Estimation of Metro Route Travel Time Based on Smart Card Data

Liang Gao
School of Transportation Southeast University

Jiadong Sun*
Traffic Management Research Institute of the Ministry of Public Security
Number 88, Qianrong Road, Wuxi City, China
253195895@qq.com

ABSTRACT

This research aims to evaluate the route travel time in metro transit network based on smart card data. By examining the route travel time of individual passengers, we found that the route travel time can be inferred by examining the links travel time (are the segments of the whole route) and the time in transfer stations. To acquire the time characterization at transfer stations, we extract the reference OD pairs which have only one dominant route with transfer stations in each OD pair. In specific transfer station, we found that the time interval between transfer time and access-exit time of passengers in the same time period is constant. By considering the passenger flow changes in a day, we calculate the route travel time in morning peak hours, off-peak hours and evening peak hours. The route travel time is calculated by adding the segmented trips time together. Tested on 14 big transfer stations with 4.9 million trips in London, the standard deviation of time interval between transfer time and access-time in the transfer stations ranges from 0.00 to 2.00.

KEYWORDS: metro network, route travel time, smart card data, transfer stations

1 INTRODUCTION

Metro auto fare collection system has been widely applied in many cities with metro transit systems. Numerous metro travelling records of individual passengers, smart card data, can be collected conveniently by the auto fare collection system. Due to the integrity and facticity of smart card data, a number of studies have analysed the metro passenger route choice (Asakura, Iryo, Nakajima, & Kusakabe, 2012; Gleason, Richter, & Sundberg, 2014), passenger flow prediction (Wei & Chen, 2012) and travel time reliability evaluation (L. Sun, Jin, Lee, & Axhausen, 2015; Y. Sun, Shi, & Schonfeld, 2016a). Route travel time is considered as a critical factor in these researches, especially in the analysis of passenger route choice. In some complex metro transit systems (such as London, Shanghai and Singapore), the potential routes between the original station and the terminal could be numerous. Passengers choose the most suitable route with the consideration of route travel time, crowd and transfer numbers (Raveau, Muñoz, & Grange, 2011). However, the smart card data only records the access time, exit time, access station and exit station. Inferring route travel time therefore can be implicit by applying smart card data due to the deficiency of transfer information. Many researches acquire the route travel time based on data collected from the on-site investigation (Gleason et al., 2014; Guo & Wilson,
This research contributes to the literature by estimating route travel time with an effective method based on the smart card data. The conventional methods to acquire the metro route travel time relies on survey data, which may lack of accuracy. The wide spatial and temporal aspects of the smart card data enable us to analyze the characteristics of route travel time to a greater extent. The time characteristic of different transfer station can be calculated based on the determined route travel time. The hypothesis is that the access time and exit time are independent identically distributed in the same time period for each station. Further, we can estimate the total travel time of every route by adding different time segments and related time in transfer station.

Another key contribution of our paper lies in the analysis of metro travel time reliability. Metro travel time has long been regarded for punctuality to timetable due to the high certainty of a train running time. However, the transfer time and waiting time could be longer when stations are crowd of passengers. In this paper, we will concentrate on relationship analysis between transfer time, access time and exit time. We think that the transfer time, access time and exit time could influence the whole travel time reliability, that will be discussed in the following paragraph.

This paper is organised as follows. We review the related literature in section 2. Section 3 describes the main opinions of our method. The data and result are introduced in section 4 and section 5. Discussion and conclusion are on the final section.

2. LITERATURE REVIEW

Large numbers of researches related to the application of metro smart card data have concentrated on the passenger route choice analysis (Asakura et al., 2012; Kusakabe, Iryo, & Asakura, 2010; Ranjan, Lal, & Susaeta, 2016) and travel time reliability evaluation (L. Sun, Jin, et al., 2015; Y. Sun, Shi, & Schonfeld, 2016b). Route choice behaviour studies the traveller's decision-making behaviour in choosing a suitable route from numerous alternatives. Route travel time is considered as a critical factor in influencing the passenger route choice behaviour together with crowd and transfers numbers. Since the smart card data only register the tap-in and tap-out time, Route travel time cannot be calculated directly. Some researches acquired the route travel time relying on the preference data. Raveau et al. (2011) and Guo (2011) acquired the route travel time in each OD pairs based on the data from RailPlan, conducted by Transport for London.

L. Sun, Lu, Jin, Lee, and Axhausen (2015) applied Markov Chain Monte Carlo (MCMC) algorithm to calculate access time, exit time and waiting time by taking the link travel time as input. They assume that link travel time follows an independent identically normal distribution. L. Sun and Jin (2015) solved the route travel time speculation issue by using expectation-maximization (EM) method, which is an extension of L. Sun, Lu, et al. (2015). However, the result of these two researches cannot be verified effectively due to lack of real data. The rationality of these methods remains to be further investigated. Similarly, Zhang and Yao (2015) also estimated the route travel time by applying MCMC algorithm. They extracted the representative OD pairs which have only one effective route to test the mobility and effectiveness of the parameters. In the process of calculation, they adopted the average travel
time of whole day as the algorithm input. Therefore, the result of this method could be biased due to the negligence of travel time difference between peak-hour and off-peak. Hong, Min, Park, Kim, and Oh (2016) found that the entry time of the passengers at original station appears uniformly distributed, while the exit time of the same passengers shows intermittent distribution at the destination. They speculated the route choice of an individual passenger by the tap-in and tap-out time based on their findings. Kim et al. (2015) also applied the same algorithm to detect the real path choice from smart card data. However, this method can only be applied to the routes where the transfer number are no more than two, and the travel time is less than 30 minutes. Therefore, this method could be biased for cities (such as London, Shanghai and Beijing) with huge metro transit network system where passengers’ travel time is more than 30 minutes. The method proposed in this paper is based only on the smart card data to calculate the route travel time. We estimate the route travel time in different time period of days with regard to the difference between morning peak-hours, evening peak-hours and off-peak hours.

3. METHODOLOGY

Definition reference OD pairs by OD group, we mean the OD pair with one dominate route among numerous alternative routes.

There are numerous routes between specific origin and destination in complex metro transit system theoretically. However, only small group of routes are selected by passengers in the real route choice. From figure 1, we can see that the original station is Marble arch and the destination is Westminster. Two feasible routes are shown in figure 1. Route 1: Marble Arch-Bond Street-Westminster. Route 2: Marble Arch-Notting Hill Gate-Westminster. It is reasonable that passengers prefer to choose route 1 instead of route 2 because the map distance of route 1 is only half of the route 2 distance. The map distance of shortest route is \( L_1 \) and second shortest route is \( L_2 \). The transfer number of route 1 is \( N_1 \) and second shortest route is \( N_2 \).

The reference OD pair need to satisfy one of these two requirements:

\[
L_1 \leq \frac{1}{2} L_2 \ (N_1 = N_2) \quad \text{Or} \quad N_1 < N_2 \quad \left( \frac{1}{4} L_2 \leq L_1 \leq \frac{5}{4} L_2 \right)
\]  

(1)

The aim of introducing reference OD pairs is to calculate route travel time in specific OD pair. Since the reference OD pair has a dominant route, the route travel time is equal to the time interval between origin station and terminal.

![Figure 1 Routes between Marble Arch and Westminster](image)
We find out the dominant route O-T-D in a reference OD pair. To have a better understanding of route travel time characteristic, the passengers who travel from station O to station T and from station T to station D are extracted to make a comparison with route O-T-D. Figure 2 and Figure 3 are the illustration of routes.

**Figure 2: An illustration of route O-T-D.**

**Figure 3: An illustration of route O-T and route T-D**

Route O-T-D is shown in Figure 2 with a transfer station T. \( t_{a,D}^{OD} \) is the access time (waiting time is included) in station O. \( t_{v}^{OT} \) is the vehicles operation time from station O to station D. \( t_{T}^{OD} \) is the transfer time in station T. \( t_{v}^{TD} \) is the vehicles operation time from station T to station D. \( t_{e,D}^{OD} \) is the exit time in station D. Therefore, the total travel time of route O-T-D is:

\[
t^{OD} = t_{a,D}^{OD} + t_{v}^{OT} + t_{T}^{OD} + t_{v}^{TD} + t_{e,D}^{OD}
\]  

(2)

Route O-T and route T-D is shown in Figure 2. In route O-T, \( t_{a,O}^{OT} \) is the access time in station O. \( t_{v}^{OT} \) is the vehicle operation time from station O to station T. \( t_{v}^{OT} \) is the exit time in station T. In route T-D, \( t_{a,T}^{TD} \) is the access time in station T. \( t_{v}^{TD} \) is the vehicles operation time from station T to station D. \( t_{e,D}^{TD} \) is the exit time in station D. The total travel time of route OT and route TD are:

\[
t^{OT} + t^{TD} = t_{a,O}^{OT} + t_{v}^{OT} + t_{v}^{OT} + t_{a,T}^{TD} + t_{v}^{TD} + t_{e,D}^{TD}
\]  

(3)

Consider three assumptions:

1. The access time and exit time of passengers in each station follow independent identically normal distribution in specific time period. It is given as:

\[
t_{a,j} \sim N_{j}(\mu_{j}, \sigma_{j}^{2})
\]  

(4)

\[
t_{e,j} \sim N_{j}(\gamma_{j}, \rho_{j}^{2})
\]  

(5)

Where j denotes the stations. \( \mu_{j} \) is the expected access time and \( \sigma_{j}^{2} \) is the variance of access time. \( \gamma_{j} \) is the expected exit time and \( \rho_{j}^{2} \) is the variance of exit time.

2. The vehicle operation time is fixed between two stations. Based on these assumptions, \( t_{a,O}^{OD} = t_{a,O}^{OT} \), \( t_{a,T}^{TD} = t_{e,D}^{TD} \). So the difference of \( t^{OD} \) and \( t^{OT} + t^{TD} \) is:
\[ t^{OT} + t^{TD} - t^{OD} = (t^{OT}_{a,O} + t^{OT}_{v} + t^{TD}_{e,T} + t^{TD}_{a,T} + t^{TD}_{e,D}) - (t^{OD}_{a,O} + t^{OD}_{v} + t^{OD}_{T} + t^{OD}_{e,D}) = (t^{OT}_{e,T} + t^{TD}_{a,D}) - t^{OD}_T \quad (6) \]

**Definition TDAET** We mean the Time Difference between Access-Exit and Transfer (TDAET) in station T. The transfer time is in the journey of station O to station D. The access time is on the journey of station T to station D. The exit time is in the journey of station O to station T. The TDAET can be denoted as:

\[
\Delta t_T = t^{OT} + t^{TD} - t^{OD} = (t^{OT}_{e,T} + t^{TD}_{a,T}) - t^{OD}_T \quad (7)
\]

The idea can be illustrated by an example. There are totally 59 reference OD pairs with 59 dominate routes passed station Westminster. The TDAETs of Westminster in morning peak hour was calculated, which is shown in Figure 4.

**Figure 4:** TDAET in Westminster in morning peak hours

The TDAETs with 0.61 standard deviation vary from 0.5mins to 3.5mins, which fluctuate in a small range. In a transfer station T, we identified three types of passengers: transfer passengers, access passengers and exit passengers. During the same period of time (such as the morning peak hours), the difference between the passenger’s transfer time, access times and exit time is a constant value due to the low standard deviation of TDAETs in the transfer station. Therefore, for arbitrary routes, the route travel time can be calculated by the segments time and TDAET:

\[
t^{OT_1T_2...T_nD} = t^{OT_1} + \sum_{i=1}^{n} t^{T_{i-1}T_i} + t^{T_nD} + \sum_{i=1}^{n} \Delta t_{Ti} \quad (8)
\]

Where:
- \( t^{OT_1T_2...T_nD} \) is the route travel time from station O to station D with the transfers: \( T_1, T_2, ..., T_n \).
- \( t^{OT_1} \) is the route travel time from station O to station \( T_1 \).
- \( \sum_{i=1}^{n} t^{T_{i-1}T_i} \) is the route travel time between \( T_1, T_2, ..., \) and \( T_n \).
- \( t^{T_nD} \) is the route travel time from station \( T_n \) to station D.
\[ \sum_{i=1}^{n} \Delta t_{O_i} \] is the total TDAETs of transfer station: \( T_1, T_2, ..., T_n \).

We assume that the TDAET of station O to station D is equal to station D to station O. Therefore, for transfer stations with only two metro lines, we only need to calculate one type of TDAET in station T. However, for the large transfer stations with many metro lines, the TDAETs between different lines need to be calculated separately. For example, the Green Park Station has three metro lines: Jubilee line, Piccadilly line and Victoria line. We need to calculate three types of TDAETs between Jubilee line, Piccadilly line and Victoria line. In summary, the procedures for using metro smart card data to calculate the arbitrary route travel time can be illustrated as following:

1. The transfer station of the objective route should be identified. Such as route A-T-B.
2. Find out the reference OD pairs with transfer station T from the map.
3. Extract the passenger travel time of route OD, OT and TD. Since the TDAET is calculated in the same period of time, so the selection of passenger smart card records are limited in the specific period of time.
4. In order to reduce the randomness of route travel time, we set the requirement that the numbers of passenger records should large than 90. If the condition is not satisfied, the reference OD pair needs to be re-selected.
5. After acquiring the route travel time OD, OT and TD in each O-T-D pairs, the TDAET of transfer station can be calculated directly.
6. In order to acquire the objective route travel time, the link travel time of A-T and T-B need to be calculated which are the parts of route AB.
7. The objective route travel time can be calculated based on the result in step (5) and step (6).

Figure 5 The flow chart of algorithm
4. DATA AND RESULT

The smart card data used in this research are collected from Transport for London (TfL). According to the requirements for peak hour partition of London (see Web-1), we calculate TDAET of each transfer station in different time period: the morning peak hours are from 6:00 am to 9:00 am, the off-peak hours are from 9:00 am to 16:00 pm, the evening peak hours are from 5:00 pm to 19:00 pm.

The critical step to calculate the TDAET of individual transfer stations is to find out reference OD pairs with transfer station T. There are totally 1,652 OTD pairs for 14 transfer stations. To avoid biased travel time observation, we removed the OTD pairs which less than 90 transactions in the process of calculating. Based on the smart card data collected from TfL, TDAETs were calculated in different time period. In order to ensure that different passengers’ transfer time was in the same time period, for example in morning peak hours, we claimed that the access time should be larger than 6:00 am and smaller than 9:00 am. Based on the OTD pairs we selected, most passenger’s travel time of station O to station T varied from 10mins to 20mins. Although a small group of passengers’ transfer time was included in morning peak hours period, the majority of passengers’ transfer time would be included. The total result is given in Table 1 which includes the mean value and standard deviation of 16 TDAETs for each transfer station. 14 transfer stations in district 1 and two transfer stations in district 2 have been selected to calculate the TDAET. There are totally 19 TDAETs for 14 transfer stations. The biggest metro station King’s Cross St. Pancras with 6 metro lines was removed, which could bias the result due to its over complexity. Most average TDAETs of transfer stations are between 0.5mins and 4mins. Average deviation and standard deviation are two measures of statistical dispersion. Table 2 gives the average deviation and standard deviation of each station, and they are shown in Figure 7. 1503 OD pairs with 3908 routes’ travel time are calculated based on the method we proposed, which are shown in Table 3.
Table 1 TDAETs of Transfer Stations in Different Time Period

<table>
<thead>
<tr>
<th>Stations</th>
<th>Morning Peak Hours</th>
<th>Off-Peak Hours</th>
<th>Evening Peak Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Numbers</td>
<td>Mean</td>
</tr>
<tr>
<td>Baker Street J-B</td>
<td>1.66</td>
<td>10</td>
<td>1.14</td>
</tr>
<tr>
<td>Baker Street J-C</td>
<td>4.86</td>
<td>24</td>
<td>4.12</td>
</tr>
<tr>
<td>Bank</td>
<td>1.10</td>
<td>17</td>
<td>1.05</td>
</tr>
<tr>
<td>Bond Street</td>
<td>2.17</td>
<td>19</td>
<td>2.16</td>
</tr>
<tr>
<td>Embankment</td>
<td>0.79</td>
<td>16</td>
<td>1.48</td>
</tr>
<tr>
<td>Green Park P-J</td>
<td>1.08</td>
<td>13</td>
<td>1.07</td>
</tr>
<tr>
<td>Green Park P-V</td>
<td>1.04</td>
<td>17</td>
<td>1.35</td>
</tr>
<tr>
<td>Green Park V-J</td>
<td>2.12</td>
<td>24</td>
<td>2.36</td>
</tr>
<tr>
<td>Holborn</td>
<td>2.42</td>
<td>64</td>
<td>2.35</td>
</tr>
<tr>
<td>Leicester Square</td>
<td>1.63</td>
<td>12</td>
<td>1.86</td>
</tr>
<tr>
<td>Mile End</td>
<td>1.31</td>
<td>29</td>
<td>1.61</td>
</tr>
<tr>
<td>Oxford Circus B-V</td>
<td>1.36</td>
<td>18</td>
<td>2.50</td>
</tr>
<tr>
<td>Oxford Circus C-B</td>
<td>1.44</td>
<td>23</td>
<td>2.64</td>
</tr>
<tr>
<td>Oxford Circus C-V</td>
<td>1.94</td>
<td>32</td>
<td>1.75</td>
</tr>
<tr>
<td>Stockwell</td>
<td>2.54</td>
<td>52</td>
<td>2.99</td>
</tr>
<tr>
<td>Tottenham Court Road</td>
<td>3.27</td>
<td>17</td>
<td>2.27</td>
</tr>
<tr>
<td>Victoria</td>
<td>1.14</td>
<td>21</td>
<td>1.17</td>
</tr>
<tr>
<td>Warren Street</td>
<td>4.15</td>
<td>30</td>
<td>3.15</td>
</tr>
<tr>
<td>Westminster</td>
<td>1.83</td>
<td>59</td>
<td>2.60</td>
</tr>
</tbody>
</table>

NOTE: J denotes the Jubilee line. C denotes the Circle line. B denotes the Bakerloo line. P denotes the Piccadilly line. V denotes the Victoria line.

Figure 7: Standard Deviations and Average Deviations of TDAET in Each Station

Table 3 The Estimated Route Travel Time 4 Route Travel Time

<table>
<thead>
<tr>
<th>Origin Station</th>
<th>Transfer Station</th>
<th>Terminal</th>
<th>Morning Peak Hours</th>
<th>Off-Peak Hours</th>
<th>Evening Peak Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank</td>
<td>Embankment</td>
<td>Paddington</td>
<td>33.210</td>
<td>38.570</td>
<td>39.502</td>
</tr>
<tr>
<td>Bank</td>
<td>Westminster</td>
<td>Paddington</td>
<td>50.212</td>
<td>42.554</td>
<td>39.237</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Green Park</td>
<td>Holborn</td>
<td>Mile End</td>
<td>23.182</td>
<td>23.432</td>
<td>25.625</td>
</tr>
<tr>
<td>Green Park</td>
<td>Oxford Circus</td>
<td>Mile End</td>
<td>11.996</td>
<td>13.219</td>
<td>13.308</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Earl's Court</td>
<td>Green Park</td>
<td>Neasden</td>
<td>30.571</td>
<td>32.472</td>
<td>32.247</td>
</tr>
<tr>
<td>Earl's Court</td>
<td>Westminster</td>
<td>Neasden</td>
<td>37.058</td>
<td>36.219</td>
<td>36.738</td>
</tr>
<tr>
<td>Hammersmith (Dis)</td>
<td>Green Park</td>
<td>Tottenham Hale</td>
<td>37.607</td>
<td>38.478</td>
<td>38.301</td>
</tr>
<tr>
<td>Hammersmith (Dis)</td>
<td>Victoria</td>
<td>Tottenham Hale</td>
<td>40.623</td>
<td>38.392</td>
<td>40.331</td>
</tr>
<tr>
<td>Marble Arch</td>
<td>Holborn</td>
<td>Piccadilly Circus</td>
<td>22.341</td>
<td>20.287</td>
<td>22.028</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

NOTE: ... means the rest.

### 5. DISCUSSION AND CONCLUSION

TDAET is defined as the time interval which equals to the access time plus the exit time minus the transfer time in each OTD air. It is shown in Table 1 that all the TDAETs of each transfer station are greater than 0, which indicates that the transfer time between metro lines is less than the total time of access and exit time. From the Table 1, we can also see that the average TDAETs of Warren Street and Baker Street (Jubilee-Bakerloo) are relatively higher than other stations. That could be the difference of station layout in different transfer stations. For trips with longer transfer distance, passengers need to walk for a longer time. Nevertheless, the access time and exit time are not influenced equally. For example the Green park station, the average TDAET of Central line to Victoria line is higher than that of Piccadilly line to Jubilee line and Piccadilly line to Victoria line, which could be the longer transfer distance of Central line to Victoria line. The stations with insufficient TDAETs are removed, such as the Baker Street (Jubilee-Bakerloo). The estimated routes travel time of different OD pairs are presented in Table 2. The route travel time in off-peak hours is less than the travel time in morning peak hours and evening peak hours, which might be the lower passenger flow in off peak hours (7).

The accuracy of our method can be tested by the statistical dispersion of the total TDAETs of OTD pairs in each transfer station. In the specific time period, TDAET is the access time of passengers on OT route plus the exit time of passengers on TD route minus the transfer time of passengers on OD trip. The small standard deviation of the TDAETs indicates that TDAETs of different OTD pairs are approaching one another. We claim that the TDAET of arbitrary route O-T-D can be denoted by the average TDAET of the corresponding transfer station. Therefore, the arbitrary route travel time can be estimated by the average TDAET on transfer stations and links travel time which are parts of the total route. From Table 2 and Figure 7, we can see that average deviation of TDAETs for transfer stations are less than 1.4mins, and the standard deviations of TDAETs range from 0 to 1.5. These indicate that TDAETs of each
transfer station fluctuates in a small range.

Smart card data only records the information of access and exit station, which is implicit to infer the route travel time. This research estimates the route travel time by combining the links travel time of related routes. In order to analyses the characteristics of the transfer time, the access time and the exit time in transfer stations, reference OD pairs are extracted each of which has only one dominant route with the responding transfer station. We calculate the TDAETs in each OTD pair, and also find that the standard deviation and average deviation of TDAETs in transfer station are in small ranges. Therefore, we claim that the TDAET of the arbitrary route can be denoted as the average TDAET of responding transfer station. The route travel time can be calculated by adding the associated link travel time and average TDAET of transfer stations.

The contribution of this research is the calculation of route travel time based only on the smart card data in the absence of passengers’ route choice. Route travel time is an important factor which influence passenger’s route choice behavior. Therefore, our research will contribute the analysis of metro route choice behavior. TDAETs is a key variable in this research which could have relations with the crowdedness, station structure and other factors, can influence the reliability of route travel time. In future researches, we will first concentrate on the probability of passenger route choice by applying gaussian mixture model based on the route travel time we calculate in this paper. The relationship between travel time reliability analysis and TDAETs of station could also be further investigated.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 71701042), the Natural Science Foundation of Jiangsu Province (Project No.BK20150615) and Open Project of Key Laboratory of Ministry of Public Security for Road Traffic Safety, Project No. 2017ZDSYSKFKT12-1

Reference

behavior with smart card transaction data. *Transportation, 37*(5), 731-749.


Web sites: