Recognition Method of Urban Residents' Travel Mode based on GPS Data

Huabin Liu[1, a], Chunfu Shao[2, b]

[1] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China;

[2] Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China.

[a]16120843@bjtu.edu.cn, [b]cfshao@bjtu.edu.cn

Abstract. The rising of big data and the Internet has brought about tremendous changes in travel. The rapid development of computer technology, network technology, wireless communication technology, portable devices, and location-based services provides an opportunity for the application of GPS technology to travel behavior survey. GPS technology has become a new technology to study urban residents' travel behavior and to identify urban residents' travel modes. This paper delivery a travel mode recognition method for urban residents' GPS travel data. Through the process of GPS data preprocessing, trajectory recognition and feature extraction, the recognition algorithm is designed to identify seven urban common travel modes, which are walking, bicycle, car, bus, taxi, subway and urban rail. In this paper, a trajectory recognition algorithm based on transition points is used to segment the trajectory of a single travel mode by identifying the transition points and pedestrian sections. The accuracy of the trajectory recognition process is about 79.8% by validating the open data set. For the extracted single trajectory, the Bagged Trees combined model is used to identify the travel mode with an accuracy of about 76.2%.

Keywords: GPS; travel mode; mode recognition; trip recognition.

1. Introduction

Resident travel survey is an important data base for traffic planning. The traditional method of travel log obtained by manual survey has many shortcomings, such as high cost of investigation, long data processing cycle, low recovery rate and reporting errors. With the popularization of advanced ICT devices and technologies such as smart phones and GPS devices, it is easier to obtain residents' dynamic travel information by using GPS technology. Through GPS travel data, it can excavate the traffic information of the participants and identify the user travel mode, which is of great practical significance to the development of traffic planning under big data and Internet+ background.

The existing methods of travel modes recognition are mainly divided into two kinds: empirical method and machine learning method. Reddy et al. (2010) [1] collect data through the GPS receiver and accelerometer built in mobile phone, discriminating walking, stationary, running, bicycle and car by Decision Tree model and First-order Discrete Hidden Markov Model. Brunauer et al. (2013) [2] take speed, acceleration and angular velocity as characteristic variables, using machine learning (Logistic Tree Model and Decision Tree) to recognize five travel modes, which are walking, bicycle, car, bus and train. Zhang Zhihua (2010) [3] conducts the first trip survey on GPS in China, using Multi-layer Perceptual Neural Network, Bayesian Network and Decision Tree Classifier to discriminate four travel modes, which are car, bus, bicycle and walking. Zhang Junfeng (2011) [4] uses the data set of volunteers' GPS travel tracking survey from the Fourth Comprehensive Traffic Survey of Beijing in 2010, extracting the characteristics of travel distance, travel speed and travel speed distribution in different travel modes and uses Fuzzy Theory model to recognize travel modes.

At present, the researches of travel survey based on GPS data in foreign countries is earlier and more fruitful, while the research based on domestic travel characteristics is less. Based on domestic urban residents' travel characteristics, this paper identifies seven urban common travel modes, which are walking, bicycle, car, bus, taxi, subway and urban rail. The method is validated by Geolife project open data set [5,6, 7].

2. Recognition Model of Travel Modes for Urban Residents

According to the GPS trajectory information of a resident's trip, this paper proposes a recognition method combining the trajectory segmentation based on the transition point and the machine learning feature recognition using machine learning algorithm. Firstly, the method identifies the transition points by stopping-interval points and stationary points, and divides the GPS trajectory into several sub-trajectory segments, which has only one travel mode. The travel characteristics are extracted from the sub-trajectory segments, and the Bagged Trees model is used to identify the travel modes of these sub-trajectory segments.

2.1 Data Preprocessing.

GPS data preprocessing is the basis and prerequisite for travel mode recognition. Commonly used GPS data preprocessing process mainly involves drift removal, format conversion, basic feature data calculation, trajectory smoothing and other steps. In this paper, the trajectory points outside the study area (longitude, latitude, elevation) and at speeds exceeding 100 m/s are removed, and the Kalman Filter algorithm is used to smooth the trajectory.

2.2 Trajectory Segmentation based on Transition Points.

In the process of traveling, residents always keep almost stationary or moving at a low speed within a certain range for a period of time before changing the travel mode. The sub-trajectory segments of a single mode of transportation can be obtained by dividing the GPS trajectory by these transition points. At present, trajectory segmentation methods mainly aim at the residence situation, and identify sub-trajectories by residence time interval, residence point set and stationary point method.

The designed trajectory segmentation method is as follows:

(1) Change the trajectory points into the minimum trajectory segment by connecting two adjacent points;

(2) Recognition the active points through the stay interval method ($T>90s$);

(3) Supplemented by the stationary point method (the track length which $V<0.6m/s$);

(4) Identify the travel segment without stopping interval by walking segment ($0.6 < V < 2m/s$).

The specific algorithm is shown in the Table 1.

Table 1. Pseudo code of Trajectory Segmentation Algorithm

Process
For each new trajectory file for each user i:
Mark the kth trajectory segment of time T > 90s as active point;
Mark the kth trajectory segment of V < 0.6m/s in the inactive point for 8 minutes as quasi-active transition point 1 (to prevent pseudo-static caused by congestion);
Mark the kth trajectory segment of V < 2m/s in the quasi-active point for 2 minutes as quasi-active transition point 2 (to prevent pseudo-walking caused by congestion);
Re-judge the unmarked points with trajectory length of quasi-active transition points of 1 and 2 segments over 4 minutes;
Reconstruct the trajectories between active and quasi-active points;
End

The traditional accuracy indexes of evaluation travel recognition are mainly the precision rate P1 and recall rate P2 proposed by Zhou et al. (2005) [8].

$$P1=RD/D\times100\%. \tag{1}$$

$$P2=RD/R\times100\%. \tag{2}$$

Where RD is the actual travel endpoint identified; R is the actual stroke endpoint; D is the identified endpoint of the stroke.

2.3 Travel Feature Extraction.

Select the velocity distribution, travel distance, travel time and other indicators to analyze from the data set. It can be found that choosing 12 characteristic variable, which are the total length of travel section, the mean of speed, the mean of positive acceleration, the probability that the speed is greater than 6m/s, parking rate, average speed, 85% speed, speed variance, acceleration variance, the maximum three speed mean, the maximum three positive acceleration mean and the speed kurtosis coefficient can better express the operation characteristics of seven modes of transportation.

The Stop Rate is defined as:

$$SR=\Sigma dis(v<0.3)/ \Sigma dis. \tag{3}$$

Where dis is the length of each minimum trajectory segment in the journey; v is the average velocity between each trajectory segment.

Velocity kurtosis coefficient vKut [9] is defined as:

$$vKut=\Sigma(v_i-v')^4/(n-1)\ \ Var(v)\ ^4-3. \tag{4}$$

Where v_i is the average velocity of the first trajectory section in the travel section; v' is the average speed of the travel section; Var(v)is the travel speed variance.

2.4 Travel Modes Recognition.

Compared with many classification algorithms, Decision Tree algorithm has the advantages of short training time, low complexity, fast prediction, easy understanding and low requirements for data processing. Especially, Decision Tree algorithm can make feasible and effective classification of large data sources in a short time, which is more suitable for the characteristics of this data set. Considering the disadvantage of over-fitting Decision Tree, the Bagged Trees model combined with Random Forest model is used to reduce the disadvantage of single decision tree. The combination model reduces the one-sidedness of single decision tree by generating N simpler decision trees.

3. Analysis of Experimental Results

The GPS data from the Geolife project collected by Microsoft Research Asia were used as experimental data. The data set recorded the travel trajectories of 182 users from April 2007 to August 2012, including a total of 7,621 trajectories.

3.1 Sub-trajectory Segmentation Accuracy.

Due to the diversity of the data, including the travel data under normal conditions and congestion conditions, the probability of false stationary or slow speed generated by different modes of transportation is different. For example, for the urban railway, the stopping time between stations is significantly higher than the bus. All these problems make it difficult to determine the time threshold of the travel endpoint. The precision rate of the test training set is 79.8% and the recall rate is 29.4%.

The result of trajectory segmentation for a extracted track file is shown in Table 2.

Table 2. Example of Sub-trajectory Segmentation Results

Trajectory Point Numble	Travel Mode	The Number of Change in Travel Mode during one Trip	Type of Quasi-active Transition Point
1	car	1	1
1449	walk	2	1
1514	walk	2	1
1896	car	3	1
2865	walk	4	1
3265	walk	4	1

The points numbered of 1, 1449, 1896 and 2865 in the above table were identified accurately, while the points 1514 and 3265 were over-identified.

3.2 Sub-trajectories Travel Mode Recognition Accuracy.

The data set is divided into training set and test set by 2:1. Twelve feature of each sub-trajectory are calculated. The Bagged Trees model is used to recognize the sub-trajectory travel mode. The accuracy of the output model is 76.2%. The results are shown in Table 3.

Table 3. Recognition Effect

Travel Mode	Bike	Bus	Car	Walk	Subway	Taxi	Train
Accuracy	68.2%	63.7%	37.9%	86.1%	66.5%	37.8%	79.4%

From the above table we can see that the recognition method proposed in this paper can basically recognize walking, bicycle and other modes, but it cannot effectively distinguish bus, car and taxi, three kinds of motor vehicle travel modes, and the existing extracted feature quantities cannot essentially distinguish these three modes.

4. Summary

Using the individual residents' travel data set as a support, this paper designs a matching algorithm to recognize seven common urban travel modes, which are walking, bicycle, car, bus, taxi, subway and urban rail, only through the GPS space-time trajectory data. The original GPS trajectory is divided into sub-trajectory segments of a single mode of transportation by using the trajectory segmentation method based on transition points, and extract the travel characteristics. The Bagged Trees combination model is used to identify each sub-trajectory, which reflects the actual effect of using GPS data alone to recognize travel modes to a certain extent.

References

[1]. Reddy S, Mun M, Burke J, et al. Using mobile phones to determine transportation modes[J]. ACM Transactions on Sensor Networks (TOSN), 2010, 6(2): 13.

[2]. Brunauer R, Hufnagl M, Rehrl K, et al. Motion pattern analysis enabling accurate travel mode detection from gps data only[C]//Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on. IEEE, 2013: 404-411.

[3]. Zhihua Zhang: Deriving Trip Information from GPS Trajectories (PhD di ss, East China Normal University, China, 2010).

[4]. Junfeng Zhang: Reaserach on OD Survey Based on the Global Positioning System (MA di ss, Beijing Jiaotong University, China, 2011).

[5]. Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain. ACM Press: 791-800.

[6]. Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma. Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea. ACM Press: 312-321.

[7]. Yu Zheng, Xing Xie, Wei-Ying Ma, GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited paper, in IEEE Data Engineering Bulletin. 33, 2, 2010, pp. 32-40.

[8]. Zhou C, Ludford P, Frankowski D, et al. An experiment in discovering personally meaningful places from location data[C]//CHI'05 Extended Abstracts on Human Factors in Computing Systems. ACM, 2005: 2029-2032.

[9]. Jianxin Chen, Transportation Mode Recognition Based on AML-BP Algorithm (MA di ss, Zhengzhou.