Estimation of Queues and Waiting Times at Bus Stops Using Wi-Fi Signal Data

Wi-Fi データによるバス停での待ち行列と待ち時間の推定に関する研究

Hyungsub Jee [*]
ジ　ヒョンソプ

[*]Chair of Transportation Engineering and Management, Intelligent Transport Systems Laboratory

1. Introduction

In cities where bus is a major means of transport such as Kyoto city, Japan, congestion often occurs at bus stops located near major points of interest (POIs) as well as POIs themselves. Traditional methods to find congestion such as counting and video camera data analysis, are expensive and time-consuming methods. There are, recently, several technologies such as automated passenger counting and smart card data analysis. But most transit data sources focus on on-board related events, few data sources to estimate waiting times at stops and boarding demand. Queue data at bus stops can be useful for evaluating the operation performance of a specific stop or a bus route in a bus system. Moreover, if it is possible to obtain the real-time queue information, preventing excessive queues in advance becomes viable by controlling the bus operation status in real-time.

2. Wi-Fi signal data collection and processing

(1) Data collection

Wi-Fi enabled devices (e.g., smartphones) send probe request packets to find networks. These probe request packets contain a timestamp, Media Access Control (MAC) address and RSSI which is signal strength and can be interpreted as an indirect representation of the distance between the Wi-Fi enabled device and the Wi-Fi packet sensor. The functions of Wi-Fi packet sensors are that sensors receive probe request packets from Wi-Fi enabled devices in the vicinity, convert the MAC address contained in the packet by a one-way hash function to anonymize the data, and upload the captured packets to a cloud storage server.

For a MAC address, it is anonymized by the sensor, but sometimes the MAC address is already randomized by a device itself and transmitted with this randomization applied every time. This randomization does not mean that the Wi-Fi packet sensors convert the MAC address by a one-way hash function to anonymize the data.

The data used for this thesis was obtained from several experiments conducted in the area of Ginkakuji and Kiyomizudera in Kyoto city, Japan. Ginkakuji area was chosen because it is an area mainly reachable by bus and slightly away from other tourist places as well as stations of railway lines. Furthermore, the Ginkakuji-mae bus stop is suitable to observe waiting passenger queues, because only two bus lines (32 and 100) serve the stop in one direction and there is one stand at the stop. The numbers of passengers who were making a queue at the bus stop were counted and the line numbers of buses were recorded with the numbers of passengers who board a bus by investigators.

(2) Data processing

The data recorded through the Wi-Fi packet sensors is used to estimate the time each passenger spent in the vicinity of bus stops. Firstly, the entire data can be classified into randomized MAC addresses and non-randomized MAC addresses. Since the record of the same device is tracked by collecting the non-randomized MAC address from data with the same MAC address, the interval is set based on the first and the last observation time. If this interval is longer than or equal to 30 minutes, it is filtered as it is considered as noise.

Furthermore, the average reception period of the probe request packet from the same device was approximately 219 seconds; thus, 100 seconds are each added before and after the interval set based on the first and last observation time considering this reception period. On the contrary, for a randomized MAC address, 100 seconds are added before and after based on the observation time of each data, and all data are assigned a 200-second interval. Subsequently, groups are divided based on the RSSI of each probe request packet.

3. Queues estimation

(1) Multiple linear regression

To estimate the queue at the bus stop, a predictive model is established after performing multiple linear regression analysis using Wi-Fi signal data and bus arrival information processed in the previous part. Multiple linear regression models have length of total queue at a stop at the time as a dependent variable.

A length of queue is expected to increase constantly after the departure of the last bus at bus stops and a cumulative time since the last bus left of each line can be calculated using bus arrival information. At Ginkakuji-mae stop, the upcoming bus arrival information is provided in real time and bus arrival time data was recorded by investigators, then the time since the last bus left is added as a dummy variable in multiple linear regression analysis. Since there is a difference in the length of the queue by time at Ginkakuji-mae stop and such difference demonstrates a similar trend on other days, the time of the day is also added as a dummy variable in multiple linear regression analysis.

Multiple linear regression analysis is conducted using groups of Wi-Fi data as well in the way of different combinations and the adjusted R-squared of best model is 0.676. Figure 1 shows the estimated length of queue by the above multiple regression model and the observed length of queue by investigators.

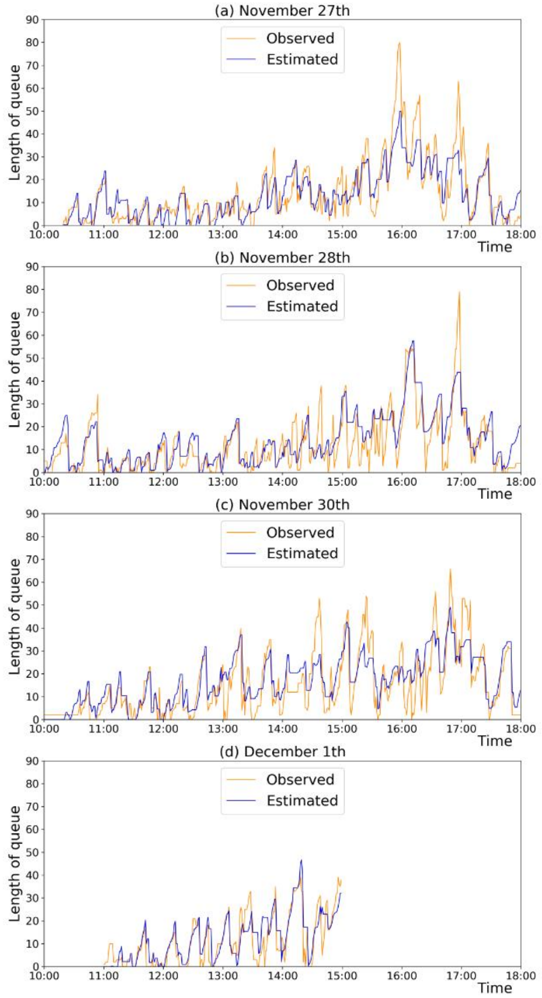

Fig. 1 Estimated length of queue by multiple linear regression

Additional analysis is also conducted to find out whether the using Wi-Fi signal data is effective at Ginkakuji-michi stop and Kiyomizu-michi stop. The results show that the same regression analysis using Wi-Fi signal data is not enough to estimate length of queue at the two stops due to various POIs, a lot of traffic, and many pedestrians around bus stops as well as few observations.

(2) Neural network

Neural network analysis is performed by using the multilayer perceptron regressor. Input and output nodes are same as independent and dependent variables of the above multiple linear regression for Ginkakuji-mae stop and number of hidden layers is 3, which consist of 1,000 neurons at the first hidden layer, 512 neurons at the second hidden layer and 4 neurons at the third hidden layer. All data is split into training dataset and test dataset at a ratio of 8:2 randomly and the model is evaluated using R-squared and Root Mean Square Error by repeating 10 times with different datasets. The model's goodness of fit of the test set has a higher value than that of the regression analysis although there is some overfitting.

4. Waiting times estimation

Waiting times at bus stops mean the time which passengers making a queue spend waiting for a bus serving the stop. The sum of all passengers' waiting time can be a factor in evaluating the performance of a given bus stop. Moreover, when multiple bus lines are running at one stop, it is possible to evaluate the performance of each line if the waiting time for each line can be obtained.

For a length of overall queue at a stop, passenger arrival rates are firstly assumed as time dependent and discrete time period of 1 minute is applied based on type of our experiments data. Then, by assuming that passenger arrival rate is constant over hour of the day, a length of overall queue at a stop at time t_h is derived as Equation (1).

$$Q(t_h) = \sum_i a_i(h)\left(t_h - \check{t}_{i,t_h}\right) \qquad (1)$$

Where,

$Q(t)$: Length of total queue at a stop at time t

$\alpha_i(h)$: Arrival rate for passengers of bus line i in hour h

t_h : Minutes in hour h

$\check{t}_{i,t_h}$: The last departure of a bus of line i at minute t_h

Given a estimated queue at a stop and time since the last bus left from bus arrival information, by linear regression model with these data and utilizing Equation (1), passenger arrival rate of each bus line can be estimated. For a case study, using Ginkakuji-mae stop data, waiting times of bus line 100 and 32 are estimated.

5. Conclusions

In this thesis, firstly, a length of queues can be estimated at bus stops using Wi-Fi signal data. Randomized MAC address data as well as non-randomized MAC address is used and the results show that it contributes analysis in static circumstances such as a queue of bus passengers unlike tracing a flow. Using the previously obtained queue data and derived formulations in this thesis, passenger arrival rates of each bus line are estimated, then the bus line specific waiting times can also be estimated.

The length of queue and bus line specific waiting time are kinds of boarding demand at bus stops. This thesis contributes to the operation planning of bus system due to boarding demand estimation and can be an early and initial step in entire management of bus system.

修士論文指導教員
山田忠史 教授，SCHMÖCKER, Jan-Dirk 准教授