TOURIST ROUTE CHOICES AND SHORT-TERM FLOW PREDICTIONS IN TOURIST AREAS BASED ON WI-FI PACKET DATA

Yuhan GAO[1] and J.-D. Schmöcker [2]

[1]Dept. of Civil Eng., Kyoto University
(Katsura C-Cluster C1-2-434, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail:gao@trans.kuciv.kyoto-u.ac.jp
[2]Member of JSCE, Assoc.Prof., Dept. of Civil Eng., Kyoto University
(Katsura C-Cluster C1-2-434, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail: schmoecker@trans.kuciv.kyoto-u.ac.jp

This research proposes a methodology to use the records from Wi-Fi packet sensors to model route choice and time spent at different locations within the tourist area. In contrast to other route choice problems, in touristic areas often "the route is the goal" so that we expect the choice behavior to be less destination oriented, or there might not even be a specific destination. A data collection survey conducted in Higashiyama Ward, Kyoto City, utilizing Wi-Fi packet sensors is described. From the survey data we first extract trip chains by tourists. We then construct a network based on map data and the location of Wi-Fi sensors in order to employ recursive logit model as the method to formulate tourists' route choice behavior in the Higashiyama area. We propose *number of POIs* as a link attribute to reflect link attractions in tourist areas. The estimation results imply that *number of POIs* helps improve the prediction.

Key Words : *Wi-Fi packet sensing, tourism, route choice, recursive logit*

1. BACKGROUND

Over 55.22 million tourists visited Kyoto City in 2016, and the travel demand in Kyoto City is still rapidly growing. Consequentially the crowding in Kyoto is believed to be a major consideration in causing tourism dissatisfaction, as reported by Kyoto City[1)]. Understanding tourist flows inside Kyoto City is the foundation of solving related problems such as estimating spatial/temporal attraction, evaluation of transportation services and evacuation planning. Tourism flow data used to be obtained by traditional traffic surveys. However, routine surveys capture tourism behavior only to a limited degree. Large-scale bespoke surveys at touristic points are rare, the last major one being carried out in Kyoto 13 years ago. They can hardly capture the dynamics in tourism due to word-of-mouth, weather conditions and other dynamics. Furthermore the understanding of tourist behavioral response due to crowding at sightseeing spots is limited.

With the rapid development and spread of smartphones, Wi-Fi packet sensors are providing new possibilities in terms of obtaining tourism flow data. The sensors are designed to detect and record all electronic devices enabling Wi-Fi function within an average radius around 40 meters. Any device detected by different sensors can be identified by an anonymized label encrypted from the device's MAC address. In addition to the hashed MAC address, Wi-Fi packet sensors also record: 1) timestamp; 2) packet sensor ID; 3) Received Signal Strength Indication (RSSI); 4) device's vender ID. By multiple detections and above attributes, routes of any individual who carries detectable devices can be observed.

Wi-Fi packet data has several advantages: 1) Rich sample size; 2) 24h monitoring; 3) real-time, which seem to be the answer to the limitations of traditional traffic survey above. More specifically, compared to other electronic data set such as GPS data, with Wi-Fi packet data it is easier to obtain unified samples for not relying on any smartphone application.

2. LITERATURE REVIEW

(1) Recursive logit model

Rather than choosing the shortest path between two sightseeing spots, tourists are more likely thinking "the route is the goal". We assume that walking tourists' route choices in touristic area are less destination-oriented than common. Thus, we formulate their route choice by a sequential route choice model.

Fosgerau et al.[2] proposed the recursive logit (RL) as a link-based model which doesn't require for generating the choice set. Individuals' route choice is decomposed into a bunch of sequential link choices, with not only instantaneous utility under given condition but downstream utility expected being taken into account. The general concept of RL is shown in Fig.1. At each link k individual n chooses the next link a among the set of outgoing links $A(k)$, maximizing the sum of instantaneous utility $u(a|k)$ and downstream utility $V_n^d(k)$. $u(a|k)$ and $V_n^d(a)$ is given as follows.

$$u(a|k) = v(a|k) + \mu\varepsilon_n(a) \tag{1}$$

$$V_n^d(k) = E\left[\max_{a\in A(k)}\left(v(a|k) + V_n^d(a) + \mu\varepsilon_n(a)\right)\right] \quad \forall k \in A \tag{2}$$

Equation (2) is known as the Bellman equation[3]. By the Markov property of traveler's sequential link choice, this dynamic choice model is proved to be equivalent to an MNL model with infinite alternatives. In the same paper, a link additive attribute "Link Size" is introduced for correcting correlation between links. Kaneko et al.[4] introduce "Link Awareness" and show that it improves the stability of model estimations. Mai et al.[5] further develop the RL model into a nested recursive logit model to relax the independence of irrelevant alternatives (IIA) by giving scale parameters specific to each link. Oyama and Hato[6] argue that decision makers have insufficient information for future decision-making states and introduce a discount factor to the expected downstream utility in a recursive logit approach for modelling people's myopic behavior in sequential route choices.

(2) Activity-based modelling

Since the Wi-Fi packet data and other localization data such as Bluetooth traces are unable to record positions continuously, most of the study using Wi-Fi packet data are activity-based. Ihoroi et al.[7] employ stochastic block models as a co-clustering method to classify the travel patterns by looking at both the type of tourists and destinations. Versichele et al.[8] analyses pedestrian flow patterns aggregately at Ghent Festival in Belgium in use of Bluetooth data.

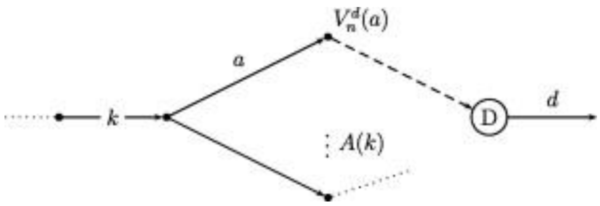

Fig.1 Concept of the recursive model

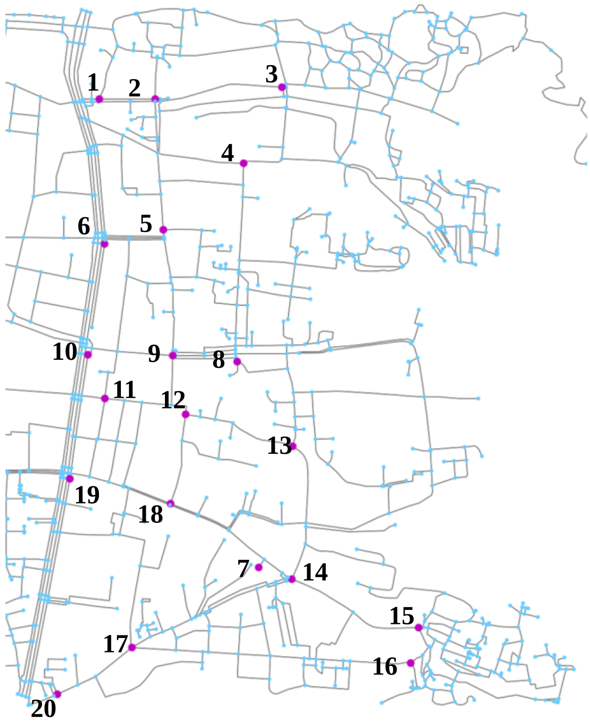

Fig.2 Survey area

3. OBJECTIVES AND METHODOLOGY

This research proposes a methodology to use the records from Wi-Fi packet sensors for understanding and predicting tourist flows in a disaggregate level. We employ RL model to formulate tourists' route choice behavior within a tourist area. In contrast to other route choice problems, for tourists in often "the route is the goal" so that we expect the choice behavior to be less destination oriented, or there might not even be a specific destination.

In order to implement the sequential link choice model in use of Wi-Fi data, the input network is specifically reconstructed in respect of locations of Wi-Fi sensors and geology in the survey area. Individuals' route choices are extracted from Wi-Fi packet data.

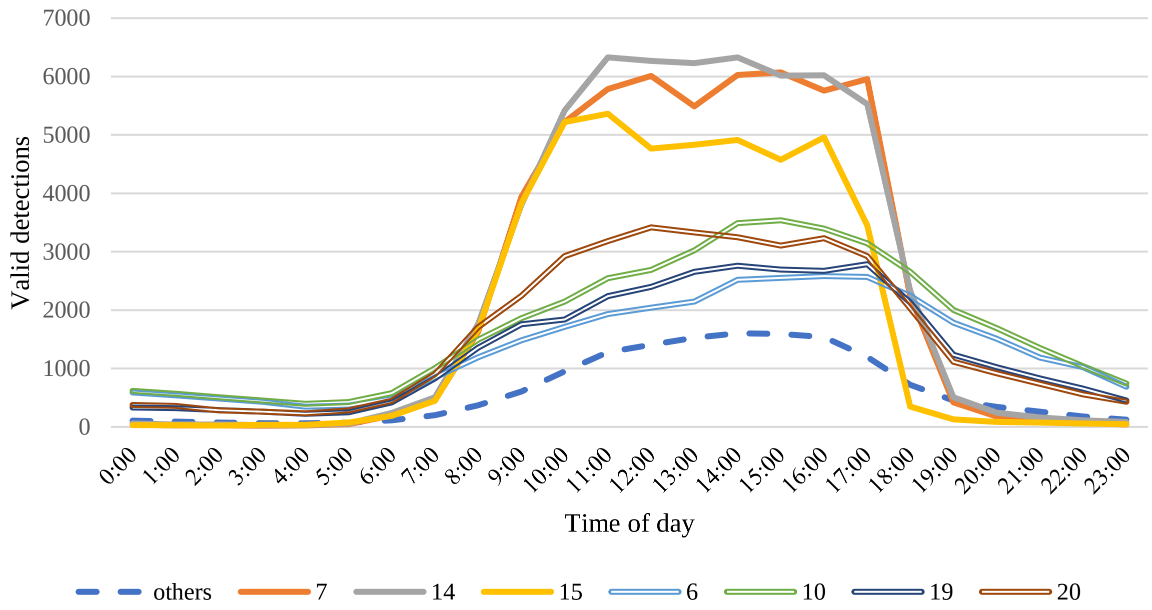

Fig.3 Daily detections along hours by sensors shown in Figure 2

4. SURVEY DESCRIPTION

(1) Survey area

The data collection experiment was conducted in Higashiyama Ward, Kyoto City from November 2017 to March 2018. 20 sensors were installed in street blocks of this area, as shown in **Fig.2** in purple. The area is about 700 meters long from west to east and 800 meters from north to south. Well-known sightseeing spots of Kyoto are located inside this area. Sensor No.1 is set 50 meters south to the entrance of Yasaka Shrine and sensor No.15 and 16 are near the entrances of Kiyomizu Temple. Since the east of the Higashiyama area is a mountainous area of Kyoto with poor associability, most of the tourists enter this area by a north-south trunk road along sensor No.6, 10, 19 and 20.

(2) Wi-Fi packet data description

Averagely around 90,000 anonymized MAC address is recorded at least twice each day. These repeated records are assumed valid for route extraction and supposed to belong to either pedestrian or driver.

Fig.3 shows the daily counts of valid detection along hours by sensors. For the sake of simplicity, one curve is used to refer the average counts from all the sensors that will not be mentioned in detail in this section. We observed that for each sensor the detection begins to increase fast around 8:00 and reach the peak around midday. At daytime the counts from sensor No.7 (a bus parking area for Kiyomizu Temple), No.14 (a path connects the parking area and Kiyomizu Temple) and No.15 (the entrance of the temple) are far higher than the counts from other sensors. These counts drop quickly in the evening after 17:00 when Kiyomizu Temple closed for visit. Also, looking at the sensors along the main road of the survey area plotted in double line in **Fig.3**, (sensor No.6, 10, 19 and 20), the counts are in the second-highest group after those sensors related to Kiyomizu Temple. The counts along the main road do not drop obviously (or even raise slightly) in the evening.

5. NETWORK CONSTRUCTION

As a link-based model, RL requires individuals' sequential link choice and an acyclic network as its input. Unlike GPS data that record devices' coordinates nearly continuously, Wi-Fi packet data only record devices when they are close enough to sensors. Thus, the footage of a device (assumed as an individual in this study) in the Wi-Fi packet dataset is a sequence of the passed sensors. Routes taken between any two sensors are unknown. Therefore, different from the studies which implement RL with GPS data on an actual map, it is necessary to extract a specific network from the origin network. This reduced network only contains nodes that correspond to sensors and dummy absorbing nodes in order to utilize Wi-Fi packet data.

(1) Network extraction

In this study, the network is reconstructed based on data from the pedestrian map provided by OpenStreetMap (OSM). The map is a multi-directed graph with more than 900 nodes and 2500 links. This original network is simplified into a directed graph with only 21 nodes and 119 links. Within these 21 nodes

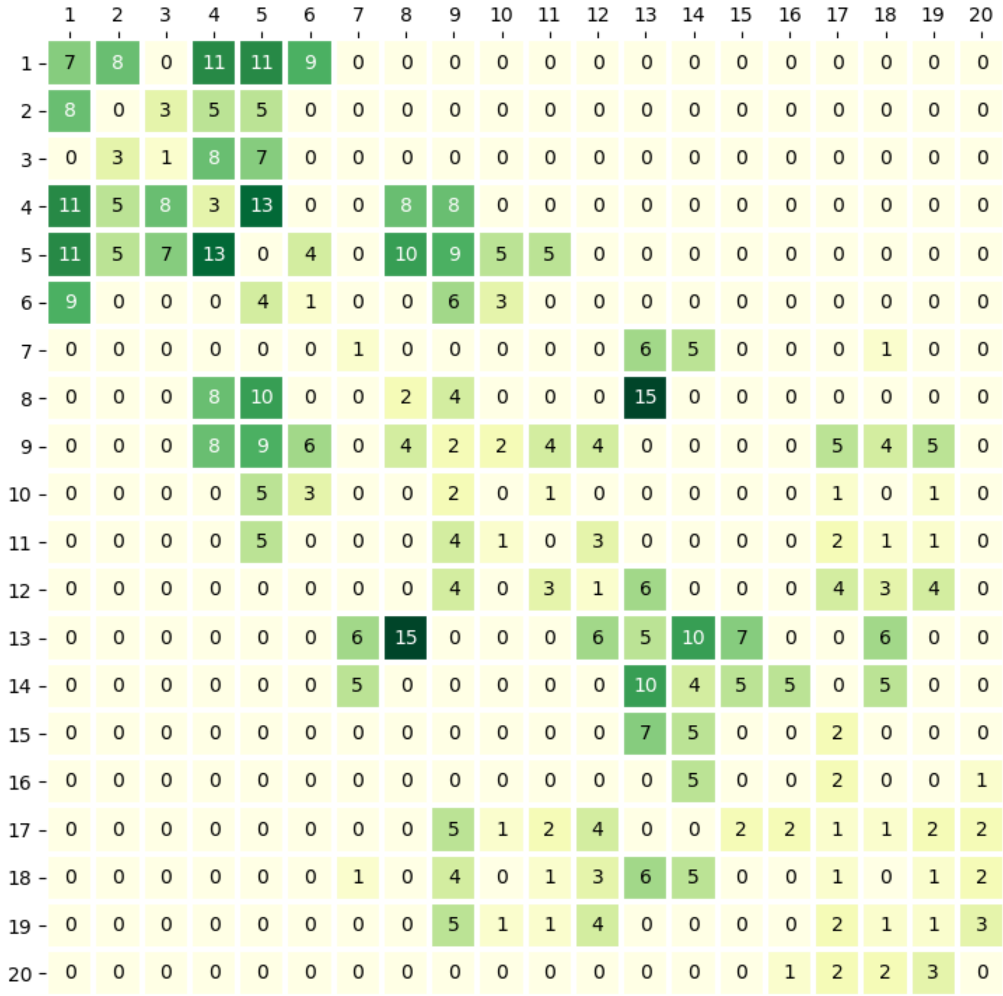

Fig.4 POI counts distribution over node pairs

and 119 links, there is one dummy node and there are two corresponding dummy links because we define arriving at Kiyomizu Temple (either sensor No.15 or 16, the front entrance or back entrance) as the absorbing stage in this study. (The model is further extendable for more destinations.) Connectivity of each node-pair in the reconstructed network is determined by whether there are connections between these sensor-nodes in the full network without passing a third sensor-node. We define all the connected nodes bidirectional connected except for the absorbing node.

Attributes of the links on the reduced network are concluded by the attributes of the associated nodes and links on the full network.

(2) Link attributes

The definition of "link attractiveness" in RL is one of the key problems in modelling. Currently in the reduced network there are three attributes on each link, *length*, *return penalty* and *number of POIs*. *Length* is defined as the shortest path distance between each sensor-node. *Return penalty* is a binary variable that equals one only if the next link is same with current link. Apart from the two widely used attributes, points of interests (POIs) such as shops and sightseeing spots en-route are considered as important decision criteria in tourism area. Thus, we introduce the *number of POIs* as one of the attributes on links. This variable reflects the number of assessable POIs referring to the actual geology between two directly connected sensor-nodes. *Number of POIs* between each sensor-node pair are obtained as follows.

(a) Load POI coordinates of the survey area from OSM (commercial facilities and sightseeing spots only, e.g. shops, restaurants, shrines/temples).

(b) On the full network, associate each POI with its nearest node. If the nearest node only connects to one undirected link, then merge the POI to its neighbor.

(c) For each link on the reduced network, its *number of POIs* adds 1 if the associated node is on the k-shortest path (computed by Yen's k-shortest path algorithm) between sensor-nodes without passing other sensors.

Distribution over node pairs is shown in **Fig.4**. One of the heat areas on the upper left (1-2-3) is Yasaka Shrine, and the second heat area of 8-13-14 includes

Ninei-zaka and Sanei-zaka, sloped streets with plenty of traditional shophouses as an approach to Kiyomizu Temple.

6. RECURSIVE LOGIT MODEL WITH POI COUNTS

Tourists are assumed to make decisions about which link to visit next when they pass any not-absorbing link on the reconstructed network. In the following we describe the data, utility function and results with the recursive logit approach to replicate this behavior.

(1) Data input

For first trial, we limit the data to the first week of the survey. Input of the model is limited to only contain detections from the entrance of Yasaka Shrine (sensor No.1) to the defined destinations, entrances of Kiyomizu Temple (sensor No.15 and 16) to ensure most of the detections and resulted routes belong to walking tourist. Also, a subsequent detection will be ignored either if it is not physically connected to the former detection or not in a reasonable time period (neither too short nor too long). Consequently, there are totally 6349 link choices within 793 extracted routes being input into the model.

(2) Utility function

Given current link k, the instantaneous utility of choosing link a as the subsequent link is:

$$v(a|k) = \beta_L L_a + \beta_{UT} UT_{a|k} + \beta_{NP} NP_a \quad (3)$$

where L_a is length of link a (m/100), $UT_{a|k}$ is a U-turn penalty associated with current link k and subsequent link a, and NP_a is the number of POIs associated with link a. The model is estimated with and without *number of POIs.* There is no significant correlation between variables.

(3) Estimation results and discussion

Estimation results are shown in **Table 1**. All estimated parameters are significant at 0.1%. *Number of POIs* improves the model and shows positive effect to link utility as expected. In both models, the results of β_L are similar, indicating significant negative impact from link length to link utility. Estimation results of U-turn penalty show that β_{UT} gives more disutility to a link once the *number of POIs* is controlled. This is possibly reasonable for walking tourists since their conflict of returning is relatively low on a link that has a variety of POIs.

Table 1 Estimation results with and without *number of POIs*.

	Without NP_a	With NP_a
β_L	-1.49 (-61.47)	-1.55 (-61.56)
β_{UT}	-0.77 (-9.31)	-0.97 (-11.33)
β_{NP}	/	0.04 (10.40)
$LL(\beta)$	-4186.96	-4134.95

7. Conclusions and further work

(1) Conclusions

In this study, we utilize Wi-Fi packet sensors to identify tourist behavior. Three steps are required for the utilization, including the construction of a specific network correspond to Wi-Fi sensors, the extraction of individuals' routes, and model estimation. We construct a simplified network based on real geodata for pedestrian route modelling based on Wi-Fi packet records. We implement the recursive logit model to study tourist's less destination-oriented route choices with introducing *number of POIs* as a variable to reflect link attractions besides distance. This variable is proved significantly effect tourist's route choice and improve the model.

(2) Further work

Firstly, the model will be extended to the entire network. Regarding the route extraction, the choice model will be combined with a clustering method distinguishing the type of pedestrian. We expect to observed different behavioral strategy between tourist and commuters by comparing the model results.

In terms of network construction, further link attributes will be included, such as slope as well as level of crowding. Apart from adding new attributes, a link of "stay" will be introduced to the reconstructed network. This dummy link with same origin and destination node is created for studying the time spent in each link by giving an additional alternative of staying to travelers.

ACKNOWLEDGMENT: This research is a part of the Japan-Israel cooperative scientific research entitled as *Tourist Flow Patterns Identification and Information Provision for Save Evacuation*. We greatly appreciated the support from Japan Science and Technology Agency, JST SICORP Grant Number JPMJSC1805, Japan

REFERENCES

1) 京都市産業観光局観光 MICE 推進室 : 平成 28 年（2016 年）京都観光総合調査, 2017.
2) Fosgerau, M., Frejinger, E. and Karlstrom, A. : A link based

network route choice model with unrestricted choice set. Transportation Research Part B: Methodological, 56, pp.70-80, 2013.
3) Bellman, R. : Dynamic programming. Science, 153(3731), pp.34-37, 1966.
4) Kaneko, N., Oka, H., Chikaraishi, M., Becker, H. and Fukuda, D. : Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-based Recursive Logit Model Featuring Link Awareness. Transportation research procedia, 34, pp.251-258, 2018.
5) Mai, T., Fosgerau, M. and Frejinger, E. : A nested recursive logit model for route choice analysis. Transportation Research Part B: Methodological, 75, pp.100-112, 2015.
6) Oyama, Y. and Hato, E. : A discounted recursive logit model for dynamic gridlock network analysis. Transportation Research Part C: Emerging Technologies, 85, pp.509-527, 2017.
7) 五百藏夏穂, 田中謙大, 中西航, 神谷大介, 菅芳樹, 福田大輔: Wi-Fiパケットセンシングによる沖縄本島観光周遊パターンの特徴抽出, 土木計画学研究・講演集, Vol. 57, 2018.
8) Versichele, M., Neutens, T., Delafontaine, M. and Van de Weghe, N. : The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. Applied Geography, 32(2), pp.208-220, 2012.