Filter Efficiency Analysis for Extracting Mobile Device Signals to Estimate Bus Passengers Population

Ryo Nishide*

*Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan

*Corresponding Author: ryo-nishide@biwako.shiga-u.ac.jp

Abstract

Congestion in the bus degrades travel experience for passengers who cannot get their seats. Uneven loading of passengers is a critical issue that may intensify the congestion in bus. To alleviate the congestion, it is necessary to balance the population of passengers inside the bus as well as those waiting at the bus stop. As passengers boarding the bus may simply be detected at the entrance door, this paper focuses on methods to analyze the population waiting at the bus stop. Wi-Fi packet sniffer has been developed to detect and record Wi-Fi activities, and the collected data is examined to precisely grasp the pedestrian flows around the bus stop. At the bus stop, unnecessary signals from passing cars and bicycles, surrounding buildings, etc. can also be detected even though they are not related to the potential passengers. To eliminate such signals, filtering parameters must be examined carefully to accurately extract only the bus passengers’ data. Linear regression is used to find the correlation between the filtered data and the ground truth data collected by observation. The results show that the appropriate parameters can be determined using correlation coefficient, which can improve the accuracy better than setting parameters only from observable information, such as distance of bus stop zone and time intervals of arriving buses.

Keywords: Automatic passenger counting, public transportation, congestion, Wi-Fi, mobile sensing.

1. Introduction

Owing to the advancement of mobile technologies, mobile devices such as smartphone or tablet PC have become a convenient and essential equipment that people carry with them in their daily lives. These devices have paved way to work as useful mobile sensors to detect and analyze people’s activities\(^{(1,2)}\).

These mobile sensing technologies have possibilities to take part in Automatic Passenger Counting\(^{(3,4)}\), in which the sensing device automatically detects and counts passengers on public transportation. Generally, camera\(^{(5)}\) or infrared sensor\(^{(6)}\) is used to detect passengers inside the bus. However, as it requires installation in the environment, passengers may feel uncomfortable with camera or sensor in sight. In addition, these devices cannot be installed at an open space with no surrounding attachments nor specific gates. Wireless transmissions have also been focused to detect passenger population and congestion in public transportation in such place as at train station\(^{(7)}\) or inside train cabin\(^{(8)}\), by analyzing data from various sensors or Wi-Fi radio sensing during transmission with passengers’ mobile device.

We have been working on estimating passenger population and congestion at a bus stop using wireless communication of mobile devices. The information can be used together with the population information on the bus so that passengers can take advantage of knowing the seat availability and congestion rate before boarding the bus. Our previous work\(^{(9)}\) employed filtering method to distinguish passenger and non-passenger data. The filtering parameters have been determined based on the RSSI (Received Signal Strength Indication) value which can roughly estimate the distance from bus stop and time intervals of arriving buses. However, the signal strength can vary depending on obstacles or surrounding environment between the devices or the amount of electric power supplied to the Wi-Fi sniffer, and the time duration that the passengers waiting at the bus stop may be longer or shorter than expected as many passengers might rush into the bus or wait for the next bus to come. Thus, the filtering parameter requires adjustment to im-
prove accuracy for extracting only the passenger data.

Based on the parameters obtained in previous work as default settings, we propose method to improve accuracy for distinguishing passenger and non-passenger data. Filtering parameters are investigated by linear regression to discover the correlation with the estimated number of passengers and the ground truth data. Linear regression is employed because the ideal situation to enable accurate estimation should be that the passengers’ population increase linearly according to the increment of the total passengers’ devices, and it can be used as a performance indicator to examine the degree of error by calculating the distance away from an optimal line. Furthermore, filtering parameters have been evaluated by comparing correlation coefficient so that the estimation result can be improved by removing or tuning the inappropriate parameters.

The rest of the paper are as follows. Section 2 briefly describes the method to estimate the number of passengers waiting at the bus stop. Section 3 provides the default filter parameters used in our method and introduces tuning scheme to enhance the parameter values. In Section 4, the evaluation is conducted to prove the efficiency of filtering method and parameter tuning. Section 5 concludes the paper along with future works.

2. Estimation Method of Bus Passengers Population from Mobile Sensing Data

The outline to estimate bus passengers’ population is provided in this section. The method consists of three parts as shown in the top part of Fig. 1. The application scenario examples are shown at the bottom in the figure.

Bus Stop Environment

In Japan, passengers usually line up at the bus stop. The bus stop zone is typically less than 12 meters at most of the bus stops. Wi-Fi sniffer is placed near the bus stop to capture Wi-Fi frames sent from passengers’ mobile devices. While capturing the data, an observer counts the number of passengers at the bus stop to obtain ground-truth data. An adequate number of ground truth data is required to accurately analyze the correlation with estimated results. For deployment in practical scenario, the observers are not needed if enough data has been collected to estimate the number of passengers accurately.

Data Filtering

The Wi-Fi sniffer generates detection log, which records the MAC address of detected device, the first and last detected time, the average signal strength (RSSI) during transmission, and the number of received packets. As the Wi-Fi sniffer collects all Wi-Fi data near the bus stop, the data must be classified to passenger and non-passenger data. Non-passenger data includes signals sent from passing pedestrians, bicycles and automobiles, or static access points (AP) from buildings. Fig. 1 shows an example of three data with red X considered as non-passenger data. The ones with low RSSI may be from distant devices, the long duration may be a static AP or a device of someone staying for different purposes, and the short duration may be someone with a device just passing by the bus stop. These data can be considered as noise (false data) and must be excluded to extract only the passenger data. The details of setting the adequate filtering parameters are discussed in the following section.

Figure 1. Overview of Proposed Method.


- Observer Collection
- Data Filtering
- Linear Regression
- Estimation

Application Scenario Examples

1. Remote passengers
2. Bus Company

Decide if the passenger is a passenger or non-passenger.

User Population Change

Decision Making

Go or wait for later on?

Fig. 1. Overview of Proposed Method.
son, we focus on the total devices of all passengers in a bus stop queue. We believe that the total of passengers’ devices correlates with the passengers’ population, especially if the passengers’ types or destination are common, in such cases as the school bus for students. Thus, if the regression line is obtained, the number of passengers may be estimated accordingly. However, the accuracy of estimated results depends on how close the data points are to the regression line, which extremely depends on how much the noises can be removed by the filter setting. Therefore, the filter parameters require tuning to improve accuracy in estimation.

[Application Scenario Examples]

We believe that our method can provide useful insights for individual bus passengers, drivers and companies. For example, remote passengers can schedule the time to depart to the bus stop if the congestion information can be obtained by their smartphones. Drivers can communicate with each other to balance the number of passengers, and the company can modify the timetable based on the daily or weekly congested time.

3. Filtering Method

3.1 Default Filtering Parameters

Bus stops are public areas with many people in the neighborhood. Wi-Fi signals from various types of devices are transmitted in every direction, and not all signals are transmitted from the passenger’s devices. To separate passenger signals from non-passenger signals, we filter out the non-passer signals from the collected Wi-Fi device activity. Our filter mechanism uses a combination of RSSI, packet frequency, and the total time duration (from the first to the last packet) which the nearby device is detected.

RSSI is used for filtering according to distance between the Wi-Fi sniffer and passengers’ devices. As the received signal strength depends on the distance between the receiver and device, the signal strength is weaker if the device is further away, and stronger if the device is close-by. The parameter is set at -79 dBm as it corresponds to the typical bus stop zone which is 12 meters long.

The number of received packets is then examined to reject passers-by devices or remote devices detected by coincidences. Fewer than two packets indicate that the device has been detected only once, and not continuously at the bus stop.

For detected time duration, two limits are set. The lower bound is set as one minute, as less time duration is too short to be identified as passengers. Those who rush into the arrived bus are not considered in this case, as it is difficult to distinguish them from passers-by. The top bound is set at 30 minutes, as the devices staying there too long cannot be regarded as passengers. The top bound can also be determined from the bus timetable, which is obvious in this case that the buses arrive in less than 30 minutes after the previous bus has left.

Fig. 2 shows the data flow used to filter passenger device data in the collected raw log data. The thresholds are determined with an adequate value stated in Fig. 2 to extract an estimate of the number of devices which belong to passengers waiting at the bus stop.

3.2 Parameter Tuning

The threshold value for each filtering condition is defined based on the preliminary investigation(9) in Section 3.1 in which the threshold for the RSSI corresponds to the length of bus stop zone, and the time frame between the arrival of two continuous buses referring to the published information of the bus timetable, etc. These parameters may be useful to determine the default setting for the experiment; however, we believe that these parameters can be modified if the correlation of the number of potential passengers’ terminals and the ground truth data is obtained.

The correlation of RSSI and distance may be slightly biased depending on the noises in surrounding environments such as walls, obstacles and natural environments such as humidity, temperature, etc. which may affect the radio transmission between devices, as well as the amount of power supplied to devices. The bus arrival may also be delayed in some cases, or many passengers may remain waiting for the next bus without boarding on the congested bus. Furthermore, each bus stop has different characteristics depending on the surrounding objects and environments, or the types of majority passengers carrying multiple or none of the devices, depending on their age, gender, occupation, etc. Therefore, the combination of filtering conditions or parameter values cannot be determined uniformly with default parameters. Thus, the filtering parameter may require tuning so that the number of filtered devices correlate to the

![Fig. 2. Process for Filtering Data.](image)

265
number of passengers for each bus stop.

Data for the past several days of the manually counted number of bus passengers and the estimation result data using the default filtering parameters must be prepared before tuning the parameter. Then, the linear regression equation and the correlation coefficient is obtained from these two data. Correlation coefficient expresses the degree that the data fits a linear model. The more the data fits a linear model, the more accurate the estimation outcome would be.

To perform tuning, several parameters higher or lower than the default parameter are investigated. Filtering is performed individually for each parameter, and then the linear regression and correlation coefficients are calculated. Among all parameters, the one with highest correlation coefficient can be considered as the optimal filtering parameter value.

4. Experiment

4.1 Bus Stop Environment and Experimental Device

To examine the usability of the four parameters in Fig. 2, we have collected data at two bus stops with different characteristics. The first one is the crowded bus terminal in the university campus, as shown in Fig. 3(a). The passengers are mostly university students and staffs. The road is inside the campus, and thus privately used only by the buses and some vehicles for campus use. On the other hand, the sidewalks near the bus stop are used frequently by pedestrians as it is the only path to enter or leave the campus.

Another one is the bus stop at the uncrowded suburban environment close to the industrial and residential area, as shown in Fig. 3(b). Most of the passengers at the bus stop are industry workers, but it is less used since the bus is sometimes crowded during daytime, and some workers prefer to walk. The road is sometimes congested with automobiles and pedestrians waiting at the traffic intersection. Note that some of the buses stop at both bus stop in Fig. 3(a) and 3(b), and passengers cannot board on the bus at suburban bus stop because it is congested with university students from the bus terminal.

Raspberry Pi is used together with a USB adapter capable of monitor mode for this experiment. Monitor mode is the Wi-Fi equivalent of promiscuous mode used in wired interfaces, which can capture all received packets disregarding their destination address. Airodump-ng is used to capture packets similarly to tcpdump but ignores massive multicast traffic sent from base stations.

4.2 Method for Efficiency Analysis

Experiment has been conducted for 9 days at crowded bus terminal, and 7 days for uncrowded suburban environment, approximately 1 hour on each day. Using those data, linear regression equation is obtained from the passengers counts by manual observation and by estimation with our method. The efficiency of the method is examined by comparing the correlation coefficient of data filtering results with different parameter settings. Meanwhile, using the linear regression equation, the number of passengers is estimated from the number of devices denoted as $P_i^{estimate}$, then the error, of $P_i^{estimate}$ is calculated from the observed passenger counts $P_i^{actual}$ using equation (1), and the average error rate is obtained (MAPE: Mean Absolute Percentage Error).

$$\text{error}_i = \frac{|P_i^{estimate} - P_i^{actual}|}{P_i^{actual}} \times 100$$

4.3 Efficiency Evaluation of Filtering Method

To verify the efficiency of filtering method, linear regression equation and correlation coefficients are obtained. The equation is then used to estimate the number of bus passengers, and the result is compared with the ground-truth data which is the number of passengers counted by manual observation.

Crowded Bus Terminal

In Fig. 4, the regression line is generated for filtered and non-filtered number of devices, and linear regression equation and correlation coefficient ($r^2$) are calculated. The x-axis is the number of detected devices in total for “Non-Filtered” and number of extracted devices by filtering process for “Filtered”, and the y-axis is the observed number of people as ground-truth data. The following equation has been obtained with each method accordingly:

Non-Filtered:

$$y = 0.7546x - 137.56$$

Filtered:
Fig. 4. Linear Regression on Crowded Bus Terminal.

![Graph showing linear regression for crowded bus terminal](image)

Table 1. Passengers Estimation on Bus Terminal.

<table>
<thead>
<tr>
<th>observed</th>
<th>non-filtered</th>
<th>estimate</th>
<th>error</th>
<th>filtered</th>
<th>estimate</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>670</td>
<td>1105</td>
<td>696</td>
<td>3.92</td>
<td>262</td>
<td>646</td>
<td>3.64</td>
</tr>
<tr>
<td>583</td>
<td>935</td>
<td>568</td>
<td>2.57</td>
<td>232</td>
<td>575</td>
<td>1.30</td>
</tr>
<tr>
<td>668</td>
<td>1111</td>
<td>701</td>
<td>4.91</td>
<td>270</td>
<td>664</td>
<td>0.56</td>
</tr>
<tr>
<td>719</td>
<td>1042</td>
<td>649</td>
<td>9.77</td>
<td>288</td>
<td>706</td>
<td>1.75</td>
</tr>
<tr>
<td>733</td>
<td>992</td>
<td>611</td>
<td>16.64</td>
<td>308</td>
<td>753</td>
<td>2.75</td>
</tr>
<tr>
<td>464</td>
<td>778</td>
<td>450</td>
<td>3.12</td>
<td>184</td>
<td>463</td>
<td>0.18</td>
</tr>
<tr>
<td>517</td>
<td>922</td>
<td>558</td>
<td>7.97</td>
<td>211</td>
<td>526</td>
<td>1.80</td>
</tr>
<tr>
<td>479</td>
<td>953</td>
<td>582</td>
<td>21.41</td>
<td>220</td>
<td>547</td>
<td>14.27</td>
</tr>
<tr>
<td>575</td>
<td>969</td>
<td>594</td>
<td>3.24</td>
<td>211</td>
<td>526</td>
<td>8.47</td>
</tr>
<tr>
<td>MAPE</td>
<td>8.17</td>
<td>MAPE</td>
<td>3.86</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ y = 2.3387x - 32.84 \]  \hspace{1cm} (3)

The correlation coefficient is 0.5798 for non-filtered and 0.8984 for filtered.

Next, equations (2) and (3) are derived with the number of devices individually to estimate the number of passengers in Table 1. “Observed” is the ground-truth data which we aim to target by estimation, and the number of detected devices and filtered devices are named as “non-filtered” and “filtered” respectively. “Estimate” is the estimated number of passengers using equations (2) and (3). “Error” is the error rate (error) in equation (1) expressing the difference between the estimated number and observed number of passengers, and MAPE is the average of error rates. The MAPE is 8.17% for “non-filtered” and 3.86% for “filtered”.

Uncrowded Suburban Bus Stop

For analysis in suburban bus stop in Fig. 5, the following equation has been obtained.

**Non-Filtered:**

\[ y = 0.2105x - 213.38 \]  \hspace{1cm} (4)

**Filtered:**

\[ y = 1.8836x - 221.52 \]  \hspace{1cm} (5)

The correlation coefficient is 0.1926 for non-filtered and 0.5272 for filtered. MAPE has been calculated in Table 2 in the same way as Table 1, which is 15.08% for non-filtered, and 11.00% for filtered.

**Discussion**

Due to the analysis, Fig. 4 and 5 show that each data points are closely gathered to a regression line for filtered data, and sparse on non-filtered data. As the correlation coefficient (R²) expresses how sparse the data point is to the regression line (in which the coefficient closer to 1 implies that it has higher correlation as the data points are closer to the line), the filtered data has significant results than non-filtered data for both crowded bus terminal and suburban bus stop cases, which indicates that the filtering method can extract data to correlate with the number of passengers with higher probability than non-filtering method, and that it may also increase the estimation accuracy for the number of passengers.

In Table 1 and 2, both cases show that the estimated number of passengers are significant with filtered data rather than the non-filtered data, and the error rates are smaller. Therefore, the effectiveness for applying filtering method has been revealed.

**4.4 Analysis on Individual Filtering Condition**

Individual parameters are investigated to verify the accuracy of filtering process and adequacy of the parameter...
values. Some filtering conditions may not be crucial as they may only remove a few non-passerger data. It may also incorrectly delete the passenger data if the parameter value is inappropriate or the filter condition itself adversely affects the outcome to hinder from extracting the passenger data correctly. To analyze the efficiency of filtering conditions and parameters, linear regression is performed individually for each condition. The correlation coefficient and MAPE are compared with results on non-filtering or filtering with all four conditions. Table 3 shows the correlation coefficient and MAPE of four individual filtering parameters for bus terminal, and Table 4 for suburban bus stop.

**Discussion**

In Table 3, filtering condition using RSSI value has the highest correlation and lowest MAPE in comparison with the other filtering conditions, which means that RSSI is an indispensable condition to obtain an accurate estimation. In contrast, the number of packets and long-term detection seem to deteriorate the estimation results. Moreover, the correlation coefficient for short-term detection seems high, while the error rate is also high, which represents that the number of devices has a high correlation, but this parameter cannot be used itself for estimation.

Table 4 shows that filtering with the number of packets worsens both the correlation coefficient and MAPE. Thus, the number of packets may not be efficient for suburban bus stop data. On the other hand, the RSSI and long-term detection seem to improve the results significantly by both the correlation coefficient and MAPE.

**4.5 Analysis on Multiple Filtering Conditions**

In this section, setting the results of using all four parameters as a goal, we verify how much each filter parameter contributes to attain the goal. We have performed remove-one filter analysis, which is conducted by removing one filter condition while using the other three conditions, to compare with the results of when all four conditions are used. That is, some filtering conditions may be useless because the other conditions may remove all or most of the non-passenger data. Thus, we aim to identify such useless parameters, considering ways to improve the results. “No Removal” refers to using all four conditions (without any of them removed), the same as “All Filter” in Table 3 and 4.

**Discussion**

Table 5 shows the result for crowded bus terminal, and Table 6 for suburban bus stop. Both results have similar characteristics when the RSSI filtering condition is removed, as the correlation coefficient is low and MAPE is high, which have been the worst results compared to the other three individual conditions. This has proved that fil-

<table>
<thead>
<tr>
<th>Filters</th>
<th>Corr. Coef. (R^2)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Filter</td>
<td>0.5798</td>
<td>8.17</td>
</tr>
<tr>
<td>RSSI (≤-79)</td>
<td>0.7372</td>
<td>7.84</td>
</tr>
<tr>
<td>Packets (&lt;2)</td>
<td>0.6185</td>
<td>8.43</td>
</tr>
<tr>
<td>Time (≥30min.)</td>
<td>0.6163</td>
<td>7.96</td>
</tr>
<tr>
<td>Time (≤1min.)</td>
<td>0.7087</td>
<td>9.46</td>
</tr>
<tr>
<td>All Filter</td>
<td>0.8984</td>
<td>3.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Filters</th>
<th>Corr. Coef. (R^2)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Filter</td>
<td>0.1926</td>
<td>15.07</td>
</tr>
<tr>
<td>RSSI (≤-79)</td>
<td>0.4099</td>
<td>12.39</td>
</tr>
<tr>
<td>Packets (&lt;2)</td>
<td>0.1778</td>
<td>15.25</td>
</tr>
<tr>
<td>Time (≥30min.)</td>
<td>0.3178</td>
<td>12.19</td>
</tr>
<tr>
<td>Time (≤1min.)</td>
<td>0.1597</td>
<td>14.60</td>
</tr>
<tr>
<td>All Filter</td>
<td>0.5272</td>
<td>11.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Removed Filter</th>
<th>Corr. Coef. (R^2)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No RSSI (&lt;-79)</td>
<td>0.7708</td>
<td>6.69</td>
</tr>
<tr>
<td>No Packets (&lt;2)</td>
<td>0.8984</td>
<td>3.86</td>
</tr>
<tr>
<td>No Time (≥30min.)</td>
<td>0.8962</td>
<td>3.73</td>
</tr>
<tr>
<td>No Time (≤1min.)</td>
<td>0.7507</td>
<td>7.62</td>
</tr>
<tr>
<td>No Removal</td>
<td>0.8984</td>
<td>3.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Removed Filter</th>
<th>Corr. Coef. (R^2)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No RSSI (&lt;-79)</td>
<td>0.2665</td>
<td>14.60</td>
</tr>
<tr>
<td>No Packets (&lt;2)</td>
<td>0.5272</td>
<td>11.05</td>
</tr>
<tr>
<td>No Time (≥30min.)</td>
<td>0.5303</td>
<td>11.18</td>
</tr>
<tr>
<td>No Time (≤1min.)</td>
<td>0.4106</td>
<td>12.36</td>
</tr>
<tr>
<td>No Removal</td>
<td>0.5272</td>
<td>11.00</td>
</tr>
</tbody>
</table>

From both experiments, it has been revealed that RSSI is the most important filtering parameter among the four parameters which contribute to extracting the accurate estimation result. High correlation coefficient and low MAPE suggest that the filtering condition is essential, but the low correlation coefficient and/or high MAPE imply that the filtering condition is useless, or the parameter value is inappropriate. Thus, before making decisions to remove such questionable conditions, it is necessary to examine if the parameter tuning would improve the estimation result.

Table 5. Remove-One Filter Analysis on Bus Terminal.

Table 6. Remove-One Filter Analysis on Bus Stop.
tering condition on RSSI contributes significantly in extracting the passenger data or estimating the number of passengers if it is included. In contrast, “No Packets” and “No Removal” have not much difference in correlation coefficient and MAPE, which is clear that these parameters do not contribute much in extracting the passenger data or estimating the number of passengers. This is because the same non-passenger data is removed with the other three filtering conditions, especially from time duration of less than or equal to 1 minute [Time (≤1 min.)], and thus, this can be removed in both types of bus stop.

Another consideration is for time duration of greater than or equal to 30 minutes [Time (≥30 min.)], in which the correlation coefficient and MAPE are not good. However, the result is slightly different from applying all four parameters, which indicates that at least a few non-passenger data is removed. In this case, we believe that the filtering parameter value is inadequate, and might have possibility to improve the result if the parameter is modified.

4.6 Parameter Tuning

We have conducted analysis to tune the four parameters. The following range has been investigated; RSSI (-71 to -90 dBm) and upper bound of time duration (5 to 70 min., for every 5 minutes). The lower bound of time duration was also investigated, but the optimum result was 1 min. for any cases. The number of packets has not been investigated because this filtering feature overlaps with time duration for less than 1 minute, as discussed in Section 4.5. Thus, we focus on changing the RSSI and upper bound of time duration (denoted as long duration) parameters.

We have focused on the following three methods to tune the parameters.

1. Apply the best correlated individual filtering parameters simultaneously
2. Apply the best correlated filtering parameters using remove-one filter, in long duration then RSSI order
3. Apply the best correlated filtering parameters using remove-one filter, in RSSI then long duration order

In Tables 7 and 8, “Mtd” representing three types of method, “RSSI”, “Long”, “Short” and “Packet” are shown for the best parameter value under four filtering conditions obtained for each method, and the correlation coefficient (Corr.) and MAPE are calculated. In Tables 7 and 8, the correlation coefficient and MAPE of methods (2) and (3) are similar, but better than method (1). This result indicates that multiple filtering conditions (Remove-One Filter) should be considered rather than the single condition when tuning the parameters, as the filtering conditions affect each other when the parameters are modified.

Table 7. Tuning result for Crowded Bus Terminal.

<table>
<thead>
<tr>
<th>Mtd</th>
<th>RSSI</th>
<th>Long</th>
<th>Short</th>
<th>Packet</th>
<th>Corr.</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-82</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>0.8752</td>
<td>4.70</td>
</tr>
<tr>
<td>(2)</td>
<td>-83</td>
<td>70</td>
<td>1</td>
<td>1</td>
<td>0.9374</td>
<td>3.70</td>
</tr>
<tr>
<td>(3)</td>
<td>-83</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>0.9380</td>
<td>3.68</td>
</tr>
</tbody>
</table>

Table 8. Tuning Result for Suburban Bus Stop.

<table>
<thead>
<tr>
<th>Mtd</th>
<th>RSSI</th>
<th>Long</th>
<th>Short</th>
<th>Packet</th>
<th>Corr.</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-74</td>
<td>45</td>
<td>1</td>
<td>1</td>
<td>0.7467</td>
<td>7.62</td>
</tr>
<tr>
<td>(2)</td>
<td>-74</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>0.7459</td>
<td>7.34</td>
</tr>
<tr>
<td>(3)</td>
<td>-74</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>0.7459</td>
<td>7.34</td>
</tr>
</tbody>
</table>

Discussion

Our experiment has shown that the estimation accuracy can be improved by tuning the parameters, as the correlation coefficient with the number of filtered devices and the observed number of passengers has increased significantly, and the error for the estimation of the number of passengers (MAPE) has been reduced. The case at bus terminal had greater accuracy because noises such as non-passenger signals inside the campus was easy to be removed.

The optimal RSSI for university bus terminal was -83 dBm as shown in Table 7 because the queue extends beyond the bus stop area, especially after the classes. On the other hand, the optimal RSSI for suburban bus stop was -74 dBm in Table 8, probably because vehicles are jammed at traffic intersection, where the line of vehicles extends to the bus stop area. In such cases, the RSSI value must be set as low as possible. Fortunately, not many bus passengers use this bus stop, so it was not a big deal in setting the RSSI.
values corresponding to less than 12 meters.

There were not many differences in terms of correlation coefficient and MAPE for most of the parameters of the long (upper bound) time duration, but if the upper bound was too low, the estimation accuracy was drastically lowered. In other words, the accuracy rate increased in a log curve according to the increase of duration time. However, depending on the location of experiment, especially at metropolitan bus stop with many access points, the upper bound setting may be critical. On the other hand, short duration (lower bound) is set to filter out many packets detected only once (in a short time duration), so there should be no other optimal values than 1.

Four filtering conditions have been investigated in this experiment, but it is necessary to consider more filtering conditions to accurately distinguish passenger signals from non-passenger signals. Furthermore, much more data collection is required to apply our method.

5. Conclusions

In this paper, we introduced linear regression analysis to estimate the number of passengers by monitoring Wi-Fi signals detected at bus stops. Based on our previous work\(^{(9)}\), the paper focused on modifying the filter parameters from the default settings. Linear regression was used to examine the best parameter to enhance the accuracy in estimation.

For future works, more data is required to evaluate the usability of the system. Our experiments were conducted with hourly basis for several days at two bus stop environments, and we are still on our way to collect more data to extend the research. For other bus stop environments, in such locations as train stations and airports, or congested downtown areas, more filtering parameters are to be explored. Linear regression was introduced in this paper, but there may be other ways to accurately estimate passengers’ population using machine learning. The counting passengers by the observer is another issue to be concerned. If the number of passengers inside the bus can be obtained using camera images\(^{(3)}\), such data may be used as a ground-truth data.

The data is collected for one hour in each day of experiment, but acquiring real-time passenger population may considerably increase usability to provide current congestion information. Referring to other sensory data or analyzing the trend of population change may also be necessary to improve the estimation accuracy in order to explore the efficient system.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17K18236. I would appreciate warm supports and valuable suggestions from Prof. Hideyuki Takada, Prof. Ian Piumarta, and Dr. Thongtat Oransirikul of my former research group at Ritsumeikan University.

References

(2) Vittorio Loreto, Muki Haklay, Andreas Hothis, Vito D.P. Servedio, Gerd Steunne, Jan Theunis, and Francesca Tria: “Participatory Sensing, Opinions and Collective Awareness”, Springer International Publishing Switzerland, 2017