AUTOMATIC ACTIVITY IDENTIFACTION FROM RAW GPS VEHICLE TRACKING DATA

Lian Huang[a,b]*, Qingquan Li[a,b], Bijun Li[a,b]

[a] State Key Laboratory of Information Engineering in Survey, Mapping and Remote Sensing, Wuhan University, No. 129, Luoyu Road, Wuhan, 430079, PR China, - (huangliansinc@hotmail.com, qqli@whu.edu.cn，Lee@whu.edu.cn)
[b] Engineering Research Center for Spatio-Temporal Data Smart Acquisition and Application, Ministry of Education of China, No.129, Luoyu Road, Wuhan, 430079 PR China

Commission I, WG I/5

KEY WORDS: Activity Identification, GPS Trajectories, Multi-Variants Analysis, Intelligent Transportation System

ABSTRACT:

Recently, activity-based analysis using GPS equipment as data collector has being a hot issue. But most researches focus on data from wearable GPS recorder for person because of easy detailed activity logging and interactive validation with users. Nevertheless, available floating car data and geographic context databases provide the possibility for activity-based analysis based on GPS vehicle tracking trajectories, this paper presents a novel and efficient approach to automatically identify activity-locations as well as activity-types from raw GPS vehicle tracking data. A number of contextual variants related to the activity locations including temporal information, spatial information and probability information are considered with the help of digital map containing Points of Interest (POIS). Taking the three aspects of inputs mentioned above into a Multi-Variants Analysis framework, detailed information of each activity will be identified. Finally, experiments using real world floating car data are conducted for evaluation and to show how this approach will help in varieties of applications in both traffic and activity-based analysis.

1. INTRODUCTION

1.1 Background

As a widely-used traffic data acquisition technique, floating car system can not only provide historic and real time traffic information but also network constraint trajectories travellers moved through. Generally floating car data (FCD) is used for traffic parameters calculation such as road segment average speed and link travel time estimation, but actually with rich background knowledge, digital urban district region map for example, more in-depth FCD based analysis becomes possible.

In recent decades, activity-based analysis using GPS equipment as data collector has being a hot issue. Most this kind of researches focus on data from wearable GPS recorder for person because of easy detailed activity logging and interactive validation with users (B. Kochan, T. Bellemans, D. Janssens and G. Wets, 2006). Nevertheless, existing huge FCD and geographic context databases provide the possibility for conductible activity-based analysis based on GPS vehicle tracking trajectories such as finding out hottest locations at certain time of a day (Stefan SCHÖNFELDER and Kay W. AXHAUSEN, 2002). However, different from conventional activities analysis applications, activity- types (at home, at work, or shopping) and activity-locations are not labelled in raw tracking data, and this information is difficult and not reliable to be manually added. With the motivation to solve this problem, this paper presents a novel and efficient approach to automatically identify activity-locations as well as activity-types from raw GPS vehicle tracking data.

Ron and Pavlos (2007) conclude and summarize current advances in activity-behaviour research. Jean Wolf , Randall Guensler , William Bachman (2001), Stefan and Kay (2003), Wendy Bohte, Kees Maat (2009) introduce basic ideas using GPS tacking data to conduct traditional travel diary research , complete extensive experiments and proposed different ways for validation. Daniel Ashbrook, Thad Starner proposed method for potential activity place clustering. Lin Liao et al (2005) use related markov networks to classify activies into six predefined types, which is suitable for trajectory data from private cars. Yu et al (2009) build framework and platform for tracking data mining and clustering, which pay more attention to aggregative analysis and data mining.

2. PROBLEM STATEMENT

For a trajectory composed of discrete GPS tracking points, conceptual data model can be used to reduce data redundancy and enrich semantic information (Lian, 2009). Then locations where activities probably occurred in a trip as well as routes between these locations will be recorded to represent original points based trajectories. In this way, the problem discussed in this paper is that: given a network constraint trajectory $Tr(CarID,T)$ with a series of potential activity-locations $Al(x, y, st,et)$, a POI (point of interest) database $AP(x, y,\varepsilon(t))$,where T is the time span of Tr including starting time and ending time; $CarID$ is a tag indicating the corresponding vehicle; x and y are horizontal coordinates; st is the time when the driver stopped; et is time when the driver left, $\varepsilon(t)$ is the time-attractiveness function of each POI, figure out possible activities $AT(Ap, st, et, Ac)$, where Ap is the POI where an activity was going on and Ac is the confidence for Ap to be assigned to AT.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

We use floating car data collected by taxies in Wuhan city, China, and thus the method proposed is based upon this type of data source.

3. METHODOLOGY

Generally candidate locations for activities identification from raw GPS tracking data are not directly available. In this case t, tracking points clustering methods (Daniel Ashbrook, 2003) will be developed to find the centre of clustered points as the location where an activity was conducted since GPS points during activities will be recorded as series of “floating” points around the place where the vehicle stopped, which results from a systematic error of GPS sensors. Then, activity duration, starting time, ending time are obtained using the time stamps of the first and last points in the cluster. Whereas, if taxies are used for floating car to collect GPS tracking points, passengers on/off information is recorded additionally and usually used for trips division. In this paper, the on/off changing points along trajectories from taxies are considered as confident places for passengers’ activities identification.

3.1 Defined Temporal and spatial rules

A detailed activity normally includes activity type like “dining”, “working”, activity spot which refers to a specific construction like “Starbucks at 5th avenue”, and duration indicates how long people stayed. However, GPS tracking data from taxies are not validate to obtained durations because in most cases taxi drivers won’t wait for the passengers they just dropped. Therefore, time of day and day of week are the only temporal factors will be used for activities identification.

Network distances from POIs those within a predefined circular buffer zone to on/off changing points are taken as spatial-aware factor. The closest POI has the greatest possibility in space to be identified as *Ap*. The radius of the buffer zone will be adaptively changed according to the density of nearby POIs, generally the more POIs around the smaller the radius is.

The available database includes three types of POIs: restaurants (coffee/tea house included), shops, and public servings. Accordingly, we define three types of activities: *dining*, *shopping*, and *others*. Table 1 and Table 2 show basic rules of temporal factors on these three types. Possibilities of “high”, “medium”, “low” will return scores of 3, 2, and 1 respectively.

Time\Type	dining	shopping	others
0:00-8:00	low	low	high
8:00-10:00	high	medium	low
10:00-12:00	medium	medium	medium
12:00-14:00	high	low	low
14:00-18:00	low	medium	medium
18:00-20:00	high	medium	medium
20:00-22:00	low	medium	medium
22:00-24:00	medium	low	medium

Table1. Temporal rules for activities: Weekdays

Time\Type	dining	shopping	others
0:00-8:00	low	low	high
8:00-10:00	high	medium	low
10:00-12:00	medium	high	medium
12:00-14:00	high	medium	low
14:00-18:00	medium	high	medium
18:00-20:00	high	high	medium
20:00-22:00	medium	high	medium
22:00-24:00	medium	low	medium

Table1. Temporal rules for activities: Weekends

3.2 Activity chains

Temporal and spatial factors are used for single activity identification. In a complete trip, the origin sometimes has a significant impact on destination and vice versa. For example, people rarely go to another restaurant immediately after a meal, so “shopping” or “other” is more likely to be assigned to the corresponding activity of destination when “dining” is assigned to previous activity. Based upon investigated information from internet, how the activity chains will affect activities identification is described in Table 3.

Activity chain	Possibility
Dining--Dining	low
Dining--Shopping	high
Dining--others	medium
Shopping--Shopping	medium
Shopping--Dining	high
Shopping--others	medium
Others--Dining	medium
Others--Shopping	medium
Others--Others	medium

Table 3. Activity chains effect on activities identification

3.3 Multi-inputs analysis method

As discussed above, temporal information, spatial information and probability information are considered for activities interference. Furthermore, POIs have different kinds of attractiveness to people which will provide a fourth aspect for identification.

Taking the four factors mentioned above into a multi-variants analysis framework, the inputs of proposed method are temporal factor *Wt*, spatial factor *Wg*, activity chain factor *Wa*, and attractiveness factor *Wε* (Equations (1)).

$$
\begin{aligned}
&Wt = F_t(st/et) \\
&Wg = F_g(Al(x,y), AP(x,y), G) \\
&Wa = \begin{cases} F_a(AT_{O/D})\_if\_AT_{O/D\_exist} \\ 0 \end{cases} \qquad (1) \\
&W\varepsilon = AP(\varepsilon(t))
\end{aligned}
$$

Where,

$F_t()$ is the function to calculate *Wt* according to temporal rules in section 3.1;

$F_g()$ returns network distance between *Al* and *AP*, G is road network;

$F_a()$ is the function to calculate *Wa* according to Table 3 in section 3.2;

In the case of taxi GPS tracking data, *st* and *et* of *Al* are the same. To collect the attractiveness function of each POI, we use information from restaurant/shopping/facilities ranking website as showed in Figure 1. The higher the POI ranks the greater value the *ε(t)*will returns.

武汉最佳餐厅 更新于2010-04-30

排名	商户	商区	口味	环境	服务	人均
1	湖锦酒楼(正银国宴店)	西北湖/新世界百货	26	26	22	￥74
2	老汉口	江汉路步行街	24	28	25	￥91
3	戈雅牛排法餐厅(青年路店)	西北湖/新世界百货	24	25	26	￥121
4	Bonjour(龙阳店)	沌口	28	23	24	￥24
5	仟吉西饼(万达店)	江汉路步行街	26	24	22	￥18
6	恭启福鲍鱼圣汤火锅(光谷街店)	光谷/鲁巷	25	24	23	￥167
7	樱花糕坊(海寿街店)	客运港/江滩	27	21	20	￥16
8	亢龙太子酒轩(金融店)	西北湖/新世界百货	22	27	22	￥77
9	天府映象火锅食府	武广/万松园	21	21	30	￥37
10	亚洲饼屋	硚口路	24	22	23	￥17
11	万丽咖啡厅	徐东大街	21	27	25	￥172
12	红鼎豆捞(水果湖店)	水果湖/东湖路	25	24	20	￥133
13	西北湖咖啡烘培工坊	西北湖/新世界百货	27	15	22	￥12
14	Yogo juice(武汉天地店)	解放公园/永清	24	22	23	￥29
15	荆州青莲酒家(中北路店)	徐东大街	23	26	21	￥147
16	周黑鸭(司门口店)	阅马场	29	16	17	￥17

Figure 1. POI ranking website_restaurant page

To figure out the confidence of each candidate POI *i*, *Ac is* defined as Equation (2).

$$Ac_i=\alpha Wt+\beta Wg+\gamma Wa+\zeta W\varepsilon \qquad (2)$$

Where,

α, β, γ, ζ are coefficients corresponding to the inputting factors.

Neuron-network are used for the configuration of α, β, γ, ζ, with a number of simple trajectories as training sample where activities identifications are obvious, for example, only two or one POI around a *Al.*

To reduce possible misjudge on activities, we select top 3 POIs with highest confidence when candidates are more than 3, and record relative confidence Ac_i' *as* Equation (3) instead of Ac_i

$$Ac_i' = Ac_i / \sum_{j=1}^{k} Ac_j \qquad (3)$$

Where,

k is total number of candidate POIs.

4. EXPERIMENTS AND EVALUATIONS

4.1 Data description and experiment setup

Trajectories data from a taxi's one-day collection along with road map of Wuhan city, China and restaurant/shopping/public serving POIs are taken for experiment setup to evaluate proposed method (Figure 2).

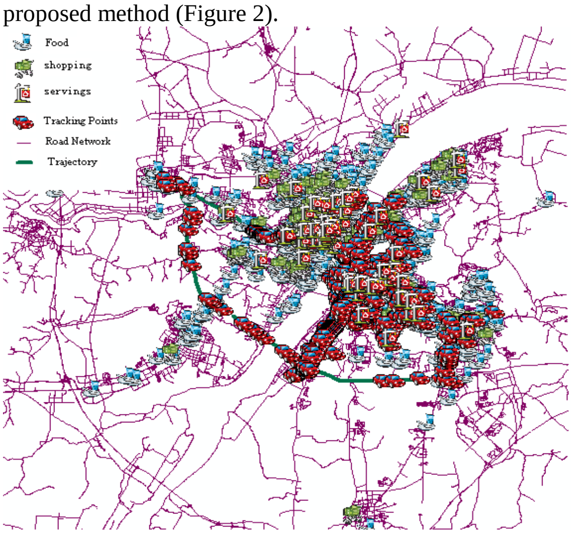


Figure 2. Tracking points and context map

The whole trajectory is divided into sub trajectories according to on/off changing information from raw tracking data. The origin and destination of each sub trajectory are used as *Als* for activities identification.

4.2 Results and evaluations

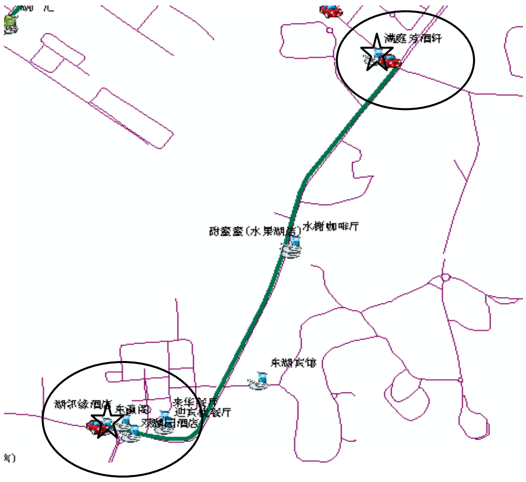


Figure 4. Experiment 1

As shown in Figure 4, only one restaurant (Mantingxuan Reataurant) POI falls into the trajectory's origin's buffer zone(upper left), thus the activity will be identified as "dining", and related *Ap* and time stamp are recorded. Whereas, four restaurant POIs fall into destination's buffer zone, then their confidences are calculated as mentioned in section 3.3. "Hulinyuan Restaurant", "Dongyingge Restaurant" and "Double Lakes Restaurant" are selected as potential *Ap* with relative confidence 51%, 34%, 15% respectively. In this case, the type of POIs are all "restaurants", thus attractiveness and *Wg* shows greater impact on the activity interference. Trip purpose based on the result can be described as "heading for a more preferable restaurant".

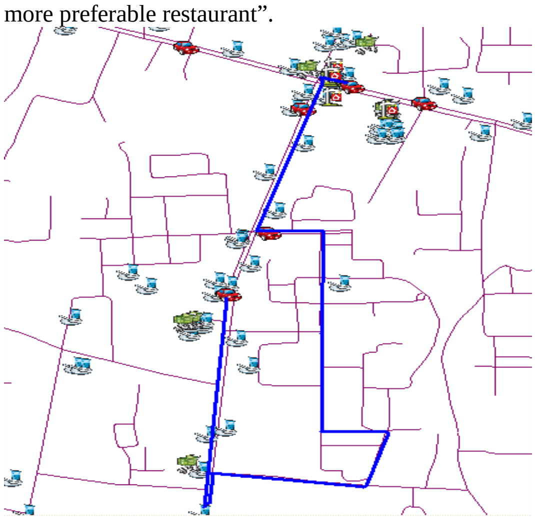
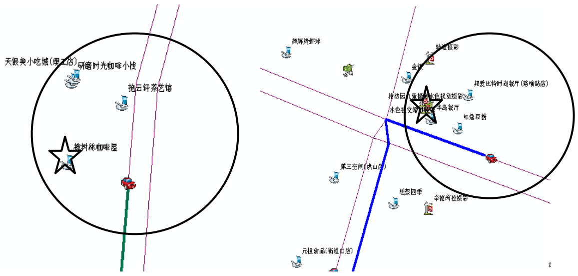


Figure 5. Overall trajectory of experiment 2



(a) (b)

Figure 6. Activity identification in complex environment

Figure 5 shows a more difficult case for activities identification where plenty of POIs fall into buffer zone at both origin and destination. Four POIs are found as candidate *Ap* when the trip started (Figure 6 (a)), and "Xiangshulin Coffee" is the one with the highest confidence. Five POIs, two shopping mall and three restaurant, show probability to be selected as *Ap to* destination (Figure (b)), and finally a public serving POI obtains the highest confidence with regards to the time stamp 16:00, and previous activity identified as "dining" even though the poi "Hongdingdoulao Restaurant" is closer to the end of the trajectory.

5. CONCLUSIONS

Booming available GPS tracking data and rich context maps provide possibility to infer activities along trajectories. This paper proposes a muti-variants inputting method considering temporal, spatial as well as probability factors, which can automatically identify activities in both simple and complex environments. Experiment with field data validates this approach. Our future work will focus on in-depth activities identification where more types POIs are available, and developing online validation system to further evaluate the effectiveness of this method.

REFERENCES

Baibing Li, 2009, Markov models for Bayesian analysis about transit route origin-destination matrices, *Transportation Research Part B: Methodological, 43(3),* pp.301-310

B. Kochan, T. Bellemans, D. Janssens and G. Wets, 2006, Dynamic activity-travel data collection using a GPS-enabled personal digital assistant, Proceedings of the ninth International Conference on AATT, pp.319-324

Daniel Ashbrook, Thad Starner, 2003, Using GPS to learn significant locations and predict movement across multiple users , *Personal and Ubiquitous Computing, 7(3), pp.275-286*

Donald J. Patterson, Lin Liao, Dieter Fox, and Kautz, 2003， Inferring High –Level Behaviour from Low-Level Sensors， Ubiquitous Computing, 5th International Conference, Seattle, WA, USA, October 12-15

http://www.dianping.com/Wuhan

Jean Wolf , Randall Guensler , William Bachman , 2001， Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data, Transportation Research Board 80th Annual Meeting, January 7-11

Kay W. AXHAUSEN, Stefan SCHÖNFELDER, J Wolf, M. Oliveira, 2003, U Samaga, 80 weeks of GPS-traces: Approaches to enriching the trip information, 83rd Transportation Research Board Meeting

Lian Huang, Qingquan Li, 2009, Floating Car Trajectories Modeling for Traffic Information Analysis, AAG2009, Las Vegas

Lin Liao, Dieter Fox and Henry Kautz, 2005, Location-Based Activity Recognition using Relational Markov Networks, Proceedings of the International Joint Conference on Artifical Intelligence (IJCAI)

Ron N.Buliung and Pavlos S.Kanaroglou, 2007, Activity-Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modelling, *Transport Reviews,*27(2), pp.151-187

Stefan SCHÖNFELDER and Kay W. AXHAUSEN, 2002, Exploring the potentials of automatically collected GPS data for travel behaviour analysis- A Swedish data source, *GI-Technologien für Verkehr und Logistik，* 13, pp.155-179

Wendy Bohte, Kees Maat, 2009, Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large scale application in the Netherlands, *Transportation Research Part C: Emerging Technologies,* 17(3),pp.285-297

Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma, 2009 Mining Interesting Locations and Travel Sequences from GPS Trajectories , Proceedings of the 18th international conference on World wide web, pp.791-800