destinations and infer activity types (eg, food, shopping, and sport) from the GPS data. We then evaluated the performance of the GPS-derived destination and activity information against the traditional diary method.

Results: Our results detected the stop locations of older adults from their GPS data with an F1 score of 87%. On average, the extracted home locations were within a 40.18-meter (SD 1.18) distance of the actual home locations. For the activity-inference algorithm, our results reached an F1 score of 86% for all participants, suggesting a reasonable accuracy against the travel diary recordings. Our results also suggest that the activity inference's accuracy measure differed by neighborhood characteristics (ie, Walk Score).

Conclusions: We conclude that GPS technology is accurate for determining semantic dimensions of outdoor mobility. However, further improvements may be needed to develop a robust application of this system that can be adopted in clinical practice.

(*JMIR Aging* 2020;3(2):e18008) doi: [10.2196/18008](https://doi.org/10.2196/18008)

KEYWORDS

outdoor mobility; older adults; GPS; life space; activity types; machine learning

Introduction

With the growing population of older adults, the notion of *aging in place*, defined as “the ability to live in one's own home and community safely, independently, and comfortably,” is of increasing importance [1]. Aging in place is desirable from both individual quality-of-life and public health perspectives. Many older adults prefer to age in place because it enables them to preserve their independence, autonomy, and social connection [2]. Furthermore, aging in place is favored by policy makers and health providers because keeping people in their homes for as long as possible can reduce institutional care costs [3]. Although most studies on aging in place focus on the home environment, there is rising recognition about the importance of neighborhoods and communities in older adults' ability to age in place [3-5]. Thus, many clinicians and researchers in the field of environmental gerontology are in search of a better understanding of older adults' travel patterns in the outdoor environment, a concept that will be referred to as *outdoor mobility* throughout the rest of this paper.

A commonly used model that focuses on older adults' outdoor mobility is *life space*, which is defined as the geographical area through which an individual moves [6]. To measure life space, the spatial environment is divided into several concentric zones (eg, bedroom, home, neighborhood, town, and out of town) and frequency of travel into each zone is recorded. Life space is traditionally measured through self-reported daily diaries or recall-based questionnaires [6,7]. The traditional life-space measures have many limitations. Travel diaries, for example, place the burden of data collection on study participants, which results in participants not recording trips that are too short or are made at inconvenient times. On the other hand, questionnaires require study participants to recall their movements during the month prior to the assessment date, which is particularly challenging for the cognitively impaired population. To overcome these limitations, more recent studies have used GPS data to create mobility indicators to assess life space, including maximum distance travelled, number of trips away from home, and area [8-11]. While most life-space studies rely only on a few indicators that represent spatial and temporal aspects of an individual's mobility [12-14], it has become clearer that mobility is a multidimensional construct [12,15,16]. Thus, in order to understand the full range of properties that emerge from mobility patterns, the construct must encapsulate the spatial, temporal, and semantic dimensions. In the current life-space mobility measures, semantic dimensions that would enable an exploration of the reasons behind the patterns that emerge in space and time are not included.

According to the *travel demand model*, travel is a derived demand, meaning that individuals travel to specific destinations in their environment in order to participate in certain activities [17]. Older adults' specific types of destinations and activities are indicative of their financial, psychosocial, physical, and cognitive status [18]. For example, participation in cognitively demanding activities, such as going to a bank, or physically

demanding activities, such as playing tennis, can relay information about the cognitive or physical health of the individual, respectively. Furthermore, an understanding of relevant destinations and activities of older adults is required to design environments that support and target effective interventions that promote older adults' outdoor mobility [18]. Therefore, examining the destinations that are relevant to older adults, as well as activities that older adults conduct at those destinations, is critical to the comprehensive understanding of their mobility behaviors.

However, the raw positioning data collected by GPS devices do not provide any additional contextual information, such as the places that people visit or the activities they perform. In travel behavior research, this information is traditionally collected through questionnaires where users are asked to annotate their trajectories. In the last few years, studies have increasingly aimed at automatically annotating raw GPS data with activities performed by users. The most popular method for automatically annotating GPS data in the existing travel behavior research is the rule-based method [19-23]. The available studies utilizing a rule-based method matched the GPS data with a series of predefined heuristic rules to determine the appropriate activity type. For example, Bohte and Maat [22] used a distance measure as a rule to determine the location that is being visited and the type of activity. They first obtained the location where a trip ends. If this location was within a radius of 50 m from a known location, it is assumed that this is the location that the user visited; otherwise, it was flagged as unknown. Furthermore, a number of studies used a probabilistic method to infer activity type [24-26]. These studies calculated the probability of participating in each of the potential activities according to a predefined measure and then selected the most probable activity type. For example, Furlotti et al [27] used car trajectories to infer the *point of interest* the user has visited. They then inferred the activity performed using the category of the point of interest and a probability measure based on the gravity law. These existing studies, however, have only focused on identifying destinations and inferring activity types at the aggregated level and have failed to examine mobility at individual levels.

Furthermore, there has been little work done to date on characterizing older adults' outdoor mobility according to its semantic aspects, such as destinations and activities. One approach entailed collecting activity-type information using self-report questionnaires [28]. This approach depends on the motivation and ability of the person to answer the questions accurately. An alternative approach has been to use ecological momentary assessments (EMAs) to classify types of activities. For example, in one study, activities of older adults within both residential and nonresidential environments were analyzed using EMAs [4]. That study was conducted during the course of 4 days, where EMA questions appeared as text on participants' iPhone screens for them to answer. The data in this study included a sample of participants' locations and activities within particular time windows and, therefore, was not comprehensive.

In addition, EMA procedures can be burdensome and time consuming if the frequency of prompts increases [29]. Finally, both traditional questionnaires and EMAs are usually completed only a few times with each person and cannot effectively provide a comprehensive and objective assessment over time [30]. To date, however, no study has used an automated GPS-based approach to characterize older adults' outdoor mobility according to its contextual information, such as destinations and activities. This study aims to develop and validate an inclusive and automated GPS-based outdoor mobility model for older adults that (1) extracts ones' destinations and (2) infers activity types conducted at the most relevant destinations.

Methods

Data Collection

A total of 5 individuals [31], 65 years of age or older, were recruited from a registry of potential research participants that is maintained at Baycrest Health Sciences in Toronto, Ontario. Exclusion criteria included cognitive impairment as determined by a Montreal Cognitive Assessment score of less than 26 [32], significant functional limitations in activities of daily living (ADL) or instrumental activities of daily living (IADL) [33], and residing in assisted-living or senior housing. Informed consent was obtained from all participants. Ethics approval was

obtained from the Baycrest Hospital Ethics Committee and the University Health Network Ethics Committee.

Participants were introduced to the SafeTracks Prime Mobile GPS device (SafeTracks GPS Canada) and instructed on how to use and charge it. The Prime Mobile device is small (3 cm × 5 cm × 2 cm) and light (less than 100 g). It automatically starts tracking when turned on, and its battery life lasts for about 8 hours of continuous tracking. Participants were asked to place these devices in their pockets, purses, or bags, or to wear them around their neck as a pendant. They were also provided with a small booklet and were instructed to record details describing all excursions from their home, including time and destination information. All participants completed 4 weeks of GPS data collection and travel diary recording. The 4-week study period was selected based on the time period used in the traditional life-space assessments [6,7].

The GPS device recorded location (ie, longitude and latitude coordinates), speed, and heading direction with corresponding time and date stamps. While in the *motion state*, the device provided sensor readings at a frequency of one reading per minute, and while in the *nonmotion state*, the device transitioned into standby mode until motion was detected. **Textbox 1** introduces the definitions and notations that are used throughout this paper.

Textbox 1. Terminology used in this paper.

Definition 1: *GPS Record (P)* is the spatiotemporal location of the user in the form of (*Lat, Lon, t, and v*), where *Lat* and *Lon* are the latitude and longitude coordinates, *t* is the date-time stamp, and *v* is the instantaneous speed.

Definition 2: *Trajectory (T)* is a set of GPS records that are ordered based on their date-time attributes *t*,

$$T = \{P_0, P_1, \dots, P_n\}, \text{ where } t_0 < t_1 < \dots < t_n$$

Definition 3: *Stop (S)* is a location at which the individual stays for more than a predefined time,

$$S = (\text{Lat}, \text{Lon}, \Delta t)$$

where Δt is the time the individual spent at the location, and the *Lat* and *Lon* coordinates are the centroid of all GPS points collected at the stop.

Since the stop measurements in the same location can vary, a stop cluster is defined.

Definition 4: *Stop Cluster* is a set of stops that belong to the same location.

Stop Detection

Algorithm

Two types of stops were considered when constructing a system to find the geographical locations of relevant destinations (ie, stops): (1) *full signal* and (2) *no signal* (see **Figure 1**). A *full-signal* stop was a stop location that contains a set of consecutive GPS records with no signal loss. The latitude and longitude of a full-signal stop was the centroid of all the GPS records in the stop. On the other hand, a *no-signal* stop was detected in two cases: (1) at areas with obstructed GPS signal and (2) when no motion is detected and the GPS device transitions into power-saving mode. This stop type consists of two GPS points: one immediately before and one immediately after the signal loss. The distance between these two adjacent points must be smaller than 150 meters. Otherwise, the signal loss can be due to underground transportation. The latitude and longitude of a *no-signal* stop was the centroid of the two GPS records in the stop. The velocity threshold was neglected

because the GPS records in a no-signal stop may belong to the trip prior to the stop and poststop.

The stop-detection method used three threshold values. First, the time threshold (δ_t) was set to 3 minutes to disregard short stops, such as stops at traffic lights, and to only detect the more meaningful destinations. Second, the distance threshold (δ_d) was selected to be 150 meters based on the average block size in the Greater Toronto Area. Third, the speed threshold (δ_v) was set to 2 m/s based on the SafeTracks GPS device speed-recording accuracy and the average walking speed for community-dwelling older adults ranging from 0.9 to 1.3 m/s [34].

To detect *full-signal* stops, initially, the first GPS record in the trajectory is added to the cluster. Then, each time a GPS record (P_i) is read, three measures are evaluated: (1) the time interval between the last GPS record (P_{i-1}) in the cluster and P_i , (2) the distance between P_{i-1} and P_i , and (3) the speed at P_i . If all values

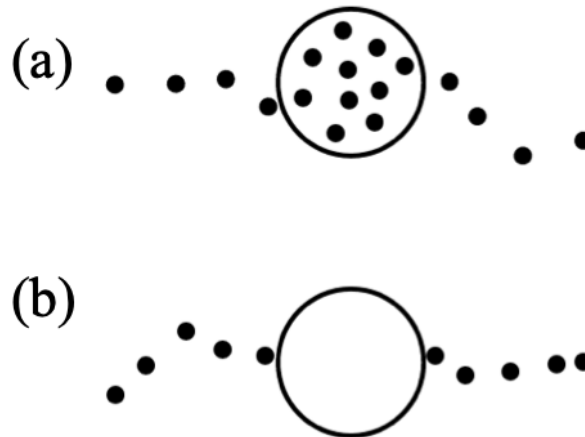
are less than their corresponding thresholds (δ_p , δ_d , and δ_v), the GPS record (P_i) is added to the cluster. Otherwise, we add the cluster to the list of stops (S), set the previous cluster to be the current cluster, and empty the current cluster.

To detect *no-signal* stops, we compute the distance and time interval between each GPS record and its previous record in the trajectory; if the extracted distance is less than the distance

threshold and the time interval is more than the time threshold, a cluster of the two GPS records is added to the stop list.

Finally, when a cluster is added to the stop list, the distance and time interval between the centroid of the cluster and the previous cluster is computed; if the values are less than the corresponding thresholds, the two clusters are merged, the previous cluster is set to the new merged cluster, and the current cluster is emptied.

Figure 1. Stop types: (a) full-signal stop and (b) no-signal stop.

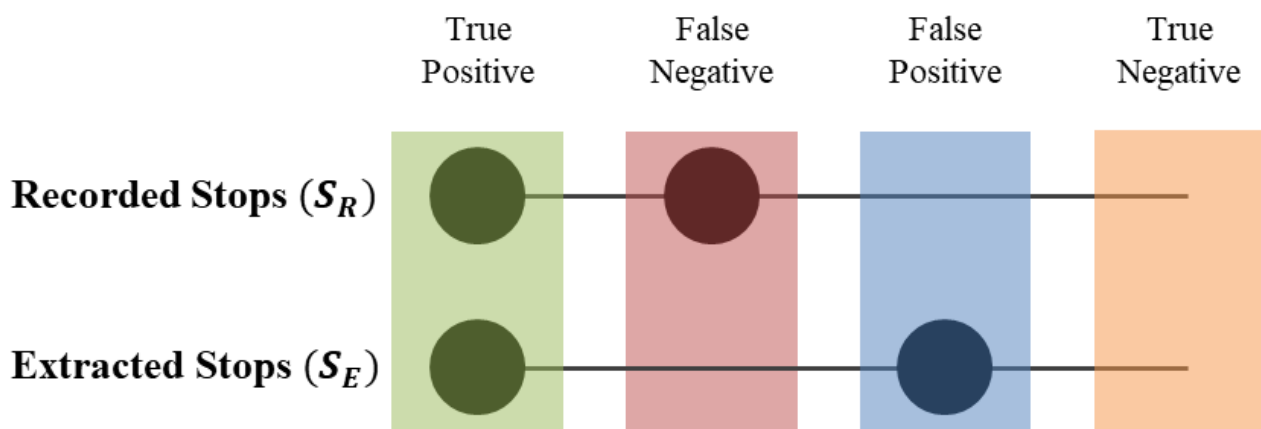


Evaluation

To evaluate the performance of the stop-detection algorithm, the stop points recorded in the travel diary (S_R) are compared with the stop points extracted from the GPS data (S_E). **Figure 2** demonstrates different stop conditions. A *true positive* (TP) refers to a stop location recorded in the travel diary that correctly matches with a stop location extracted from the GPS data. We

find a match if the distance between the recorded stop and the extracted stop is smaller than 150 meters. A *false negative* (FN) stands for a stop point recorded in the travel diary that is not extracted from the GPS data (ie, the algorithm considers it to be part of a trip or the user forgot to take the GPS device). A *false positive* (FP) refers to a stop point extracted from the GPS data but not recorded in the travel diary. Finally, a *true negative* (TN) occurs when no stop is extracted or recorded.

Figure 2. Comparison of the stops recorded in the travel diary versus the stops extracted from the GPS data.



Using the four stop conditions (ie, TP, TN, FP, and FN), the F1 score is determined to evaluate the performance of the stop-detection algorithm:

$$F1\ score = (2 \times precision \times recall) / (precision + recall)$$

where precision is $TP / (TP + FP)$ and recall is $TP / (TP + FN)$.

Activity Inference

Algorithm

Home

To infer the home location of each participant, the algorithm for density-based spatial clustering of applications with noise (DBSCAN) was used [35]. DBSCAN has been successfully used to find stop points with the most visits in GPS trajectories [36]. Two parameters that affect the results in DBSCAN are the

cluster radius (*Eps*) and the minimum number of points required to form a cluster (MinPts). The MinPts parameter was set to 4 as suggested in Ester et al [35] for 2D data. To obtain the optimal *Eps* value for each participant, we draw a k-distance plot for $k=4$ and find the *knee*, which corresponds to a sharp

change of gradient along the curve [37]. The fourth-nearest neighbor distance plot for participant 1 is presented in Figure 3. The knee point, which represents a change in density among the stop points, is selected as the optimal *Eps*. Table 1 demonstrates the optimal *Eps* value for each participant.

Figure 3. Participant 1's fourth-nearest neighbor distance. *Eps*: cluster radius.

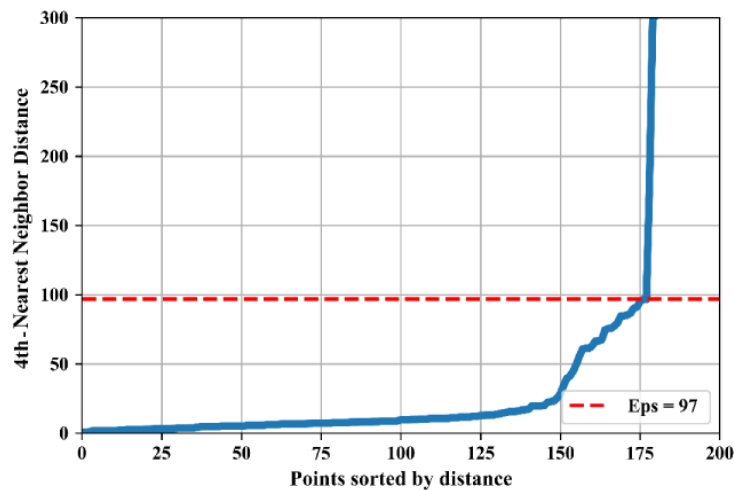


Table 1. Optimal cluster radius (*Eps*) value for each participant.

| Participant | <i>Eps</i> , m |
|-------------|----------------|
| 1 | 97 |
| 2 | 94 |
| 3 | 59 |
| 4 | 32 |
| 5 | 95 |

Other Activities

The activity-inference algorithm takes as input the list of stops and outputs the most probable activity type at each stop location. For each stop point (S_i), a Nearby Search was invoked from the Google Places application programming interface (API). The search returned a list of places within a 150-meter radius of the S_i . Then, for each place, a Place Details search was invoked, which returned details including place name, type, and opening hours. The place was only considered if it was open during the stop. Next, the place types were mapped into activity types

according to Table 2. If the Nearby Search returns nothing or if the place type does not fall into any of the listed categories, it would be placed into the *Other* category. To find the most probable activity type at each stop, the gravity model was implemented [27]. This model associated a probability to each possible activity by taking into account the distance of each place from the stop and the general characteristics of the stop location. For example, if a stop is in an area with many places that are mapped to food and few places that are mapped to medical services, the gravity model gives more weight to food as compared to medical services.

Table 2. Mapping between place types and activities.

| Activity category | Place type |
|-------------------|--|
| Food | Bakery, bar, cafe, food, meal takeaway, restaurant, and meal delivery |
| Daily shopping | Grocery and supermarket |
| Shopping | Bookstore, clothing store, convenience store, hardware store, electronics store, furniture store, shopping mall, liquor store, pet store, shoe store, and department store |
| Services | ATM (automatic teller machine), bank, car rental, car repair, finance, insurance, gas station, travel agency, post office, accounting, beauty salon, courthouse, and laundry |
| Leisure | Bowling alley, casino, library, movie rental, movie theater, museum, park, spa, stadium, and lodging |
| Medical services | Dentist, hospital, pharmacy, physiotherapist, chiropractor, psychologist, naturopath, walk-in clinic, sleep lab, LifeLabs ^a , and Dynacare ^a |
| Religious | Church, Hindu temple, synagogue, and mosque |
| Sport | Gym and YMCA |

^aLifeLabs and Dynacare are medical laboratory services companies based in Ontario.

Evaluation

From the TP stops, the ones annotated by the activity-inference algorithm are compared to the ones declared by the participant in the travel diary. For the activities annotated as *Activity at Home*, the distances between the participant's actual home location and extracted home locations are computed.

Results

Participants

A total of 5 cognitively intact, community-dwelling individuals completed 4 weeks (ie, 28 days) of GPS monitoring and travel diary recording. [Table 3](#) presents the summary statistics of the travel diary recordings for each participant.

The average age of participants in the study sample was 73 years (SD 6) (range 68-80); participants lived within the Greater Toronto Area and were all active drivers. Study participants had an average Montreal Cognitive Assessment (MoCA) score of 27.8 (SD 1.8). All 5 participants received a score of 6 in ADL and 8 in IADL, indicating the highest level of function. For more details of the sample's demographic characteristics, including the Walk Score [38], refer to [Table 4](#).

The sample of 5 consisted of 4 retired older adults (80%) (Participants 2-5) and 1 older adult (20%) with a part-time job (Participant 1). It should be noted that although employed, Participant 1 was not working during the 4 weeks of the study.

Table 3. Summary statistics of 4 weeks of travel diary recordings for the participants.

| Participant | Number of stops | Stops per day, mean (SD) |
|-------------|-----------------|--------------------------|
| 1 | 117 | 5.6 (2.5) |
| 2 | 89 | 3.4 (1.5) |
| 3 | 88 | 3.5 (1.8) |
| 4 | 124 | 4.6 (1.9) |
| 5 | 108 | 4.2 (1.9) |

Table 4. Sample demographic characteristics.

| Participant | Age in years | MoCA ^a score | Sex | ADL ^b score | IADL ^c score | Employment status | Driving status | Walk Score ^d |
|-------------|--------------|-------------------------|--------|------------------------|-------------------------|--------------------|----------------|-------------------------|
| 1 | 68 | 26 | Female | 6 | 8 | Employed part time | Driving | 65 |
| 2 | 70 | 26 | Female | 6 | 8 | Retired | Driving | 30 |
| 3 | 78 | 30 | Female | 6 | 8 | Retired | Driving | 57 |
| 4 | 68 | 29 | Male | 6 | 8 | Retired | Driving | 94 |
| 5 | 80 | 28 | Female | 6 | 8 | Retired | Driving | 69 |

^aMoCA: Montreal Cognitive Assessment.

^bADL: activities of daily living. ADL scores ranged from 0 (lowest level of function) to 6 (highest level of function).

^cIADL: instrumental activities of daily living. IADL scores ranged from 0 (lowest level of function) to 8 (highest level of function).

^dWalk Score is a measure of access to walkable amenities, ranging from 0 (Car-Dependent) to 100 (Walker's Paradise) [38].

Stop Detection

We reached a global stop-detection F1 score of 87% for all participants. The F1 score of each participant was computed and presented in Figure 4. The scores suggest that destinations of individuals can be detected with reasonable accuracy.

We further evaluated the stop detection by analyzing the mobility patterns. Figure 5 shows the distribution of stop

locations recorded in the travel diary (left) versus the ones extracted from GPS data (right) for Participant 1 over the 4 weeks of the study. Additionally, an envelope was built around all stop points using the convex hull algorithm to represent the extent of travel into the environment (ie, life-space area). From Figure 5, it is clear that the two distributions are similar, although some stop points are missing in the diary plot.

Figure 4. Stop-detection F1 scores for participants 1 to 5 (P1-P5).

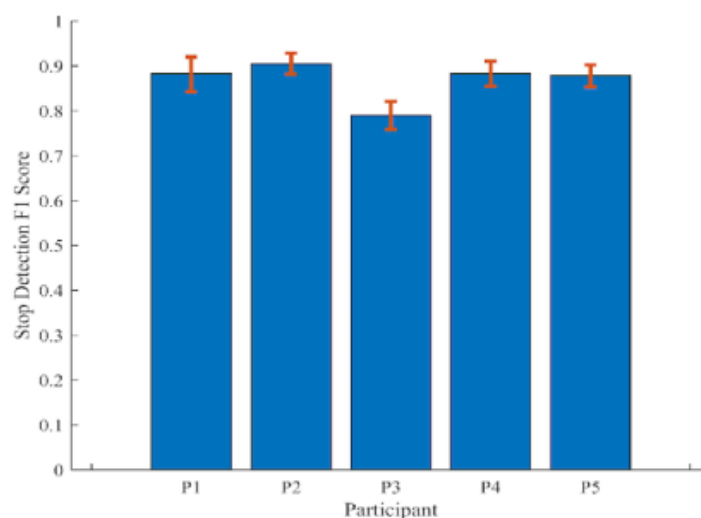
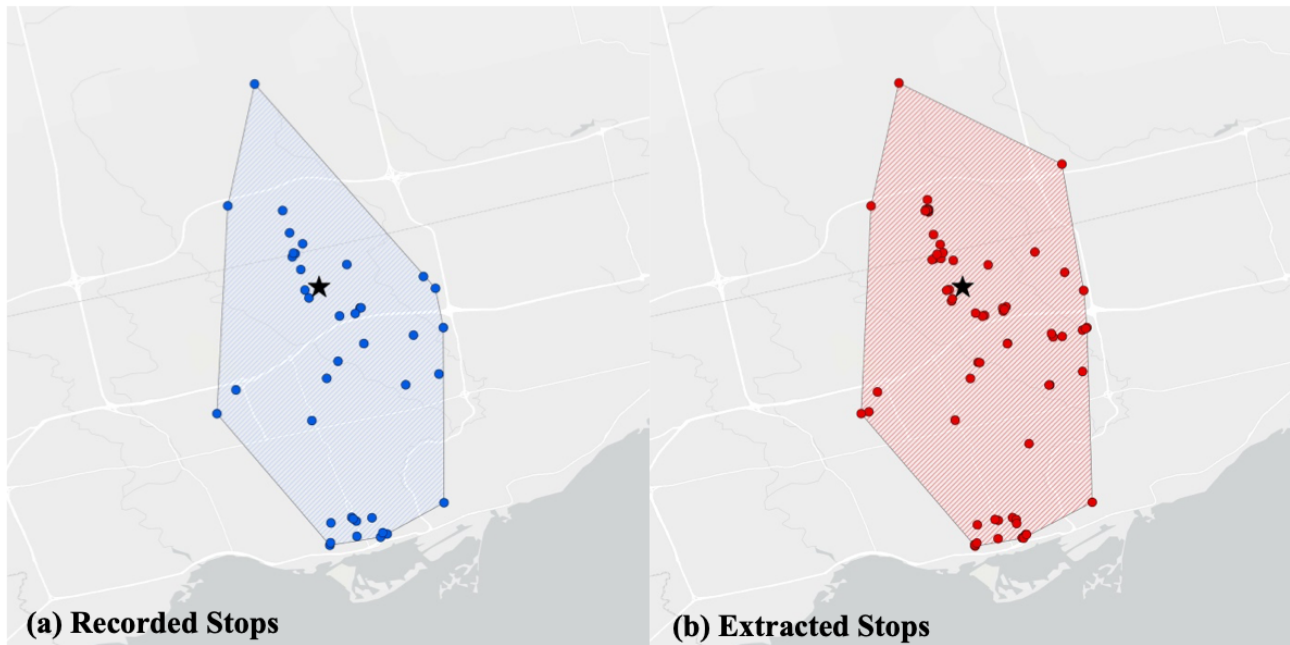


Figure 5. Comparison of (a) stops recorded in the travel diary versus (b) stops extracted from the GPS data for Participant 1.



Home and Activity Inference

Table 5 presents the distances between the participants' actual home locations and extracted home locations. On average, the extracted home locations were within a 40.18-meter (SD 1.18) distance of the actual home locations.

For all activity types including activities at *home*, F1 scores were obtained, which determine the percentage of activities correctly classified with respect to the activities recorded in the travel diary. Each participant's score is presented in Table 6. A global F1 score of 86% was reached for all participants.

Table 5. Distances between the actual home locations and extracted home locations.

| Participant | Distance, m |
|-------------|-------------|
| 1 | 6.60 |
| 2 | 50.3 |
| 3 | 19.9 |
| 4 | 20.4 |
| 5 | 76.7 |

Table 6. Activity inference's F1 score.

| Participant | F1 score | Number of stops |
|-------------|----------|-----------------|
| 1 | 0.89 | 74 |
| 2 | 0.90 | 79 |
| 3 | 0.86 | 101 |
| 4 | 0.79 | 95 |
| 5 | 0.86 | 87 |

The scores suggest that we can infer activity types with reasonable accuracy for all participants. It can be demonstrated that the algorithm reached its lowest score for Participant 4. To understand the reason behind this, the neighborhood characteristics of each participant's destinations were examined using Walk Score, a measure of access to walkable amenities, ranging from 0 (Car-Dependent) to 100 (Walker's Paradise) [38]; the distribution of scores are shown in Figure 6. It can be

demonstrated that for Participant 4, all destinations are located in neighborhoods with Walk Score values of greater than 80.

Furthermore, for personalized evaluation of the activity-inference algorithm, plots showing the percentage of activities that were inferred from the GPS data and declared in the travel diary for all 5 participants and each activity category are presented in Figure 7. In general, the percentage of inferred activities and percentage of declared activities follow a similar trend.

Figure 6. Distribution of Walk Scores of each participant's (P) destinations, as calculated by Walk Score [38]. A score of 0-24 or 25-49 is considered Car-Dependent, a score of 50-69 is considered Somewhat Walkable, a score of 70-89 is considered Very Walkable, and a score of 90-100 is considered Walker's Paradise.

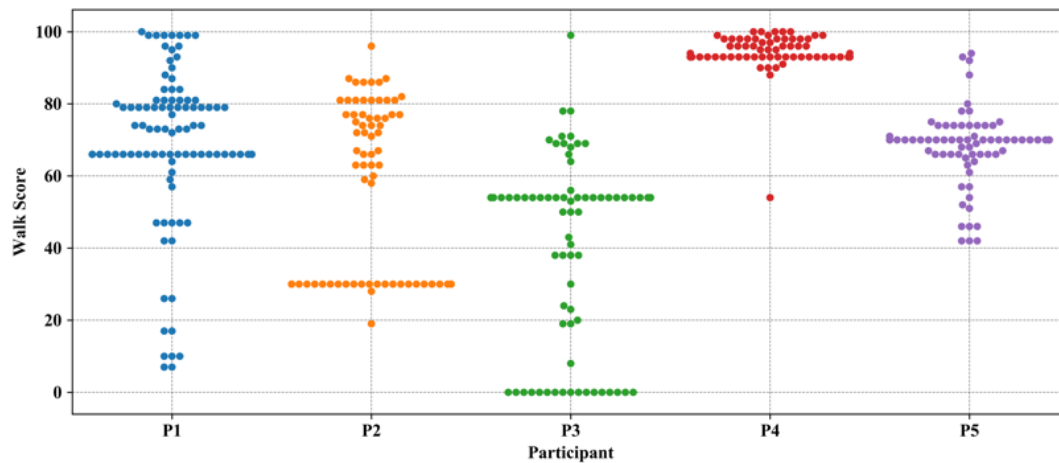
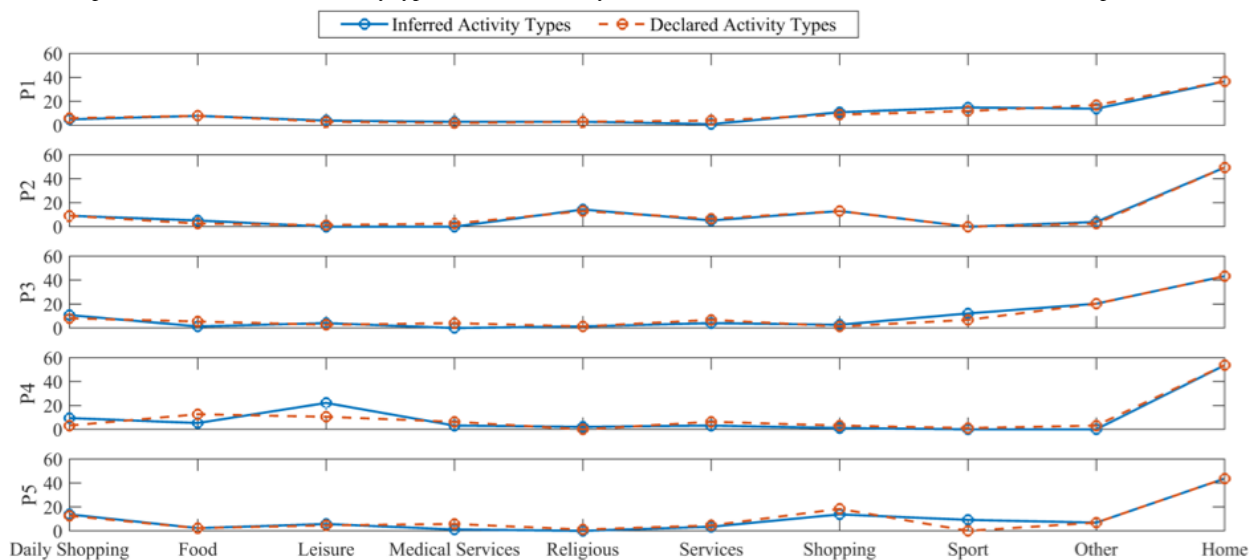


Figure 7. Comparison of the number of activity types in the travel diary versus the ones inferred from the GPS data for Participants 1 to 5 (P1-P5).



In order to further demonstrate the effects of neighborhood characteristics on the performance of activity inference, the out-of-home destinations were first placed into five groups according to their Walk Score [39]: (1) Walker's Paradise (90-100), where most errands can be accomplished on foot, and many people get by without owning a car; (2) Very Walkable (70-89) areas, where it is possible to accomplish most errands without driving; (3) Somewhat Walkable (50-69) areas, where some amenities are within walking distance, but many daily errands still rely on public transportation or driving; (4)

Car-Dependent (25-49) areas, where a few destinations are within walking distance, but most activities require driving or public transportation; and (5) Very Car-Dependent (0-24), where no neighborhood destinations are within walking range. Then, the average F1 score of the activity-inference algorithm in each group was determined. Table 7 illustrates that by moving from neighborhoods with a high density of amenities to neighborhoods with a low density of amenities, the performance of the activity-inference algorithm improves.

Table 7. Effects of Walk Score on performance of the activity-inference algorithm.

| Category | Walk Score ^a | F1 score | Number of stops |
|--------------------|-------------------------|----------|-----------------|
| Walker's Paradise | 90-100 | 0.58 | 58 |
| Very Walkable | 70-89 | 0.76 | 87 |
| Somewhat Walkable | 50-69 | 0.83 | 46 |
| Car-Dependent | 25-49 | 0.79 | 22 |
| Very Car-Dependent | 0-24 | 0.96 | 24 |

^aWalk Score is a measure of access to walkable amenities [38].

Discussion

Principal Findings

This paper presents a framework that allows classification of semantic aspects of outdoor mobility including destination and activity type. This framework is complementary to the GPS-based mobility indicators used in the literature to assess spatial and temporal facets of outdoor mobility and can enhance the understanding of older adults' outdoor mobility behavior.

The results show that, on average, 86% of activity types, including home activities, recorded in the travel diaries were correctly inferred. The performance of the activity-inference algorithm, however, depends on the neighborhood characteristics of destinations. If a destination is in a neighborhood with a high Walk Score (ie, Very Walkable or Walker's Paradise) where places are densely positioned, the algorithm can have difficulty identifying the correct places and the corresponding activity types; however, for destinations in less-walkable neighborhoods with a low Walk Score (ie, Very Car-Dependent, Car-Dependent, or Somewhat Walkable), the algorithm performed better. This pattern further hinders the performance of the algorithm when a participant's primary residence is in a high-Walk Score neighborhood, where their destinations are mostly concentrated within the same area, as illustrated with Participant 4's results in Figure 6. This renders the algorithm unable to precisely draw inferences, given the high-density life space, ultimately resulting in a lower F1 score. Furthermore, looking into the Walk Score of each participant's destinations (see Figure 6), it can be observed that among our 5 participants, Participant 4's destinations were positioned in locations with a higher Walk Score, and the algorithm reached the lowest activity-inference F1 score for this participant.

It is important to note that there are challenges associated with using travel diaries as ground truth. Participants' compliance with the diaries tends to decline over time. Due to the burden of manual data recording, participants may avoid or forget to record trips that are too short or that are made at inconvenient times. By examining the distributions of recorded and extracted stop locations in Figure 5, it can be observed that more stops were extracted from the GPS data in comparison to the stops recorded in the diary, suggesting that some stops in the GPS data were not recorded in the travel diary.

Furthermore, travel diaries included the street addresses of the destinations. In order to build ground truth of the destinations, we first geocoded each street address using the Google Maps

API. Given a street address, Google Maps API returned the latitude and longitude coordinates. However, this process is not always straightforward. For example, in the cases where the recorded address in the diary is incomplete, the correct coordinate may not be determined. To minimize these ambiguities, during the digitizing process of the travel diaries, the recorded addresses were verified. Since a verified address corresponds to a unique geographical coordinate, the destinations that were visited multiple times had the exact same latitude and longitude coordinates and are completely overlapping in Figure 5. This is not the case for the stops extracted from the GPS data. The GPS device does not record exactly the same coordinate for a unique destination. Even if the participant is staying at the same location for some time, the GPS records can vary by up to 15 meters. Therefore, in the GPS data, the coordinates corresponding to the same physical location can vary. This can be clearly observed in Figure 5 where there are multiple stops in close proximity belonging to the same location.

The mapping between the place types and activity categories is another important issue. The activity categories are inferred based on the place types available on the Google Maps. However, there are some occasions where the activities conducted at a specific place cannot be uniquely determined. For example, playing golf with some friends can be considered as *Sport* or *Leisure* activity. On the other hand, due to the limitations of GPSs in indoor settings, for some multipurpose locations the correct activity type cannot be inferred. For example, a movie theater inside a shopping center can be identified as a *Shopping* activity instead of a *Leisure* activity. This issue can be addressed in future studies by having a hybrid system that relies on a Wi-Fi positioning system for indoor tracking and GPS for outdoor tracking.

Our proposed model outperforms the existing activity-inference techniques. One major advantage of the proposed model is that it does not require any prior information regarding users' most-visited destinations. Most of the available techniques for inferring activity types from raw trajectories involve some initial data collection at the time of recruitment. These studies collect the addresses of each participant's most-visited locations, such as home, workplace, school, or the frequently used grocery stores [20-22], and claim that these locations yield more than 60% of each participant's activities [21]. Our presented model, however, is fully autonomous and uses a density-based clustering method to extract the most-visited destinations of each participant. This improvement is particularly important

for exploring activity profiles of the older adult population because their most-visited destinations are not always an educational establishment or workplace. Another advantage is that while most of the prevailing techniques use a small list of points of interest to infer activity types [20,22], our model takes advantage of the Google Places API system, which is currently the most comprehensive dataset of points of interest. Additionally, while most studies use a uniform probability measure to select the most appropriate activity category [21,22,25], our model implemented a gravity-based approach similar to the method introduced in Furletti et al [27] to select the most probable activity category for each destination.

Furthermore, most studies that present empirical results of activity-inference methods only focus on car trips [24,26,27]. The GPS data used in this study were collected by participants during their out-of-home trips using various modes of transportation. It is worth noting that although the gravity-based probability measure used to infer activity types in our study was similar to the approach introduced in Furletti et al [27], our results achieved a higher accuracy. This is because in Furletti et al [27], they used positioning data collected by tracking devices installed in cars. In this case, the identification of the stop locations can become problematic because a car usually cannot enter inside the stop location, meaning that the participant needs to park the car and then walk to the destination, which will not be tracked by the GPS device. In our study, however, the GPS devices were being carried by the participants inside all the stop locations they visited. Furthermore, our activity-inference algorithm also discusses the detection of the home location, which was not included in Furletti et al [27].

Limitations

The findings should be understood in light of some inherent limitations within this study, which can be addressed in future research. First, the small sample size in this study, although effective for demonstrating the feasibility of the developed algorithms [31], is not representative of the population and is, thus, not suitable for statistical analysis of the population. Further studies on larger sample sizes are required in order to comprehensively analyze the activity profiles of older adult populations. Second, a cutoff of 65 years of age was used in this study, resulting in a sample age range of 68–80 years. This wide age range along with the small sample size prohibited analysis on age-related changes in the activity profiles of older adults. Future studies can divide the older adult population into

three subgroups—the young-old (65–74 years), the middle-old (75–84 years), and the old-old (over 85 years)—and investigate the mobility and activity profile of each subgroup separately. Third, all the participants were from the Greater Toronto Area. Further research is required on older adults from a set of representative locations that reflect the climatic, socioeconomic, and geographic diversity of the older adult population. Finally, due to recruitment challenges, our sample contained an unbalanced sex distribution (ie, 1 male and 4 females), which prohibited any statistical analysis by sex or gender. Future studies should examine gender differences, since they may be attributable to a variety of interrelated factors, including differences in perceptions of safety and cultural norms regarding outdoor mobility.

Conclusions

In environmental gerontology research, GPS devices are becoming increasingly more common to accurately and continuously measure older adults' outdoor mobility, thereby addressing limitations of traditional self-reported measures, such as recall biases. Outdoor mobility, however, is a multidimensional concept and it is challenging to characterize it comprehensively with only spatiotemporal indicators derived from GPS data.

In this paper, we extend the literature on older adults' mobility models through development and validation of a framework that relies on GPS data to capture older adults' travel destinations (ie, stop points) and activity types. We have performed a comparison with ground truth based on travel diaries, and we have evaluated in detail the performance of the implemented stop-detection and activity-inference algorithms. Our results indicate that it is possible to extract destinations and infer activity types from GPS data with reasonable accuracy.

This paper encourages incorporation of GPS-based mobility indicators that reflect the semantic dimension of individuals' outdoor mobility into future health- and aging-related research. This approach fosters a better understanding of what aspects of mobility are key to healthy aging. It also shows great potential in examining the impact of interventions and long-term monitoring of social connection, functionality, and quality of life. Future research should aim to utilize GPS technology to assess older adults' transportation modes in order to provide insights about different ways of conceptualizing older adults' environmental exposure.

Acknowledgments

This research was supported with funding from the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the AGE-WELL NCE (Aging Gracefully across Environments using Technology to Support Wellness, Engagement and Long Life Networks of Centres of Excellence). The CCNA is supported by a grant from the Canadian Institutes of Health Research (CAN 137794), with funding from several partners. This research was initiated as a collaborative study conducted by Baycrest Hospital, Sunnybrook Health Sciences Centre, and the University of Toronto. GN was supported by the George, Margaret and Gary Hunt Family Chair in Geriatric Medicine, University of Toronto. MJR receives research salary support from the Sunnybrook Psychiatry Partnership.

Conflicts of Interest

None declared.

References

1. Healthy places terminology. Centers for Disease Control and Prevention. 2009 Oct 15. URL: <https://www.cdc.gov/healthyplaces/terminology.htm> [accessed 2020-07-10]
2. Iecovich E. Aging in place: From theory to practice. *Anthropol Notebooks* 2014;20(1):21-33 [FREE Full text]
3. Wiles JL, Leibing A, Guberman N, Reeve J, Allen RES. The meaning of "aging in place" to older people. *Gerontologist* 2012 Jun;52(3):357-366. [doi: [10.1093/geront/gnr098](https://doi.org/10.1093/geront/gnr098)] [Medline: [21983126](https://pubmed.ncbi.nlm.nih.gov/21983126/)]
4. York Cornwell E, Cagney KA. Aging in activity space: Results from smartphone-based GPS-tracking of urban seniors. *J Gerontol B Psychol Sci Soc Sci* 2017 Sep 01;72(5):864-875 [FREE Full text] [doi: [10.1093/geronb/gbx063](https://doi.org/10.1093/geronb/gbx063)] [Medline: [28586475](https://pubmed.ncbi.nlm.nih.gov/28586475/)]
5. Lui C, Everingham J, Warburton J, Cuthill M, Bartlett H. What makes a community age-friendly: A review of international literature. *Australas J Ageing* 2009 Sep;28(3):116-121. [doi: [10.1111/j.1741-6612.2009.00355.x](https://doi.org/10.1111/j.1741-6612.2009.00355.x)] [Medline: [19845650](https://pubmed.ncbi.nlm.nih.gov/19845650/)]
6. Stalvey BT, Owsley C, Sloane ME, Ball K. The Life Space Questionnaire: A measure of the extent of mobility of older adults. *J Appl Gerontol* 2016 Jun 29;18(4):460-478. [doi: [10.1177/073346489901800404](https://doi.org/10.1177/073346489901800404)]
7. Peel C, Sawyer Baker P, Roth DL, Brown CJ, Brodner EV, Allman RM. Assessing mobility in older adults: The UAB Study of Aging Life-Space Assessment. *Phys Ther* 2005 Oct;85(10):1008-1119. [Medline: [16180950](https://pubmed.ncbi.nlm.nih.gov/16180950/)]
8. Schenk AK, Witbrodt BC, Hoarty CA, Carlson RH, Goulding EH, Potter JF, et al. Cellular telephones measure activity and lifespaces in community-dwelling adults: Proof of principle. *J Am Geriatr Soc* 2011 Feb;59(2):345-352 [FREE Full text] [doi: [10.1111/j.1532-5415.2010.03267.x](https://doi.org/10.1111/j.1532-5415.2010.03267.x)] [Medline: [21288235](https://pubmed.ncbi.nlm.nih.gov/21288235/)]
9. Boissy P, Brière S, Hamel M, Jog M, Speechley M, Karelis A, EMAP Group. Wireless inertial measurement unit with GPS (WIMU-GPS): Wearable monitoring platform for ecological assessment of lifespaces and mobility in aging and disease. *Conf Proc IEEE Eng Med Biol Soc* 2011;2011:5815-5819. [doi: [10.1109/IEMBS.2011.6091439](https://doi.org/10.1109/IEMBS.2011.6091439)] [Medline: [22255662](https://pubmed.ncbi.nlm.nih.gov/22255662/)]
10. Wettstein M, Wahl H, Shoval N, Auslander G, Oswald F, Heinik J. Identifying mobility types in cognitively heterogeneous older adults based on GPS-tracking: What discriminates best? *J Appl Gerontol* 2015 Dec;34(8):1001-1027. [doi: [10.1177/0733464813512897](https://doi.org/10.1177/0733464813512897)] [Medline: [24652916](https://pubmed.ncbi.nlm.nih.gov/24652916/)]
11. Brusilovskiy E, Klein LA, Salzer MS. Using global positioning systems to study health-related mobility and participation. *Soc Sci Med* 2016 Jul;161:134-142. [doi: [10.1016/j.socscimed.2016.06.001](https://doi.org/10.1016/j.socscimed.2016.06.001)] [Medline: [27289268](https://pubmed.ncbi.nlm.nih.gov/27289268/)]
12. Fillekes MP, Giannouli E, Kim E, Zijlstra W, Weibel R. Towards a comprehensive set of GPS-based indicators reflecting the multidimensional nature of daily mobility for applications in health and aging research. *Int J Health Geogr* 2019 Jul 24;18(1):17 [FREE Full text] [doi: [10.1186/s12942-019-0181-0](https://doi.org/10.1186/s12942-019-0181-0)] [Medline: [31340812](https://pubmed.ncbi.nlm.nih.gov/31340812/)]
13. Giannouli E, Bock O, Mellone S, Zijlstra W. Mobility in old age: Capacity is not performance. *Biomed Res Int* 2016;2016:1-8 [FREE Full text] [doi: [10.1155/2016/3261567](https://doi.org/10.1155/2016/3261567)] [Medline: [27034932](https://pubmed.ncbi.nlm.nih.gov/27034932/)]
14. Cuiagnet T, Perchoux C, Caruso G, Klein O, Klein S, Chaix B, et al. Mobility among older adults: Deconstructing the effects of motility and movement on well-being. *Urban Stud* 2019 Jul 16;57(2):383-401. [doi: [10.1177/0042098019852033](https://doi.org/10.1177/0042098019852033)]
15. Webber SC, Porter MM, Menec VH. Mobility in older adults: A comprehensive framework. *Gerontologist* 2010 Aug;50(4):443-450. [doi: [10.1093/geront/gnq013](https://doi.org/10.1093/geront/gnq013)] [Medline: [20145017](https://pubmed.ncbi.nlm.nih.gov/20145017/)]
16. Jin M, Claramunt C. A semantic model for human mobility in an urban region. *J Data Semant* 2018 Aug 25;7(3):171-187. [doi: [10.1007/s13740-018-0092-4](https://doi.org/10.1007/s13740-018-0092-4)]
17. Jones PM. A New Approach to Understanding Travel Behaviour and its Implications for Transportation Planning [doctoral thesis]. London, UK: University of London, Imperial College of Science and Technology; 1983. URL: <https://spiral.imperial.ac.uk/bitstream/10044/1/36320/2/Jones-PM-1983-PhD-Thesis.pdf> [accessed 2020-07-10]
18. Chudyk AM, Winters M, Moniruzzaman M, Ashe MC, Gould JS, McKay H. Destinations matter: The association between where older adults live and their travel behavior. *J Transp Health* 2015 Mar;2(1):50-57 [FREE Full text] [doi: [10.1016/j.jth.2014.09.008](https://doi.org/10.1016/j.jth.2014.09.008)] [Medline: [27104147](https://pubmed.ncbi.nlm.nih.gov/27104147/)]
19. Wolf J, Guensler R, Bachman W. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transp Res Rec* 2001 Jan;1768(1):125-134. [doi: [10.3141/1768-15](https://doi.org/10.3141/1768-15)]
20. Wolf J, Schönfelder S, Samaga U, Oliveira M, Axhausen K. Eighty weeks of global positioning system traces: Approaches to enriching trip information. *Transp Res Rec* 2004 Jan;1870(1):46-54. [doi: [10.3141/1870-06](https://doi.org/10.3141/1870-06)]
21. Stopher P, FitzGerald C, Zhang J. Search for a global positioning system device to measure person travel. *Transp Res Part C Emerg Technol* 2008 Jun;16(3):350-369 [FREE Full text] [doi: [10.1016/j.trc.2007.10.002](https://doi.org/10.1016/j.trc.2007.10.002)]
22. Bohte W, Maat K. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transp Res Part C Emerg Technol* 2009 Jun;17(3):285-297. [doi: [10.1016/j.trc.2008.11.004](https://doi.org/10.1016/j.trc.2008.11.004)]
23. Chen C, Gong H, Lawson C, Bialostozky E. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transp Res Part A Policy Pract* 2010 Dec;44(10):830-840. [doi: [10.1016/j.tra.2010.08.004](https://doi.org/10.1016/j.tra.2010.08.004)]
24. Axhausen KW, Schönfelder S, Wolf J, Oliveira M, Samaga U. 80 Weeks of GPS-Traces: Approaches to Enriching the Trip Information. Zürich, Switzerland: ETH Zürich; 2003 Aug. URL: <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/147519/eth-26551-01.pdf?sequence=1&isAllowed=y> [accessed 2020-07-10]

25. Spinsanti L, Celli F, Renso C. Where you stop is who you are: Understanding people's activities by places visited. In: Proceedings of the 5th Workshop on Behaviour Monitoring and Interpretation (BMI'10). 2010 Presented at: 5th Workshop on Behaviour Monitoring and Interpretation (BMI'10); September 21, 2010; Karlsruhe, Germany URL: https://pdfs.semanticscholar.org/dcbe/65f3e7baf9d654aa36edb7c9cef5de498207.pdf?_ga=2.66874086.916907263.1594433898-245535511.1590015772
26. Li A, Axhausen KW. Trip purpose imputation for taxi data. In: Proceedings of the 18th Swiss Transport Research Conference (STRC 2018). 2018 Presented at: 18th Swiss Transport Research Conference (STRC 2018); May 16-18, 2018; Ascona, Switzerland URL: <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/264721/ab1341.pdf?sequence=1&isAllowed=y> [doi: [10.3929/ethz-b-000264721](https://doi.org/10.3929/ethz-b-000264721)]
27. Furletti B, Cintia P, Renso C, Spinsanti L. Inferring human activities from GPS tracks. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing. New York, NY: Association for Computing Machinery; 2013 Presented at: 2nd ACM SIGKDD International Workshop on Urban Computing; August 11-14, 2013; Chicago, IL p. 1-8. [doi: [10.1145/2505821.2505830](https://doi.org/10.1145/2505821.2505830)]
28. Wettstein M, Wahl H, Shoval N, Oswald F, Voss E, Seidl U, et al. Out-of-home behavior and cognitive impairment in older adults: Findings of the SenTra Project. *J Appl Gerontol* 2015 Feb;34(1):3-25. [doi: [10.1177/0733464812459373](https://doi.org/10.1177/0733464812459373)] [Medline: [25548086](https://pubmed.ncbi.nlm.nih.gov/25548086/)]
29. Yang YS, Ryu GW, Han I, Oh S, Choi M. Ecological momentary assessment using smartphone-based mobile application for affect and stress assessment. *Healthc Inform Res* 2018 Oct;24(4):381-386 [FREE Full text] [doi: [10.4258/hir.2018.24.4.381](https://doi.org/10.4258/hir.2018.24.4.381)] [Medline: [30443427](https://pubmed.ncbi.nlm.nih.gov/30443427/)]
30. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. *Cogn Behav Pract* 2015 Feb;22(1):49-59 [FREE Full text] [doi: [10.1016/j.cbpra.2014.01.010](https://doi.org/10.1016/j.cbpra.2014.01.010)] [Medline: [27330267](https://pubmed.ncbi.nlm.nih.gov/27330267/)]
31. Nielsen J. How many test users in a usability study? Nielsen Norman Group. 2012 Jun 03. URL: <http://www.nngroup.com/articles/how-many-test-users/> [accessed 2013-03-21]
32. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005 Apr;53(4):695-699. [doi: [10.1111/j.1532-5415.2005.53221.x](https://doi.org/10.1111/j.1532-5415.2005.53221.x)] [Medline: [15817019](https://pubmed.ncbi.nlm.nih.gov/15817019/)]
33. Lawton MP, Brody EM. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist* 1969 Sep 01;9(3 Part 1):179-186. [doi: [10.1093/geront/9.3_part_1.179](https://doi.org/10.1093/geront/9.3_part_1.179)]
34. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: Reference values and determinants. *Age Ageing* 1997 Jan;26(1):15-19. [doi: [10.1093/ageing/26.1.15](https://doi.org/10.1093/ageing/26.1.15)] [Medline: [9143432](https://pubmed.ncbi.nlm.nih.gov/9143432/)]
35. Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96). Palo Alto, CA: AAAI Press; 1996 Presented at: Second International Conference on Knowledge Discovery and Data Mining (KDD'96); August 2-4, 1996; Portland, OR p. 226-231. [doi: [10.5555/3001460.3001507](https://doi.org/10.5555/3001460.3001507)]
36. Tang J, Jiang H, Li Z, Li M, Liu F, Wang Y. A two-layer model for taxi customer searching behaviors using GPS trajectory data. *IEEE trans Intell Transp Syst* 2016 Nov;17(11):3318-3324. [doi: [10.1109/tits.2016.2544140](https://doi.org/10.1109/tits.2016.2544140)]
37. Gaonkar MN, Sawant K. AutoEpsDBSCAN: DBSCAN with Eps automatic for large dataset. *Int J Adv Comput Theory Eng* 2013;2(2):2319-2526 [FREE Full text]
38. Walk Score methodology. Walk Score. 2020. URL: <https://www.walkscore.com/methodology.shtml> [accessed 2020-01-28]
39. Carr LJ, Dunsiger SI, Marcus BH. Validation of Walk Score for estimating access to walkable amenities. *Br J Sports Med* 2011 Nov;45(14):1144-1148 [FREE Full text] [doi: [10.1136/bjism.2009.069609](https://doi.org/10.1136/bjism.2009.069609)] [Medline: [20418525](https://pubmed.ncbi.nlm.nih.gov/20418525/)]

Abbreviations

ADL: activities of daily living

AGE-WELL NCE: Aging Gracefully across Environments using Technology to Support Wellness, Engagement and Long Life Networks of Centres of Excellence

API: application programming interface

ATM: automatic teller machine

CCNA: Canadian Consortium on Neurodegeneration in Aging

DBSCAN: density-based spatial clustering of applications with noise

EMA: ecological momentary assessment

Eps: cluster radius

FN: false negative

FP: false positive

IADL: instrumental activities of daily living

MinPts: minimum number of points required to form a cluster

MoCA: Montreal Cognitive Assessment

TN: true negative

TP: true positive

Edited by R Casado Arroyo; submitted 31.01.20; peer-reviewed by V Costela-Ruiz, M Hawley; comments to author 20.02.20; revised version received 19.03.20; accepted 01.05.20; published 28.07.20

Please cite as:

*Bayat S, Naglie G, Rapoport MJ, Stasiulis E, Chikhaoui B, Mihailidis A
Inferring Destinations and Activity Types of Older Adults From GPS Data: Algorithm Development and Validation
JMIR Aging 2020;3(2):e18008
URL: <http://aging.jmir.org/2020/2/e18008/>
doi: [10.2196/18008](https://doi.org/10.2196/18008)
PMID:*

©Sayeh Bayat, Gary Naglie, Mark J Rapoport, Elaine Stasiulis, Belkacem Chikhaoui, Alex Mihailidis. Originally published in JMIR Aging (<http://aging.jmir.org>), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Aging, is properly cited. The complete bibliographic information, a link to the original publication on <http://aging.jmir.org>, as well as this copyright and license information must be included.