Extracting Significant Places from Mobile User GPS Trajectories: A Bearing Change Based Approach

Tanusri Bhattacharya, Lars Kulik, James Bailey
The University of Melbourne
Department of Computing and Information Systems, Melbourne, Australia
tanusrib@student.unimelb.edu.au,{lkulik, baileyj}@unimelb.edu.au

ABSTRACT

Moving object data, in particular of mobile users, is becoming widely available. A GPS trajectory of a moving object is a time-stamped sequence of latitude and longitude coordinates. The analysis and extraction of knowledge from GPS trajectories is important for a range of applications. Existing studies have extracted knowledge from trajectory patterns for both single and multiple GPS trajectories. However, few works have taken into account the unreliability of GPS measurements for mobile devices or focused on the extraction of fine-grained events from a user's GPS trajectory, such as waiting in traffic, at an intersection, or at a bus stop. In this paper, we develop and experimentally evaluate a novel algorithm that analyses a mobile user's bearing change distribution, together with speed and acceleration, to extract significant places of events from their GPS trajectory.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Spatial databases and GIS

General Terms

Algorithms, Design

Keywords

Trajectory mining, Significant Places, GPS trajectories

1. INTRODUCTION

GPS enabled devices, such as smart phones, have gained large popularity in the last few years. It allows people to record their outdoor activities through GPS trajectories, which enables researchers to infer further knowledge about the moving behaviour of mobile users. An important theme in analysing GPS data is the identification of significant or interesting places visited by users [1, 2, 6, 8, 9]. Such data may be sourced from a single or multiple GPS trajectories.

GPS enabled devices can record user trajectories as time-stamped sequences of latitude and longitude coordinates. We call a location where a user stays for an extended time a *significant place*. Significant places of a mobile user are those segments (consisting of consecutive GPS points) from a GPS trajectory which are (1) bound by some time and/or distance thresholds [6, 8, 9] or (2) intersect with the geometry of a Point of Interest (POI) [1, 2]. This task is difficult due to the unreliable GPS measurements of mobile devices, e.g., caused by battery outage, signal loss or inaccuracy of the GPS receiver. This is exacerbated by the unpredictable behaviour of mobile users, e.g., switching off the phone. The inaccuracy of GPS devices makes it also difficult to differentiate whether a user is stationary at a place such as a bus-stop or slowly walking.

In the existing literature, the word "significant" mainly refers to places where the user actively or purposefully spends some time, such as home, office, tourist attraction, shopping mall, or a park. However, there is another class of places along a trajectory where a user has to stop or wait, e.g., at an intersection, at a bus-stop, or in a traffic jam. The significance of these places depends on the length of the time period and on the application: while for some applications, e.g., interesting places for a tourist, they may be insignificant, for other applications such as traffic and transportation management systems, they can provide important information to increase the performance of the system [3].

To identify significant places (where the user either actively or forcefully spends some time) from unreliable GPS measurements while considering unpredictable movements of mobile users, we develop a novel algorithm that takes speed, acceleration, bearing and bearing change into account. We categorize the significant places as either *point-like* (e.g. at an office location, a restaurant or a bus-stop where a user is at rest) or *extended* (e.g., a market or a shopping mall that requires slow movement to explore).

The two key elements of our algorithm are (a) direction or bearing and (b) speed and acceleration. The main insight is that when a user is at a point-like or extended place the bearing is random in nature, which results in a large average bearing change. Whereas typically places are identified through the use of speed (and acceleration), which is approximately zero when a user is at a *point-like* place or small when a user is at an *extended* place, we suggest that the use of bearing change is equally important. Even if a user is at rest, the GPS readings will still introduce a "pseudo" movement due to GPS noise and result in a random bearing change behaviour. A user exploring a place such as a market also generates large average bearing changes, because the user has no distinct bearing. However, when a user is moving along a path, the user follows a particular direction, which results in small average bearing changes.

As a result, we use a two stage algorithm to extract significant places. In the first stage, we bucket speed and acceleration to identify all those segments of a GPS trajectory that are likely to be *point-like* and *extended* places. This technique may still return

ACM SIGSPATIAL GIS '12, November 6-9, 2012. Redondo Beach, CA, USA

segments of the trajectory where the mobile user, for example, is slowly driving by car or walking on a path. In the second stage, we use a bearing based approach to differentiate the moving segments from the actual significant places. In our bearing based approach, we particularly look at the distribution of bearing change and compute the standard deviation for each extracted segment.

Our main contributions are as follows: (1) instead of relying on a predetermined distance or time threshold, or considering additional resources such as geometry of Point Of Interests (POIs) or map-matching techniques, to identify a place as significant, we propose a technique that automatically extracts all significant places and their corresponding durations. (2) Our approach can extract significant places considering the unreliability of GPS measurements for mobile devices. (3) To our knowledge, we are the first to consider information about a user's bearing change distribution for identifying significant places from user's GPS trajectory.

2. EXTRACTING SIGNIFICANT PLACES

The extraction of significant places requires two concepts: direction and speed. Conceptually, a significant point-like place such as an office location is a part of a GPS trajectory where the user does not move and his bearing could have any value due to GPS noise. Similarly, for a significant extended place like a market, the user typically moves slowly but in random directions, resulting in significant bearing changes between consecutive GPS points. In summary, the significant places are those parts of a user's trajectory where the bearing is random (undirected) and the speed and acceleration is either close to zero (for *point-like* places) or the speed is within the walking speed range (for *extended* places). Thus, we will use direction (bearing) and speed (and acceleration) to define the parts of a user's trajectory that describe significant places. To capture the random movements at point-like and extended places, we will use the standard deviation of the bearing change to describe the relative bearing change. If the bearing change is large, the standard deviation will also be large, i.e., significantly larger than zero. We will denote the set of bearing changes that can be computed from a set of bearings B as $\Delta B = \{\Delta b_{12}, \Delta b_{23}, \ldots, \Delta b_{(n-1)n}\}$.

There are other parts of the trajectory that are important for significant places, in particular those trips that lead to or start from a significant place. These parts of a trajectory can be similar to the segments representing a *point-like* place in terms of movement speed and acceleration behaviour, i.e., the user is walking very slowly. However, as the user is walking along a given direction, e.g., towards a significant place, the bearing change will be small and the standard deviation will also be close to zero. Similarly, there are other parts of the trajectory where the user slowly moves on a road or path, which can be similar to a segment representing an *extended* place in terms of speed and acceleration. However, as the user is moving along a road or path, the bearing change will again be small (except for the cases where the road or path itself alters or the user changes direction) and the standard deviation will also be close to zero. We call these parts as directed movements.

We next define a GPS trajectory as all other definitions are based on it. We then define different segments of a trajectory representing significant places or directed movements.

2.1 Definitions

GPS Trajectory: A GPS Trajectory T is a time-stamped consecutive sequence of GPS data points $p_i \in P, P = \{p_1, p_2, ..., p_n\}$ such that $\forall i \in [1,n]$, $p_i = (x_i, y_i, t_i)$ and $t_i < t_{i+1}$, where x_i, y_i and t_i represent latitude, longitude and time-stamp, respectively.

Undirected Static Segment: Let T be a user's trajectory of n points and $B = \{b_1, b_2, \ldots, b_n\}$ its bearing values. A partial GPS trajectory $T' \subset T$ is called an *undirected static segment* if the speed (velocity) v and acceleration a is ~ 0 for all data points of T' and the standard deviation of the bearing change $\Delta B'$, where $B' \subset B$, is $\sigma(\Delta B') \gg 0$.

An undirected static segment represents a *point-like* place such as home, office, bus-stop, or intersection. In practice, values of 0 for v and a are typically not possible either because a user is moving a little at a bus-stop or due to the errors in GPS measurements and should be replaced by upper bounds, i.e., $v \leq v_{\min}$ and $|a| \leq a_{\min}$.

Undirected Slow-motion Segment: Let T be a user's trajectory of n points and $B = \{b_1, b_2, \ldots, b_n\}$ its bearing values. A partial GPS trajectory $T' \subset T$ is called an *undirected slow-motion segment* if the speed satisfies $v > v_{\min}$ for most data points and $v \leq v_{\max}$ for all data points of T' and the standard deviation of the bearing change $\Delta B'$, where $B' \subset B$, is $\sigma(\Delta B') \gg 0$.

An undirected slow-motion segment represents *extended* places such as a market or a park. The definition has no constraint for acceleration. We found that in practice the speed alone is sufficient. Extended places typically involve the user to stop for some time (e.g., while doing shopping in a market) and to mostly move otherwise (e.g., while moving between shops) with an upper bound for the speed ($v_{\max}$).

Directed Static Segment: Let T be a user's trajectory of n points and $B = \{b_1, b_2, \ldots, b_n\}$ its bearing values. A partial GPS trajectory $T' \subset T$ is called an *directed static segment* if the speed (velocity) v and acceleration a is ~ 0 for all data points of T' and the standard deviation of the bearing change $\Delta B'$, where $B' \subset B$, is $\sigma(\Delta B') \sim 0$.

Directed slow-motion segment: Let T be a user's trajectory of n points and $B = \{b_1, b_2, \ldots, b_n\}$ its bearing values. A partial GPS trajectory $T' \subset T$ is called an *directed slow-motion segment* if the speed satisfies $v > v_{\min}$ for most data points and $v \leq v_{\max}$ for all data points of T' and the standard deviation of the bearing change $\Delta B'$, where $B' \subset B$, is $\sigma(\Delta B') \sim 0$.

2.2 Problem Statement

Given a user's GPS Trajectory T of n consecutive GPS points and a time-duration t_{sig}, we need to find all significant places, i.e., all undirected static and slow-motion segments $T' \subset T$ such that $(t_r - t_q) \geq t_{sig}$, where t_q and t_r are the first and last time-stamp of T', respectively.

As in [3], we empirically define the speed and acceleration range in Table 1 for directed and undirected static segments. Since the average walking speed of a human being varies around 4-5km/hr [4], we set the speed range of directed and undirected slow-motion segments as shown in Table 1.

Table 1: Speed/acceleration for static and slow-motion segments

Segment	Speed	Acceleration
Static segment	$0-3.6km/hr$	$\pm 1km/hr^2$
Slow-motion segment	$0-7.5km/hr$	"N/A"

2.3 Algorithm

As mentioned earlier, we propose a two stage algorithm, named as ExtractSigPlaces, to extract significant places from GPS trajectory. In the first stage, we extract all static and slow motion segments (both directed and undirected) using a speed and acceleration based bucketing technique. In the second stage, we use a bearing based technique to identify the actual significant places. The algorithm takes two parameters as inputs: a raw GPS trajectory T and a time duration t_{sig}, produces a set of significant places L_{sig} as output. Note that the value of t_{sig} is application-specific and needs to be supplied accordingly. Figure 1 shows the architecture of our approach. For each point in the trajectory, we compute the speed, acceleration, bearing and bearing change as in [7].

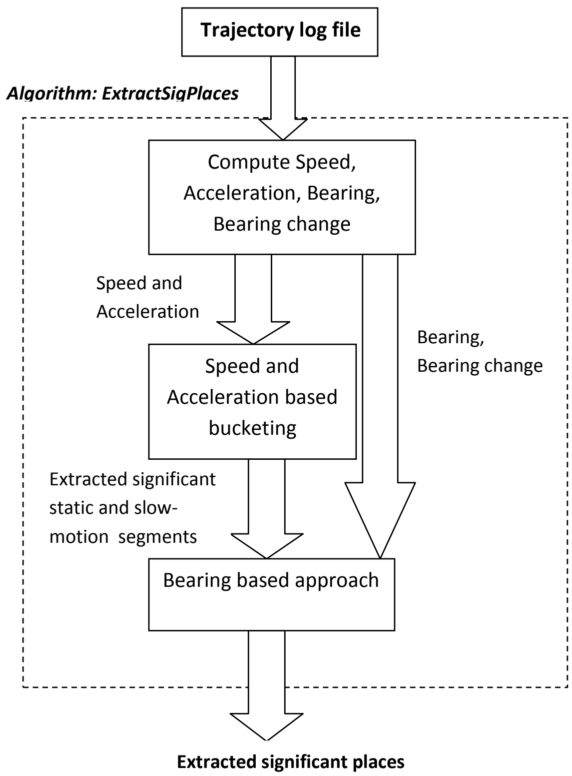

Figure 1: System Architecture

Speed and acceleration based bucketing: In our speed and acceleration based bucketing technique, we first extract all significant directed and undirected static segments of the trajectory, where the speed and acceleration of consecutive GPS points of those segments remain within the static range (as in Table 1) and the duration of stay is $>= t_{sig}$. Note that there may be no data between two consecutive GPS points in the trajectory for some time due to the loss of the GPS signal or user switching off the mobile phone: if both speed and acceleration at those two points fulfil the criteria for a static segment, we can also retrieve those points as a probable *point-like* location. This technique will not retrieve locations due to the signal loss when the mobile user is traveling inside a tunnel, since some distance will be covered in this case.

We next extract all significant directed and undirected slow-motion segments of the GPS trajectory where the speed and acceleration of consecutive GPS points remain within the slow-motion speed range (as in Table 1) and the duration of the segment is $>= t_{sig}$.

Bearing based technique: In the previous stage, we extract all the significant static and slow-motion segments representing either significant places (undirected static or slow-motion segments) or directed movements. To retrieve the significant places, we use a bearing based technique. We compute the standard deviation (sd) of bearing change for all the extracted segments. If the sd value of a segment is greater than S_{th} (for static segments) or D_{th} (for slow-motion segments), we consider that segment representing a significant *point-like* or *extended* place respectively. We obtain the centroid of the GPS points for each segment as in [9] and plot them using Google Map to identify the locations.

3. EXPERIMENTAL EVALUATION

We conducted a proof of concept experimental evaluation of our approach by gathering our own real world GPS data. For data collection, we used two SAMSUNG Galaxy2 GPS enabled Android mobile phones. Data was logged for two users in Melbourne, Australia, for a selection of times during a three month period. The data from each user was then (human) annotated, to serve as a ground truth for validation. Annotations recorded locations of stays, transport modes and significant events during a trip.

We present our findings for one of the trajectory data files that contained 259 data points, logged at 10 second intervals. The summary of the ground truth is shown in Table 2. The user was engaged in a variety of activities, centred around the task of taking a child to kindergarten. We provide comparisons of our algorithm's output to the ground truth, and to a well-known baseline technique [9].

Table 2: Ground Truth - User's Annotations

Time Duration	Annotations
08:40:51 - 08:50:11	car
08:50:24 - 08:53:37	walk
08:53:52 - 08:57:44	kindergarten
08:58:23 - 09:01:03	walking
09:01:16 - 09:08:50	car
09:09:05 - 11:37:26	home (GPS turned off)
11:37:26 - 11:43:51	car
11:44:06 - 11:47:48	walk
11:48:04 - 11:51:57	kindergarten
11:52:15 - 12:00:25	walk
12:00:41 - 12:07:42	car

Extracting significant places: Using a significance threshold of $t_{sig} = 1$ minute, we first processed the data using our speed and acceleration based bucketing method, which yielded 11 significant static segments, but no significant slow-motion segment.

We processed the data further using our bearing based technique. In particular, we checked if the standard deviation (sd) of bearing change for each of the extracted static segments exceeded the static threshold limit S_{th}. In order to select a value for S_{th}, we employed a simple machine learning approach. A completely separate set of data was used as a training set, containing GPS traces collected when i) the user was walking slowly to/from kindergarten (4 segments) and ii) when the user was inside the kindergarten (8 segments). Using this training data, all candidate values for S_{th} were evaluated and the value which yielded the best accuracy for distinguishing between slow walking behavior and being inside behaviour was chosen. The most predictive value is $S_{th} = 108.12$ (with an accuracy of 91.67%).

Table 3: Extracted significant *point-like* places

Extracted Places	Duration	Arrival Time
car park (kindergarten)	1min 16sec	08:50:11
kindergarten	3min 2sec	08:55:21
home	149min 19sec	09:08:35
waiting in traffic	1min 15sec	11:38:19
car-pickup (kindergarten)	3min 11sec	11:58:32

The significant places that were output are shown in Table 3. A semantic label for each location was obtained by plotting it via Google Maps and showing it to the user and requesting a short description. Automatic generation of semantic labels is an interesting issue, but outside the focus of this paper, which concentrates

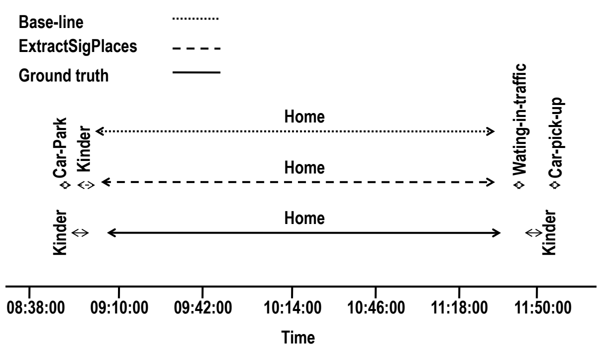

Figure 2: Results against ground truth and baseline

on the necessary first step: identifying the co-ordinates and relevant window of time for a significant place. Comparing Table 3 with the ground truth (Table 2), we see that our algorithm identifies all but one significant place. Interestingly, it can also identify some finer grained additional places, such as car-parking, car-pick-up and waiting at an intersection. The significant place that was not identified, is the second kindergarten place where a user picks up their child. In this case, the standard deviation of bearing change was below the S_{th} threshold value. The additional identified places, included two car park places – at which car was parked when dropping off/picking up the child at kindergarten. In both cases the user was with his child and the bearing change between consecutive GPS points was scattered, a behaviour recognised by our algorithm as being indicative of a significant place.

Figure 2 provides an overlay comparison of our algorithm's output against ground truth. We compare in Figure 2 our algorithm against the baseline technique in [9], using a distance threshold value of 200 metres and time threshold value of 20 minutes to identify stay locations. We use the same parameter settings. Figure 2 shows that the output of our algorithm is close to ground truth. The baseline approach, however, could only identify the "home" place and missed the other places. We also investigated the baseline approach for a time threshold of 5 minutes and distance threshold of 200 metres. In this case, the baseline approach could identify the kindergarten places correctly, but could not identify waiting in traffic, car-parked and car-pickup places. A further reduction of the time threshold to 1 minute and distance threshold to 100 metres for the baseline approach produced an explosion of places: 32 different locations, among which only "home" was identified correctly. This highlights the inherent difficulty of trying to identify significant places only using time and distances thresholds.

Distribution analysis using the KS test: We further analyzed the significance of our results using a two-sample Kolmogorov-Smirnov (KS) Test [5]. We compared the bearing change distribution for the places of events identified by our algorithm (except for "home", since there was only one data point) versus the bearing change distribution for the training kindergarten data and training slow-walk data. Table 4 provides the KS test p-values. The null hypothesis for the KS-test is that the two bearing change distributions are the same. The table shows that the p values output when comparing the distribution for the significant places output by our algorithm against the kindergarten data distribution are higher, indicating greater consistency with the null hypothesis. Conversely, when comparing against the slow-walk data distribution, the p-values are lower. This provides further evidence that the places output by our algorithm are more consistent with a kindergarten theme, as opposed to a slow walking theme.

Table 4: KS Test Result comparing bearing change distribution for significant *point-like* places against bearing change distribution for kindergarten and for slow-walk. Higher p-values indicate that the two distributions have greater similarity

KS Test	P-value
car park(kindergarten) vs. kindergarten	0.796
car park(kindergarten)vs. slow-walk	0.213
kindergarten vs. kindergarten	0.465
kindergarten vs. slow-walk	0.089
waiting in traffic vs. kindergarten	0.7
waiting in traffic vs. slow-walk	0.153
car-pickup(kindergarten) vs. kindergarten	0.996
car-pickup(kindergarten) vs. slow-walk	0.022

4. CONCLUSIONS

We have developed an algorithm based on a mobile user's speed, acceleration, bearing and bearing change, to extract significant places from a GPS trajectory. To our knowledge, we are the first to consider bearing change distribution to extract significant places from a trajectory, while considering the unreliability of GPS measurements. The results are encouraging and indicate that the use of bearing change offers considerable advantages over a traditional time and distance thresholding approach.

5. ACKNOWLEDGEMENTS

This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP110100757).

6. REFERENCES

[1] L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans, and A. Vaisman. A model for enriching trajectories with semantic geographical information. In *Proc. of ACM GIS*, 2007.

[2] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares. A clustering-based approach for discovering interesting places in trajectories. In *Proc. of ACM SAC*, pages 863–868, 2008.

[3] K. Rehrl, S. Leitinger, S. Krampe, and R. Stumptner. An approach to semantic processing of gps traces. In *Proc. of the MPA*, pages 136–142, Zurich, 2010.

[4] I. Road Engineering Journal, TranSafety. Study compares older and younger pedestrian walking speeds. `http://www.usroads.com/journals/p/rej/9710/re971001.htm`. Accessed: 28/06/2012.

[5] D. Sheskin. *Handbook of Parametric and Nonparametric Statistical Procedures*. Champa and Hall/CRC, third edition, 2004.

[6] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *Proc. of WWW*, pages 1029–1038, 2010.

[7] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proc. of UbiComp*, pages 312–321, 2008.

[8] Y. Zheng, L. Zhang*, Z. Ma, X. Xie, and W.-Y. Ma. Recommending friends and locations based on individual location history. *In ACM TWEB*, 5(1), 2011.

[9] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proc. of WWW*, pages 791–800, 2009.