Check for updates

IET Intelligent Transport Systems

Research Article

Identifying activities and trips with GPS data

ISSN 1751-956X
Received on 11th December 2017
Revised 8th April 2018
Accepted on 8th May 2018
E-First on 19th June 2018
doi: 10.1049/iet-its.2017.0405
www.ietdl.org

Zong Fang[1], Lv Jian-yu[1], Tang Jin-jun[2] ✉, Wang Xiao[1], Gao Fei[1]

[1]Department of Transportation Planning and Management, College of Transportation, Jilin University, College of Transportation, Changchun, Jilin, People's Republic of China

[2]Key Laboratory of Smart Transport, School of Traffic and Transportation Engineering, Central South University, Hunan Province, Changsha, Hunan, People's Republic of China

✉ *E-mail: jinjuntang@csu.edu.cn*

Abstract: This study designs a process for identifying trips and activities based on global positioning system (GPS) survey data. The proposed identification process is composed of four steps, namely determining status segments, detecting activities, identifying trips, and recognising short-time activities. The results indicate that the proposed algorithm shows a high level of identification accuracy compared with the travel diaries reported in the paper-form travel survey. By providing the identification method of the short-time activities, this study resolves the problem of overlooking short-time activities in conventional travel surveys and increases the accuracy of trip detection. This work also facilitates the study of the spatial and temporal distributions of short-time activities related to travel behaviours such as temporary parking. By proposing a method for identifying trips and activities from GPS data, the findings provide a research scheme for detecting other travel information based on GPS data such as travel mode and trip purpose, reasonable decisions for urban transportation planning and management.

1 Introduction

By analysing resident travel and activity data, urban travel behaviour and patterns can be explored, which is the basic goal of transportation planning and management. Most travel data were collected via large-scale resident travel surveys such as the first–fifth resident travel survey conducted in Beijing City in 1986, 2000, 2005, 2010, and 2014, respectively. Statistics indicate that trip frequency and trip intensity are increasing with the development of society and economy. To understand the dynamic change of travel behaviour, high-frequency and large-scale travel surveys are needed. However, it is infeasible to conduct high-frequency surveys by using the conventional travel survey method. Moreover, conventional methods such as mail or phone surveys that collect travel information based on participant recall, require respondents to fill in many pages of travel diaries. Quite often, respondents overlook short trips and short-time activities, report trips out of sequence, or approximate the departure and arrival times [1, 2].

Adopting mass global positioning system (GPS) trajectory data to explore the travel patterns of urban residents can ease a participant's burden and enhance the efficiency and accuracy of travel survey. This ensures that the acquisition of high-frequency and large-scale travel data is possible [3–6]. Nevertheless, the raw GPS trajectory data cannot be applied to travel data analysis directly. A mathematical model and systematic software are needed to extract travel information from GPS trajectory data. This paper will introduce an identification process that can extract activity and trip data based on GPS survey data. Compared to previous studies, this method tends to enhance the identification accuracy by designing an algorithm for short-time activities identification specifically and collecting a larger sample size of GPS records.

The remainder of this paper is organised as follows. In Section 2, we review the relevant literature on recognising trips and activities based on GPS data in general. Section 3 provides a description of the data and the study area. In Section 4, the process of recognising trips and activities from GPS data is characterised and designed. This paper concludes with Section 5, where we summarise our findings, discuss our study limitations and identify directions for future research.

2 Literature review

Being a crucial step in GPS-based travel behaviour, trip identification and activity recognition have been investigated by many researchers. Some of the leading methods that previous researchers have developed or investigated include methods identifying activities based on the detection of stationary points, points omitted by GPS devices, heading changes in road networks, and the spatial distribution of stationary points.

According to the detection method based on stationary points, studies took a group of consecutive stationary points, with duration times greater than a pre-determined dwell-time threshold, as an activity. A stationary point is defined as a point with speeds less than the walking speed. Du and Aultman-hall [7] used 45, 120, 140, and 180 s as the dwell-time thresholds and defined 0.51 m/s as the maximum speed limit of the stationary points. Qiu [8] detected the stationary point by considering the average speed of a set of consecutive points instead of a single point.

As we know, satellite signals are lost in some circumstances such as a signal blockage. These kinds of data gaps are more likely to happen near urban canyons in central business district (CBD) areas. Since activities may likely occur during the data gaps, the data gaps can be used to recognise activities. If the signal loss is detected between adjacent points, i and $i+1$, the average speed between point i and point $i+1$ can be calculated. This is then compared with the average speed of the points before point i as well as the points after point $i+1$. These deviations are then applied to determine whether there is an activity during the period of signal loss. Du and Aultman-hall [7] compared the average speed of the data gap with that of 20 points before and after the signal losses, respectively.

Some short-time activities are also characterised by heading changes in road networks. For example, activities such as dropping off or picking up goods are characterised by a 180° heading change in the GPS records. Therefore, the extent of the heading change can be used to identify activities. For example, Du and Aultman-hall [7] employed a heading change to identify short-time activities. Forrest and Pearson [9] set different thresholds for the dwell time according to different extents of the heading changes. Previous studies also identified activities by calculating the distance that

© The Institution of Engineering and Technology 2018

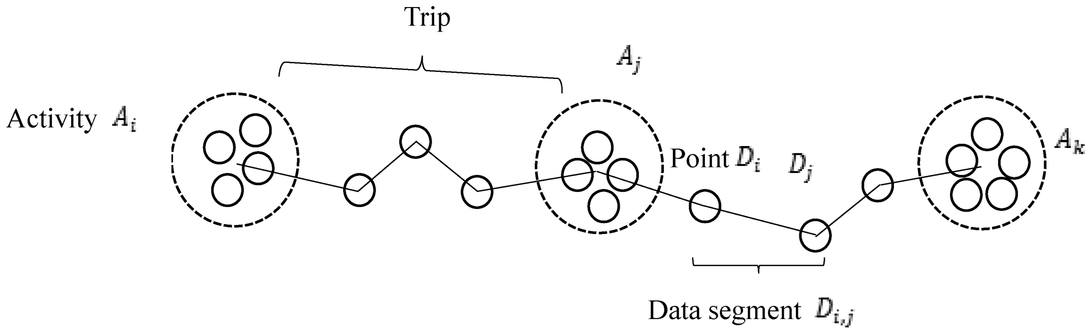

Fig. 1 *Example of a daily trip and activity chain*

GPS points deviates from a road network. For example, Du and Aultman-hall [7] denoted the GPS traces deviating from road networks, for a distance of 15–30 m, as missing points, and identified whether the data gap, composed of the missing points, was the dwell time for activities or not by comparing the dwell time to the time threshold of activity duration.

Another previous method is proposed to recognise activities based on the spatial distribution of stationary points. The reason is that when an activity occurs, several stationary points will demonstrate an aggregated distribution to some extent [10, 11]. Zhou *et al.* [12] proposed a different density and join-based clustering (DJ-cluster) algorithm to investigate clustering stationary points as potential activity points.

Although previous studies proposed many methods for trip and activity identification, much of the attention was paid to the identification of long-time activities. The studies such as the study by Du and Aultman-hall [7] that investigate short-time activities by examining heading changes are more applicable to a car trip but may not be appropriate for a pedestrian trip. However, identification of short-time activities is also important to travel behaviour detection. One of the reasons is that most of the short-time activities are coming up with temporary parking, which is the key point of traffic congestion diagnosis and management. Moreover, some of the greatest advantages of GPS-based travel surveys are that the methods can detect short-time activities in a more feasible and accurate way [13, 14]. Therefore, developing a high-accuracy programme for short-time activities detection is one of the crucial missions of this paper. In addition, this paper will try to improve the identification accuracy by means of employing GPS survey data with a larger sample size compared with most of the previous studies of which the sample sizes were <200. Moreover, except for trip and activity, many works studied on detection of other travel features such as travel mode and trip purpose. For example, Zong *et al.* [15] developed a hybrid procedure for travel mode identification based on GPS travel data and geographic information system information of transit network. It improved the detection accuracy of subway mode.

3 Data and the study area

The GPS-based travel survey data collected in Beijing City in 2015 is used in this paper. The data sample consists of ∼950 respondents and ∼1,040,000 GPS records. Totally, 1949 trips and 1835 activities are contained in the data sample, with ∼293 short-time activities. During the survey period from 10th August 2015 to 15th September 2015, 950 respondents also reported their socio-demographic information and travel diaries by filling in paper forms.

Each record in the dataset represents a GPS signal that was captured consecutively at 5 s intervals by the GPS device (i-gotU GT-600). The main items in each record include date, time, latitude, longitude, altitude (m), distance (m), speed (m/h), course (deg), estimated horizontal position error (EHPE, cm), and satellite ID.

By the end of 2015, the Beijing metropolitan area covered 16,412 km^2 and had a permanent registered population of 21.7 million. The residential travel survey conducted in Beijing in 2014 indicates that the total number of trips per day is 44.5 million. Compared to 2010, the number of trips grew by 14.7%, with an average annual growth rate of 3.5%. The average daily trip frequency per person is 2.75, which is 2.5% less than that in 2010. About 8.05% of the activities are short-time activities.

After transferring the geodesic coordinates into geographical coordinates, and changing the date and time from universal time coordinated to local time for each GPS record, we created the following rules which were applied to identify records to retain in the dataset [16–18].

(i) Satellite ID ≥ 3

When a GPS device receives signals from at least three satellites, the position of the device can be determined with an accuracy of within ∼10 m. Therefore, all records with fewer than three satellites were dropped.

(ii) EHPE ≤ 100 m

EHPE refers to the level positioning accuracy of a GPS point. The lower the value is, the greater the positioning accuracy is. According to the positioning accuracy of the GPS device, 100 m is set as the threshold of EHPE for data filtering.

(iii) Speed ≤ 200 km/h

Speed refers to a traveller's instantaneous velocity. A possible maximum value of instantaneous velocity is set to be the maximal threshold of speed [19, 20].

(iv) According to the geographical position and regional extent of Beijing city, the range of altitude, longitude, and latitude is set as, 0 m ≤ altitude ≤ 2350 m, 115.25° ≤ longitude ≤ 117.37°, and 39.26° ≤ latitude ≤ 41.03° [21].

4 Identifying trips and activities

To take part in an activity, a traveller should move from one place to another. This moving process is denoted as a trip. At the end of a trip, the traveller will consume a certain dwell time to conduct an activity. This is then defined as an activity. Trip and activity recognition is a decomposition process that divides the GPS traces into segments of trips or activities. In this paper, we define a group of stationary points at the end of a trip as an activity and denote it A_i. Then, the points between adjacent trip activities are shown in Fig. 1. If activities can be identified, trips can be extracted easily. Therefore, the first step of trip and activity identification is detecting the activities. In this paper, we design a trip and activity identification process, which is composed of three steps, i.e. dividing the status segments, identifying activities, and recognising trips. Each step will be introduced in detail in the following sections.

4.1 Dividing status segments

In this step, all the GPS traces will be divided into status segments. Each is composed of four sub-steps: converting GPS traces into data segments, identifying the status of the data segments, merging adjacent data segments with the same status and generating status segments, and adjusting the status of status segments.

4.1.1 Converting GPS traces into data segments: The traces that the GPS devices recorded are a sequence of points, which are represented in chronological order, as shown in Table 1. Each point

© The Institution of Engineering and Technology 2018

17519578, 2018, 8, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2017.0405 by INASP/HINARI - ARMENIA, Wiley Online Library on [17/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

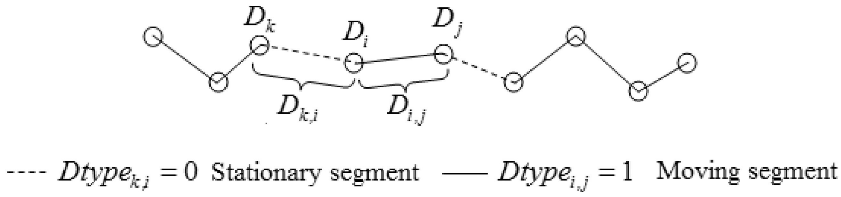

Fig. 2 *Status of data segments*

Table 1 Sample GPS records in point form

Number of travellers	Number of GPS records	Date	Time	Distance, m
—	—	—	—	—
10,011	9	17 September 2015	9:14:50	16.0206
10,011	10	17 September 2015	9:14:55	18.3438
10,011	11	17 September 2015	9:15:00	52.2385
10,011	12	17 September 2015	9:15:05	16.1318
—	—	—	—	—

Table 2 Samples records in the data-segment form

Number of travellers	Number of points D_i	Number of points D_j	Duration, s	Distance, m
—	—	—	—	—
10,011	9	10	5	2.5206
10,011	10	11	5	2.3438
10,011	11	12	5	52.2384
10,011	12	13	5	16.1317
—	—	—	—	—

Table 3 Samples of records in the status-segment form

Number of points D_m	Number of points D_n	$t_{m,n}$, s	$d_{m,n}$, m	$K\text{type}_{m,n}$
—	—	—	—	—
9	11	10	4.8644	0
11	13	10	68.3703	1
—	—	—	—	—

indicates an instantaneous position. If two adjacent points join, this in turn can convert the data from point form to segment form. Consequently, each data segment shows the status from one point to the next one. Each data segment is associated with several attributes such as distance, duration, and average speed. An arbitrary data point is defined as D_i, the data segment consists of adjacent GPS traces D_i and D_j, which is denoted as $D_{i,j}$ (Fig. 2). Then, the attributes of $D_{i,j}$ are calculated as follows: distance, i.e. denoted by the distance between D_i and D_j; durations, which is the time gap between D_i and D_j; and average speed, which is the ratio of distance to duration. Some samples of records in point and segment form are shown in Tables 1 and 2, respectively.

4.1.2 Identifying status of data segments: The status (Dtype) of each data segment is defined as stationary versus moving, which is determined by comparing the average speed of the data segment to the speed threshold. According to Wang and Li, the range of walking speed is 0.83–1.67 m/s [16]. We define the speed threshold (V_{thresh}) as 0.6 m/s, which denotes the minimum speed of a moving segment. The average speed of each data segment can be calculated as follows:

$$v_{i,j} = d_{i,j}/t_{i,j} \tag{1}$$

where $d_{i,j}$ is the distance between two adjacent points D_i and D_j and $t_{i,j}$ is the duration between them.

If $v_{i,j}$ is greater than V_{thresh}, $D_{i,j}$ is identified as a moving segment, and its status $D\text{type}_{i,j}$ is recorded as 1. Otherwise, $D_{i,j}$ is identified as a stationary segment and its $D\text{type}_{i,j}$ is recorded as 0

$$D\text{type}_{i,j} = \begin{cases} 1 & v_{i,j} > V_{\text{thresh}} \\ 0 & v_{i,j} \le V_{\text{thresh}} \end{cases} \tag{2}$$

$$d_{i,j} = \sum_{k=i,\, i+1,\, \dots,\, j-1} d_{k,k+1} \tag{3}$$

where $d_{k,k+1}$ is the distance between two adjacent points D_k and D_{k+1}.

4.1.3 Merging adjacent data segments with the same status: After identifying the status of each data segment, e.g. stationary or moving, the adjacent data segments may show the same status. Thus, we need to merge the adjacent data segments with the same status into one data segment. Accordingly, the statuses ($D\text{type}_{i-1,j-1}$ and $D\text{type}_{i,j}$) of arbitrary two adjacent data segments ($D_{i-1,j-1}$ and $D_{i,j}$) are compared. If $D\text{type}_{i-1,j-1}$ equals to $D\text{type}_{i,j}$, we merge $D_{i-1,j-1}$ and $D_{i,j}$ into $D_{i-1,j}$. Then, all the adjacent data segments with the same status are merged, until the status of adjacent data segments is distinct from one another. Thus, the stationary status and the moving status alternatively appear. After conducting this merging sub-step, we call an arbitrary data segment ($D_{m,n}$) as a status segment $K_{m,n}$ (m and n are the marks of two endpoints of the data segment). We also define a parameter $K\text{Type}_{m,n}$ to denote the status of $K_{m,n}$. Similarly, $K\text{Type}_{m,n} = 1$ or $K\text{Type}_{m,n} = 0$ denote that the $K_{m,n}$ is in a moving or stationary status. The duration ($t_{m,n}$) and distance ($d_{m,n}$) of the status segment $K_{m,n}$ can be calculated by

$$t_{m,n} = T_n - T_m \tag{4}$$

$$d_{m,n} = \sum_{k=m,\, m+1,\, \dots,\, n-1} d_{k,k+1} \tag{5}$$

where T_n and T_m are the recording times of data points D_n and D_m, respectively. Some samples of status segments are shown in Table 3.

4.1.4 Adjusting the status of status segments: On the basis of the definition of activity and trip, only the stationary or moving segments, with a certain duration or distance can be regarded as an activity or a trip. We define the time threshold (T_{thresh}) as 30 s and the distance threshold (D_{thresh}) as 300 m [22–24]. Then, we compare the duration ($t_{m,n}$) of the stationary segment with T_{thresh} as well as the distance ($d_{m,n}$) of moving segment with D_{thresh}. If the $t_{m,n}$ or $d_{m,n}$ of an arbitrary segment $K_{m,n}$ is less than T_{thresh} or D_{thresh}, then the status of this segment is converted into the opposite one. The adjusting process is shown in Fig. 3.

After completing this adjustment process, the status of adjacent status segments will be the same again. Therefore, we need to merge adjacent data segments with the same status as we do in sub-step 3, until we get alternating stationary and moving status segments. According to the four sub-steps, the data processing stage of the process is completed.

© The Institution of Engineering and Technology 2018

17519578, 2018, 8, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2017.0405 by INASP/HINARI - ARMENIA, Wiley Online Library on [17/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

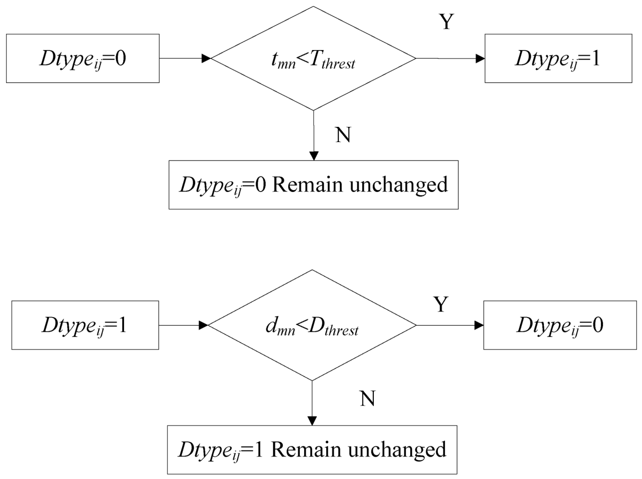

Fig. 3 *Adjusting status of status segments*

4.2 Identifying activities

A threshold time S_{thresh} is defined as the minimum dwell time of an activity. The stationary segment, whose duration is greater than S_{thresh}, is recognised as an activity (A_i). Zmud and Wolf [25] proposed a S_{thresh} of 2 min to identify activities. In determining the value of S_{thresh}, we should distinguish between an activity from a short-time stop during a trip due to traffic congestion, delays at an intersection and transfer time between travel modes. According to the travel data collected in Beijing in 2014, the minimum duration of activities, except for short-time stops due to traffic congestions or delays at intersections and mode transfer, is 7 min. Therefore, S_{thresh} is set to be 7 min in this paper. Accordingly, a stationary segment $K_{m,n}$, whose duration is >7 min, is identified as an activity.

4.3 Identifying trips

After identifying the activities, the points between the two adjacent activities form trips. Thus, the preliminary identification of trips and activities is completed. The recognition process mentioned above can be expressed as Fig. 4. Taking respondent #15342 as an example, the identification results are shown in Fig. 5.

4.4 Identification results of trips and activities

By applying the identification process presented above, we obtained the trips and activities of 950 respondents. Identification results are then compared with real travel records that respondents completed, in survey form, to examine the identification accuracy. With respect to the number and duration of activities, the identification accuracy is calculated by using the following equations.

4.4.1 Accuracy concerning the number of activities:

(i) The precision rate (P_1) refers to the number of activities which really occurred and were recognised as the number of activities identified. P_1 is formulated in the equation below:

$$P_1 = \frac{\text{NOI}}{\text{NI}} \times 100\% \tag{6}$$

where NOI denotes the number of activities which really occurred and were identified, while NI is the number of activities identified [26].
(ii) The recall rate (P_2), is the ratio of NOI to the number of activities really occurred

$$P_2 = \frac{\text{NOI}}{\text{NO}} \times 100\% \tag{7}$$

where NO is the number of activities really occurred, and NOI = NI ∩NO.

4.4.2 Accuracy concerning the duration of activities:

(i) The relative error of activity duration (δ) can be calculated as follows:

$$\delta = \left| \frac{t' - t}{t} \right| \times 100\% \tag{8}$$

where t' is the duration of an identified activity and t is the duration of a real activity.
(ii) The average differences of starting times ($\bar{x}$) refer to the average value of the differences between the starting times of the identified activities and the real activities. $\bar{x}$ can be calculated by using the following equation:

$$\bar{x} = \frac{|st'_1 - st_1| + |st'_2 - st_2| + \cdots + |st'_i - st_i| + \cdots + |st'_n - st_n|}{n} \tag{9}$$
$$i = 1, 2, \ldots, n$$

where st_i' is the starting time of identified activity i, whereas st_i is the starting time of real activity i, and n indicates the total number of real activities for which there is a corresponding assessed activity.

By using (6)–(9), the accuracy of the identification results is calculated. The results show that P_1 and P_2 are 75.55 and 90.04%, respectively, and $\bar{x}$ is 5 min. The maximum, minimum and average value of δ are 20.71, 0.00, and 5.4%, respectively. This indicates that, though the accuracy of the identification results is acceptable, short-time demonstrates a big promotion probability. Nevertheless, travel data collected in Beijing in 2014 indicates that 2.66% of the total number of activities is an under-20 min activity and under-7 min activities account for ∼16.34% of all the intermediate stops conducted during commute trips. We then conducted additional accuracy determinations for activities over-20 min and under-20 min. The results indicate that the P_1 and P_2 of the over-20 min activities are 85.48 and 93.62%, whereas they are 65.62 and 86.45% for the under-20 min activities, respectively. This reveals a relatively unsatisfactory performance of short-time activity identification. Therefore, further investigations into the under-20 min activities are conducted.

4.5 Identifying short-time activities

Survey results show that, in addition to the major activities, many travellers conduct short-time activities during a trip such as a stop for breakfast on the way to the work place. Moreover, the survey results propose that many respondents forget or overlook these short-time activities when filling in the survey questionnaires. This issue can be fixed by identifying short-time activities by using GPS trajectory data, as GPS data have the merit of recording continuous traces that the traveller covered.

On the basis of the identification results of activities and trips, activities that lasted for more than 20 min were extracted from the activity data. Then, if the algorithm finds an activity between two adjacent over-20 min activities (A_i and A_j), whose duration is shorter than 20 min, we will identify the trips and activities between A_i and A_j once again. Otherwise, the characterisation of two adjacent, over-20 min activities will be final.

Then, the duration threshold S_{thresh} of activities, which have been used in Section 3, is shortened to a certain value for identifying short-time activities. The travel survey data in Beijing City in 2014 indicates that the duration of most of the short-time activities is between 4 and 20 min. The duration of some short-time activities such as picking someone up or buying breakfast is between 4 and 7 min. Therefore, taking 7 min as S_{thresh} (the threshold duration of an activity) means that we will omit some short-time activities with durations between 4 and 7 min. To

© The Institution of Engineering and Technology 2018

17519578, 2018, 8, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2017.0405 by INASP/HINARI - ARMENIA, Wiley Online Library on [17/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

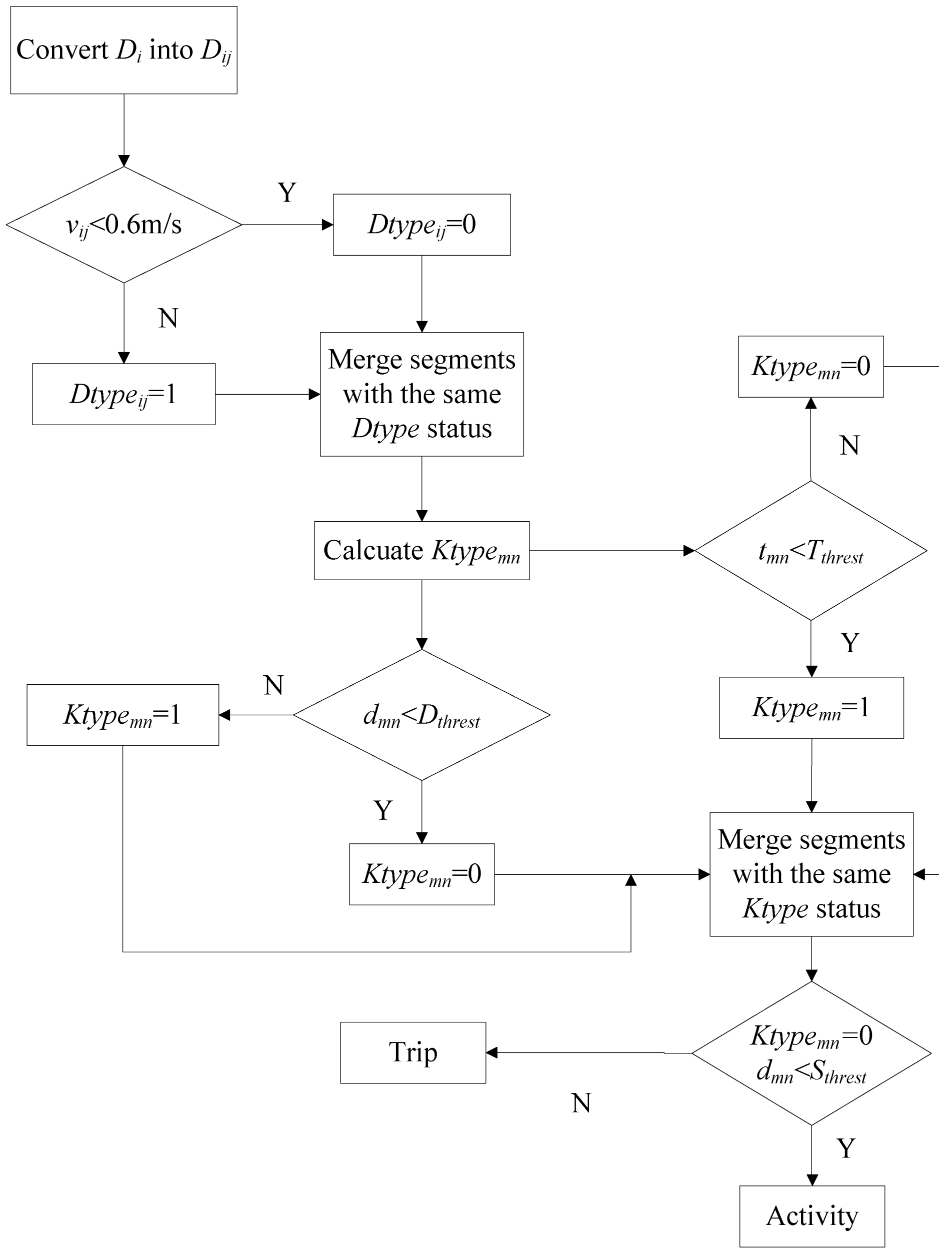

Fig. 4 *Procedure of trip and activity identification*

identify these short-time activities, the duration threshold of short-time activities is set to 4 min. In other words, the stationary segment, whose duration is >4 min, is defined as a short-time activity. Then, the segment between two newly identified short-time activities or between a newly identified short-time activity and an originally identified activity are identified as a trip. The identification process of activities and trips are then completed after this final step.

The identification results of short-time activities indicate that with the introduction of the short-time activity identification step, P_1 of the short-time activities increases from 65.62 to 85.68%, and P_2 increases from 86.45 to 92.13%. This improvement shows the effectiveness of the identification algorithm for short-time activities. Identification accuracy is likely more positive, as the 'real' trip and activity records are applied to the accuracy calculation, which was obtained from the traditional paper-form travel survey, where respondents may have overlooked short trips and short-time activities.

4.6 Overall identification results

By using the overall identification process presented above, all the trips and activities for each respondent in the GPS data are identified. The daily trip and activity chain identified for respondent #13841 is shown in Fig. 6.

The distribution of activity duration based on real survey data and identification results are shown in Figs. 7*a* and *b*, respectively. This indicates that, concerning activity duration, the recognition results resemble the survey data. Moreover, compared to the travel records that respondents filled in, we recognised more under-20 min activities by using our identification algorithm. This indicates that the identification process proposed in this paper can help us to obtain accurate travel information by recognising many short-time activities or trips that survey respondents overlooked due to the burden resulting from the survey length. This is a significant improvement not only for identifying activities and trips from GPS survey data but also for obtaining travel information from the traditional paper-form travel survey.

By comparing to the previous studies (such as those by Du and Aultman-hall [7]) that employed heading changes and used the distance from a road network as parameters for activity identification, this paper utilised the dwell time as the single detection parameter. This ensured the method could be applied to not only car trips but also to trips by other modes. The method also made the geographical environment information and process of map matching unnecessary. In addition, this paper obtained the identification accuracy of 92.13% (the recall rate P_2), which is

© The Institution of Engineering and Technology 2018

17519578, 2018, 8, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2017.0405 by INASP/HINARI - ARMENIA, Wiley Online Library on [17/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

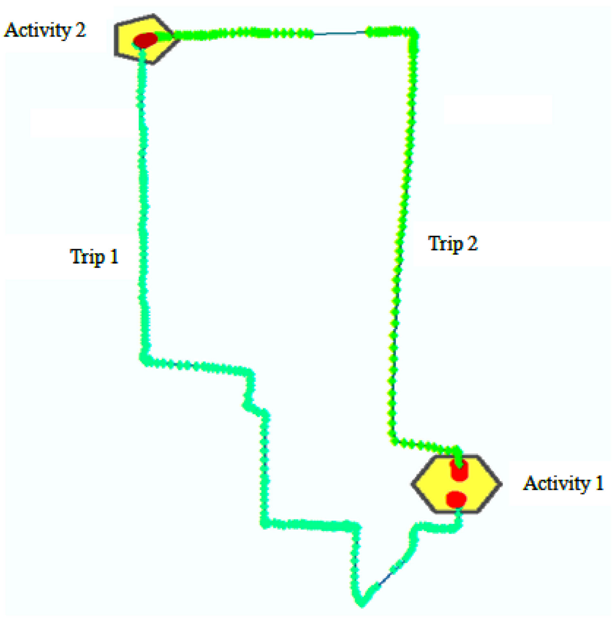

Fig. 5 *Example of identification results*

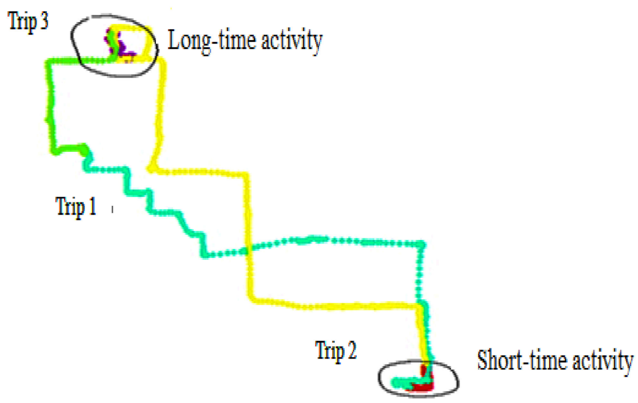

Fig. 6 *Daily trip and activity chain of respondent #13841*

higher than Du and Aultman-hall's study, which is 88% when omitting the parameter of heading change and limits the method in only car trip detection [7].

5 Conclusions

This paper designed a process for identifying activities and trips based on GPS survey data. By analysing the dwell time of stationary points, the proposed method recognised activities to determine the status of continuous data segments. A two-stage process for activity detection was proposed: stage 1 identified activities with a dwell-time threshold of 7 min and stage 2 identified short-time activities with a dwell-time threshold of 4 min. The results of the proposed identification algorithm demonstrated acceptable performance. By detecting short-time activities, this paper resolves the problem of overlooked short-time activities in conventional travel surveys and increases the accuracy of the trip detection. The method can also be applied to trip and activity detection with mobile phone data. By proposing methods for identifying trips and activities from GPS data, the findings provide a study basis for detecting additional travel information based on GPS data such as the travel mode and trip purpose and contributed to urban transportation planning and management.

The data filtering and repair methods conducted by this paper are not adequate for dealing with missing records due to urban canyon and subways. In addition, only 1 day GPS survey data are used in this paper. In a further study, we will pay attention to the methods of data filtering and repair as well as conducting a multi-day GPS survey and travel information identification. On the basis of the identification results of trips and activities, detection of

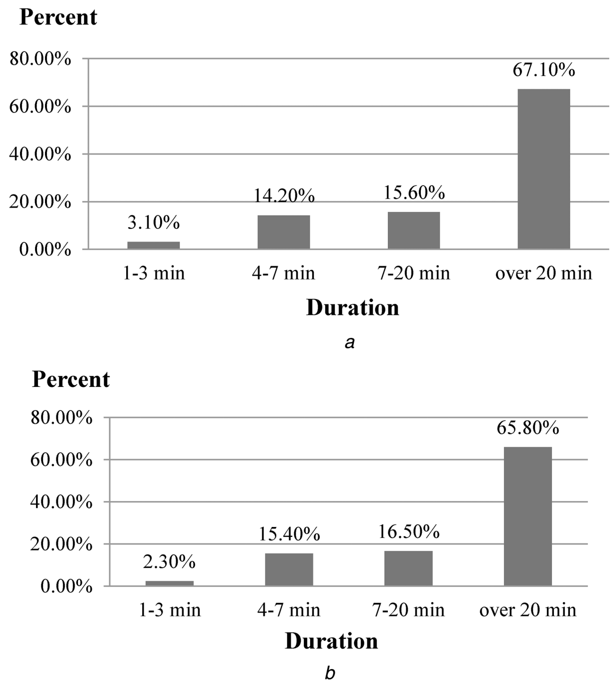

Fig. 7 *Distribution of activity duration*
(a) Distribution of activity duration for survey data, ***(b)*** Distribution of activity duration for identification results

travel mode, trip purpose, and other travel information will also be focused on in further studies. Moreover, except for the rule-based detecting method applied in this paper, there is also machine learning method which can be used in trip identification based on GPS data [27]. The follow-up study will focus on developing detection models with machine learning method.

6 Acknowledgments

We would express our sincere appreciation to the three reviewers for their constructive comments. The research is funded by the National Natural Science Foundation of China (71701215), the Humanity and Social Science Youth Foundation of Ministry of Education (14YJC630225), and Jilin Science and Technology Development Project (20150204021SF).

7 References

[1] Stopher, P., Fitzgerald, C., Zhang, J.: 'Search for a global positioning system device to measure personal travel', *Transp. Res. C*, 2008, **16**, (3), pp. 350–369

[2] Wolf, J.: 'Using GPS data loggers to replace travel diaries in the collection of travel data'. PhD thesis, Georgia Institute of Technology, 2000

[3] Tang, J. J., Liu, F., Wang, Y. H., *et al.*: 'Uncovering urban human mobility from large scale taxi GPS data', *Physica A*, 2015, **438**, (15), pp. 140–153

[4] Tang, J.J., Zhang, S., Liu, F., *et al.*: 'Statistical properties of urban mobility from location-based travel networks', *Physica A*, 2016, **461**, pp. 694–707

[5] Zhang, S., Tang, J.J., Wang, H.X., *et al.*: 'Revealing intra-urban travel patterns and service ranges from taxi trajectories', *J. Transp. Geography*, 2017, **61**, pp. 72–86

[6] Tang, J. J., Jiang, H., Li, Z. B., *et al.*: 'A two-layer model for taxi customer searching behaviors using GPS trajectory data', *IEEE Trans. Intell. Transp. Syst.*, 2016, **17**, (11), pp. 3318–3324

[7] Du, J.H., Aultman-hall, L.: 'Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: automatic trip end identification issues', *Transp. Res. A*, 2007, **41**, (3), pp. 220–232

[8] Qiu, P. Y.: 'Research on spatio-temporal mining method for trajectory data of urban travelers'. MS thesis, Beijing University of Civil Engineering and Architecture, 2012

[9] Forrest, T.L., Pearson, D.F.: 'Comparison of trip determination methods in household travel surveys enhanced by a global positioning system', *Transp. Res. Rec., J. Transp. Res. Board 1917*, 2005, **1**, pp. 63–71

[10] Hariharan, R., Toyama, K.: 'Project Lachesis: parsing and modeling location histories', *Geogr. Inf. Sci.*, 2004, **3204**, pp. 106–124

[11] Liu, J.H., Wolfson, O., Yin, H.B.: 'Extracting semantic location from outdoor positioning systems'. The Seventh Int. Conf. Mobile Data Management, Chicago, June 2006, p 73

[12] Zhou, C. Q., Franlkowski, D., Ludford, P.: 'Discovering personal gazetteers: an interactive clustering approach'. Proc. 12th Annual ACM Int. Workshop on Geographic Information Systems, Washington, January 2004, pp. 266–273

© The Institution of Engineering and Technology 2018

17519578, 2018, 8, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2017.0405 by INASP/HINARI - ARMENIA, Wiley Online Library on [17/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[13] Ma, X.L., Wang, Y., McCormack, E.*, et al.*: 'Understanding freight trip chaining behavior using spatial data mining approach with GPS data', *Transp. Res. Rec., J. Transp. Res. Board*, 2016, **2596**, pp. 44–54
[14] Ma, X. L., Dai, Z., He, Z. B.*, et al.*: 'Learning traffic as images: a deep convolution neural network for large-scale transportation network speed prediction', *Sensors*, 2017, **17**, p. 818
[15] Zong, F., Yuan, Y.X., Liu, J.F.: 'Identifying travel mode with GPS data', *Transp. Plan. Technol.*, 2017, **40**, (2), pp. 242–255
[16] Wang, Y.H., Li, F.C.: 'The analysis of the pedestrian traffic facilities and the traffic characteristics', *Stand. Traffic*, 2007, **10**, pp. 49–53
[17] Tang, J.J., Zhang, S., Chen, X.Q.*, et al.*: 'Taxi trips distribution modeling based on entropy-maximizing theory: a case study in Harbin city – China', *Physica A*, 2018, **493**, pp. 430–443
[18] Tang, J.J., Liu, F., Zhang, W.H.*, et al.*: 'Lane-changes prediction based on adaptive fuzzy neural network', *Expert Syst. Appl.*, 2018, **91**, pp. 452–463
[19] Juan, Z.C., Xianyu, J.C.: 'Daily travel time analysis with duration model', *J. Transp. Syst. Eng. Intell.*, 2010, **10**, (4), pp. 62–67
[20] Zhang, L., Wang, Y.P., Sun, J.*, et al.*: 'The sightseeing bus schedule optimization under park and ride system in tourist attractions', *Ann. Oper. Res.*, 2016, doi: 10.1007/s10479-016-2364-4
[21] Ma, X.L., Chen, X., Ding, C.*, et al.*: 'Sustainable station-level planning: an integrated transport and land use design model for transit-oriented development', *J. Clean Prod.*, 2018, **170**, (1), pp. 1052–1063
[22] Zong, F.: 'Understanding taxi driver's cruising pattern with GPS data', *J. Central South Univ. Technol.*, 2014, **21**, (8), pp. 3404–3410
[23] Yao, B.Z., Chen, C., Cao, Q.D.*, et al.*: 'Short-term traffic speed prediction for an urban corridor', *Comput.-Aided Civ. Infrastruct. Eng.*, 2017, **32**, (2), pp. 154–169
[24] Yao, B.Z., Chen, C., Song, X.L.*, et al.*: 'Fresh seafood delivery routing problem using an improved ant colony optimization', *Ann. Oper. Res.*, 2017, doi: 10.1007/s10479-017-2531-2
[25] Zmud, J., Wolf, J.: 'Identifying the correlates of trip misreporting-results from the California statewide household travel survey GPS study'. The Physical and Social Dimensions of Travel, Tenth Int. Conf. Travel Behavior Research, Lucerne, June 2003
[26] Wang, T.T., Chen, C.: 'Attitudes, mode switching behavior and built environment: a longitudinal study in the Puget sound region', *Transp. Res. A, Policy Pract.*, 2012, **46**, (10), pp. 1594–1607
[27] Yu, B., Wang, H.Z., Shan, W.X.*, et al.*: 'Prediction of bus travel time using random forests based on near neighbors', *Comput.-Aided Civ. Infrastruct. Eng.*, 2017, DOI: 10.1111/mice.12315

© The Institution of Engineering and Technology 2018

17519578, 2018, 8, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2017.0405 by INASP/HINARI - ARMENIA, Wiley Online Library on [17/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License