Identifying the Passenger Travel Behavior of the Metro System Using Wi-Fi Information

Xinyu Zhang
School of Highway, Chang'an University
Chang'an University, Xi’an, China
619588854@qq.com

Yan Li
School of Highway, Chang'an University
Chang’an University, Xi’an, China
2017121153@chd.edu.cn

Fu Chen
School of Highway, Chang’an University
Chang’an University, Xi’an, China
396120858@qq.com

Abstract

In order to accurately identify the passenger travel characteristics in urban rail transit with the network operation environment, a distributed traffic behavior identification system with Wi-Fi information is designed. The information detection technology can effectively identify the passenger travel behavior in urban rail transit, avoiding a great deviation between the deductive travel behavior and the actual travel behavior. This paper utilizes Wi-Fi information detection technology to obtain the passenger travel information on urban rail transit stations. The travel characteristics, such as travel time distribution and travel route choice inferred from the Wi-Fi information detection data is shown in field experiment in Xi'an Urban Rail Transit to be consistent with those inferred from AFC data. The result shows travel characteristics can be identified by Wi-Fi information technology with relatively high accuracy, Sampling rate and matching degree, which might provide some basic evidence for improving metro operation management and fare clearing.

KEYWORDS: Wi-Fi information; urban rail transit; travel characteristics; travel time distribution

1 Introduction

Passenger route guidance, operation management and fare clearing all depend on the accurate grasp of the passenger travel route in urban rail transit system (Lai et al., 2013). At present, the research on travel route choice mainly focuses on the prediction methods of passenger flow distribution based on the principle of user equilibrium (Wu and Liu, 2004), proportion estimation method based on AFC data (Shi et al., 2015 & Xu et al., 2018), route choice based on logit model (Zhang et al., 2016) and probability model based on normal distribution (Wei et al., 2009 & Yang et al., 2009), however these prediction accuracy isn't relatively high. Therefore, it is urgent to establish a method that can accurately identify passenger travel characteristics, and provide some evidence for improving operation management in urban rail transit system.

Modern traffic information technology can accurately obtain the location and time stamp of passengers, which provides a new way to accurately identify passenger travel characteristics. GPS devices can be used to detect traffic conditions (HOFLEITNER A et al., 2012), Bluetooth devices can match vehicle ID (LIP et al., 2016) or match Mac addresses to predict travel time (RICHARDSON J
K et al., 2011 & BAKULA C et al., 2012 & HAINEN A M et al., 2011). Mobile phone positioning technology (QIU Z et al., 2009) can also obtain information such as vehicle travel time and road operation condition. However, the GPS information cannot be obtained in the underground space, and the sampling rate of the road Bluetooth device is about 1-3% (BAKULA C et al., 2012), and the positioning accuracy is low. Although the mobile phone signalling data can be applied to the identification of subway passengers travel route (Lai et al., 2013), it contains privacy information and is difficult to obtain, thus it is inconvenient to be applied to the identification of passenger travel characteristics. With the widespread use of Wi-Fi mobile devices, it can be used as an effective source of data information (FIGUERA C et al., 2011) to identify passengers travel characteristics. By developing a detecting device to obtain the unique MAC (Media Access Control) information of the passenger Wi-Fi device, the travel characteristics can be accurately identified. This method does not need a Wi-Fi device to establish a connection with the detecting device, and has the advantages of high speed, high security, low cost, high precision, high sampling rate, strong compatibility and expandability (DU P et al., 2011), and is an ideal technology for identifying passenger travel characteristics in urban rail transit system.

Based on previous research, this paper utilizes Wi-Fi information detection technology to detect the mobile phone Wi-Fi information with the unique MAC address on underground station, identify the travel characteristics by matching and analysing the detection data of each station, and compare with the AFC data to assess the reliability of the Wi-Fi information detection technology.

The rest of the paper is organized in three sections. Section 2 introduce the passenger travel route identification technology, including the basic principle of WI-FI information detection technology and Wi-Fi information data returned by device. Section 3 deals with the identification of passenger travel characteristics including (i) time distribution of platform passenger flow at each subway station, and (ii) passenger travel route choice between stations. In Section 4, Conducting a field experiment in Xi'an Urban Rail Transit to test and verify Wi-Fi information and compare to AFC data. The conclusion follows.

2 Passenger Travel Route Identification Technology

2.1 The basic principle of Wi-Fi information detection technology

The basic structure of WLAN is point-to-point or point-to-multipoint communication between two wireless devices. The information detection system is mainly composed of embedded arithmetic programming to realize software function and hardware equipment to build supporting equipment. The design of the information detection system uses Wi-Fi broadcasting technology. Its principle is that the broadcasting is transmitted through Wi-Fi as a medium, so that the device scanning the Wi-Fi signal can receive the broadcasting content, and only the holder of the Wi-Fi can send the broadcasting.

The mobile device in the wireless local area network search mode uploads the Mac address inherent to the mobile device to the central data platform through the traffic behavior detecting device by searching for the wireless communication device, as shown in Fig.1. The process for the detecting device to obtain distributed travel information is as follows:

1) The mobile device turns on the wireless local area network search mode and does not connect to Wi-Fi;
2) The information detecting device is powered on and in a working state;
3) The information detecting device is connected to the Cloud Server;
4) The mobile device is within the detection radiation range of the information detecting device.
Fig. 1. Data transmission process of information detecting device

2.2 Wi-Fi information data

The Wi-Fi information data returned by the Wi-Fi information detecting device to the Cloud Server mainly includes: Date, Time stamp, Rssi (Received Signal Strength Indication), Mac address, and detecting device name. The information is shown in Table 1.

<table>
<thead>
<tr>
<th>date</th>
<th>time</th>
<th>Rssi</th>
<th>MAC address</th>
<th>device name</th>
</tr>
</thead>
</table>

The Rssi is related to the distance between the mobile device and the information detecting device. After testing and analysis, the relationship between the Rssi and the distance is as shown in equation (1).

\[
y = \begin{cases} 
0.0274x^2 + 3.2649x + 101.49, & x \in (-95, -55) \\
0.0571e^{-0.07x}, & x \in (-55, -40) 
\end{cases}
\]  

(1)

In equation (1), \( y \) is the distance between the mobile device and the Wi-Fi information detecting device, and \( x \) is the Rssi from the mobile device.

3 Identification of Passenger Travel Characteristics

In this section, to work on the identification of passenger travel characteristics, the detection data computed from the detecting devices is analysed in two parts: the time distribution of passenger flow on each station and the travel route choice between the two stations.

3.1 Time distribution of passenger flow

By analysing the data detected at the stations, the time distribution of passenger flow throughout the day and the distribution of passenger positions within the station can be obtained. The detecting device collects Wi-Fi information data one time every 30 seconds, and the passengers stay at the station for different length of time. Therefore, the device will collect a large number of Wi-Fi information data with repeated Mac addresses. Therefore, it is necessary to eliminate the data of repeated Mac addresses to analyse the time distribution of passenger flow throughout the day.

Data set is denoted \( X, X = \{x_{ij} | i = 1, 2, \ldots, l, j = 1, 2, \ldots, k\} \), where \( i \) is the Station, \( l \) is the total station, \( j \) is the detecting device of the deployed at the current station. The Wi-Fi information returned by the detecting device \( j \) on station \( i \) is denoted by \( x_{ij} = (s, m, t) \), where \( s \) is Rssi, \( m \) is MAC address, \( t \) is the time stamp (the instant when the data is detected). For rapid cleaning of large amounts of repeated and redundant data, the pseudo code for filtering valid data is written which is as follows. The time distribution of passenger flow on each station can be obtained by performing frequency statistics on the cleaned data set.

Begin
Read \( x_{ij}(t) \);
for \( i \leftarrow 1 \) to \( m \)
s ← detect
for h ← t -1800 to t
  if (m(t)=m(h))
    then h ← h + 1;
  else s ← s + 1;
  while (s=1800)
    add x_i(t) to X;
  i ← i + 1;
return X;
End

3.2 Travel route choice between stations

3.2.1 Travel time between two stations
The total travel time spent by passenger’s travelling from one station to another includes travel
time, waiting time, transfer time, etc. The travel time between the two stations can be calculated by
equation (2), Where \( t_a \) is the total travel time, \( l \) is the distance between the two stations, \( v \) is the travel
average speed of train, \( a \) is the number of stations between two stations, \( b \) is transfer times of
passenger, \( t_t \) is the transfer time, \( t_d \) is the dwell time, \( t_w \) is the waiting time.

\[
(t_a) = t_u + t_v + t_t = t_u + \frac{l}{v} + (a - b) \cdot t_d + b \cdot t_t
\]  

(2)

Previous research has proved transfer time \( (t_t) \) of metro station fit the lognormal distribution, as
shown in equation (3).

\[
f(t_t; \mu, \sigma) = \frac{1}{t_t \sqrt{2\pi} \sigma} e^{-\frac{(\ln t_t - \mu)^2}{2\sigma^2}}
\]  

(3)

In equation (3), \( \mu \) is the mean value of transfer time, and \( \sigma \) is variance of transfer time.

3.2.2 Travel route choice
According to the calculation formula of travel time shown in equation (2), the minimum \( t_{min} \) and
maximum \( t_{max} \) of travel time between the two stations can be calculated. Assume that train
departure interval time is \( \tau \), the transfer time \( (t_t) \) is generally between 0 and 10 minutes. When
passengers may fail to board the current train due to the inadequate capacity of the train coach, \( 0 < t_t < 10 + 2 \tau \). When there are transfer times \( b \) between two stations, the \( t_{min} \) and \( t_{max} \) are calculated by
equation (4).

\[
\begin{align*}
t_{max} &= \frac{l}{v} + 10b + (\tau - t_d) \cdot b + (a - b) \cdot t_d = \frac{l}{v} + (10 + \tau - 2t_d)b + at_d \\
t_{min} &= \frac{l}{v} + (a - b) \cdot t_d
\end{align*}
\]  

(4)

When the travel time between the two stations computed from detection data is in the range of
\( (t_{min}, t_{max}) \), the travel time is reliable, that is to say the travel route through the corresponding two
stations is accurately matched. The number of passengers who chose this travel route can be counted
by the matched data of travel route.

4 Xi’an Subway Passenger Travel Route Detection
In order to evaluate the accuracy of the passenger travel characteristics inferred by detection
data, a field experiment have been conducted in Xi’an Urban Rail Transit Line 1 and Line 2 on
October 27, 2017, and compare to AFC data provided by Xi’an Urban Rail Transit Operation
Company. In this experiment, the detecting devices are installed on the platform floors of WuLukou
Station and SaJinQiao Station in Line 1, ZhongLou Station and AnYuanMen Station in Line 2, and
installing more than two detecting devices at each station to ensure detecting range covers the entire
station platform. BeiDaJie Station is a transfer station for Line 1 and Line 2. Installation layout of
detecting device installation layout is as shown in Fig.2.

Fig. 2. Installation layout of detecting device

4.1 AFC data

According to the OD table data computed by AFC data, the cross-section passenger flows on
the ZhongLou Station, AnYuanMen Station, SaJinQiao Station and WuLuKou Station, and the
number of Passengers choosing the travel route which contains these four stations are counted. The
passenger flows in each direction is shown in Fig.3.

Fig. 3. Passenger flows in each direction on four stations

4.2 Time distribution of passenger flow

After cleaning the Wi-Fi information detection data returned by device on the subway station,
comparing with the AFC data, it can be seen that the cleaned data obtained by the detecting device
accounts for 50-60% of the AFC data, and the sampling rate of the detecting device on each station is
shown in Table 2.
Table 2 The sampling rate of the detecting device

<table>
<thead>
<tr>
<th>Station</th>
<th>Cleaned detection data</th>
<th>AFC data</th>
<th>Sampling rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WuLuKou</td>
<td>96344</td>
<td>191082</td>
<td>50.42</td>
</tr>
<tr>
<td>SaJinQiao</td>
<td>96594</td>
<td>163576</td>
<td>59.05</td>
</tr>
<tr>
<td>AnYuanMen</td>
<td>154294</td>
<td>288454</td>
<td>53.49</td>
</tr>
<tr>
<td>ZhongLou</td>
<td>180032</td>
<td>314852</td>
<td>57.18</td>
</tr>
</tbody>
</table>

The time distribution of passenger flow is plotted by statistical analysis of the detection data which has been cleaned and filtered in Fig. 4. It can be seen that the time distribution of passenger flow on ZhongLou Station and AnYuanMen Station is similar, reaching the morning peak between 7:00 and 9:00, and the evening peak between 17:30 and 19:00. The passenger flow on Wulukou Station and SaJinQiao Station is significantly lower than that of Zhonglou Station and AnYuanMen Station. The peak hour of Wulukou Station is similar to that of Zhonglou and AnYuanMen Station. There is no obvious peak hour at the SaJinQiao Station, which is consistent with the actual passenger flow distribution.

![Fig. 4. The time distribution of passenger flow](image)

4.3 Travel route choice between stations

4.3.1 Travel time between two stations

The average travel speed of the subway is 35km/h, and the dwell time on each station is 20–35 seconds. The departure interval time of Line 1 is 3.5 minutes, and Line 2 is 3.0 minutes. The transfer time on BeiDaJie Station needs 2–10 minutes. The travel time can be calculated by the formula (4), and the calculation results are shown in Table 3. The $t_{\text{max}}$ and $t_{\text{min}}$ in the Table 3 mean respectively the maximum and the minimum of travel time between two stations. The abbreviation of Wu-Sa in the Table 3 means travelling from WuLuKou to SaJinQiao Station, and other abbreviations are similar.

Table 3 Travel time between two stations

<table>
<thead>
<tr>
<th>Travel time</th>
<th>Wu-Sa</th>
<th>Wu-An</th>
<th>Wu-Zhong</th>
<th>Sa-An</th>
<th>Sa-Zhong</th>
<th>An-Zhong</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{max}}$ (min)</td>
<td>9.5</td>
<td>20.41</td>
<td>20.11</td>
<td>20.11</td>
<td>19.81</td>
<td>7.25</td>
</tr>
<tr>
<td>$t_{\text{min}}$ (min)</td>
<td>3.9</td>
<td>6.8</td>
<td>6.2</td>
<td>6.2</td>
<td>5.7</td>
<td>4.01</td>
</tr>
</tbody>
</table>

The travel time between two stations is plotted by statistical analysis of the detection data which has been cleaned and filtered in Fig. 5. When no transfer stations in the travel route of passenger, such as travel directly from WuLuKou to SaJinQiao Station, travel time between two stations spent by passenger reaches nearly in the range between 5~7 min, and When transfer stations
exist in the travel route, the travel time reaches nearly in the range between 9~13 min. The result shows the travel time containing transfer time is more than the other, which corresponds to realistic situation.

The travel time includes transfer time and train travel time. As is mentioned in Section 3.2, transfer time fit the lognormal distribution in urban rail transit. To assess the travel time between two stations inferred from the detection data, the samples of transfer time are tested by Kolmogorov-Smirnov test with 95% confidence using SPSS software, and the test result shown in Fig. 6 indicates the transfer time fit the lognormal distribution. Therefore, the travel characteristics computed by the detecting device, such as the passenger travel route, travel time, and transfer time are consistent with the actual situation.

The matching data of the number of passengers choosing the travel route between two stations can be counted according to the method in Section 3.2. The number of passengers whose travel route go through AnYuanMen and ZhongLou is maximum, which is consistent with the AFC data. The matching degree is shown in Table 4, and the average matching degree reaches at 52.45%. The five matching degrees more than 50% shows the detection technology has quite high accuracy.
Table 4 The matching degree of the travel route

<table>
<thead>
<tr>
<th>Travel route</th>
<th>Matching data</th>
<th>AFC data</th>
<th>matching degree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wu-Sa</td>
<td>44014</td>
<td>85398</td>
<td>51.54</td>
</tr>
<tr>
<td>Wu-An</td>
<td>24477</td>
<td>44199</td>
<td>55.38</td>
</tr>
<tr>
<td>Wu-Zhong</td>
<td>33651</td>
<td>61485</td>
<td>54.73</td>
</tr>
<tr>
<td>Sa-An</td>
<td>18296</td>
<td>34533</td>
<td>52.98</td>
</tr>
<tr>
<td>Sa-Zhong</td>
<td>21290</td>
<td>43645</td>
<td>48.78</td>
</tr>
<tr>
<td>An-Zhong</td>
<td>107524</td>
<td>209722</td>
<td>51.27</td>
</tr>
</tbody>
</table>

## 5 Conclusion

The Wi-Fi information detecting devices can collect unique MAC addresses of mobile Wi-Fi devices and use the information to identify the passenger travel characteristics. The field experiment verification of Xi’an rail transit system shows that the information detection system has high Sampling rate, which reaches at 52.45% and can accurately identify the travel characteristics of rail passengers, such as travel route, travel time, transfer behaviour, which is consistent with those inferred from AFC data. Compared with AFC data, especially in the network-operated urban rail transit, this technology can accurately identify passengers route choice. Also, the result shows application of this technology in fare clearing has good potential and prospects due to it can provide some relevant evidence for the metro operating department. Furthermore, WI-FI information detection technology can also be applied to detect road traffic flow, detect information such as travel time, vehicle steering in road network, and carry on research about the identification of vehicle, pedestrian and non-motor vehicle trajectories at urban road intersections.

## REFERENCES


