Impact Factor Value 4.046 e-ISSN: 2456-3463
International Journal of Innovations in Engineering and Science, Vol. 3, No.5 2018
www.ijies.net

Assigning Passenger Flows on a Metro Network Based on Automatic Fare Collection Data and Timetable in Shenzhen Metro System

Rohini D S , Roja G , Gayithri S , Namratha G

Department of Computer Science and Engineering,
Visvesvaraya Technological University, Belgavi, Karnataka, India,

rohinisiddharth097@gmail.com,rojarenuka2@gmail.com,14namratha@gmail.com,gayithrigowda97@gmail.com

Abstract: *Assigning passenger flows on a metro network plays an important role in passenger flow analysis that is the foundation of metro operation. Traditional transit assignment models are becoming increasingly complex and inefficient. These models may even not be valid in case of sudden changes in the time table or disruptions in the metro system We propose a methodology for assigning passenger flow on ametro based on Automatic fare collection(AFC) data and realized time table. We find that the routes connecting a given Origin and Destination(O-D) pair are related to their observed travel times (OTTs) especially there pure travel times (PTTs)abstracted from AFC data combined with realized time table.A novel clustering Algorithm is used to cluster trips between a given O-D pair based on PPTs /OTTs and complete the assignment .An intial application to categorical O-D pairs on the Shenzhen metro system, which is one of the largest system in the world, shows that the proposed methodology works well .Accompanying the initial application,an interesting approach, is also provided for determining the theoretical maximum accuracy of the new assignment model.*

I-INTRODUCTION

As an efficient transport system, the metro system is now the mainstay of urban passenger transport in many megacities, especially in highly populated areas. passenger flow is the foundation of making and coordinating operation plans for a metro network plays an important role in analyzing passenger flows. A number of studies have developed passenger flow assignment models .However ,these models are becoming increasingly complex because of many diverse parameter types .In the case of sudden changes in the timetable or disruptions in the metro system ,these models may not be valid.

1. Different from private cars ,a metro system is operated according to the timetable ,which is an important constraint for a passengers travel .New technologies are widely introduced into metro systems ,resulting in improvements in passenger flow assignment .For example the automatic fare collection(AFC) system has become the main method for collecting metro fares in many cities in the world .This system records the origin and destination station of a trip and their corresponding time stamps .The transaction data obtained through these AFC system contain a vast amount of archived information on how passenger use a metro system. Up to date ,however, there are limited studies on AFC data or how to assign passenger flows efficiently by combining these data with the timetable.
2. This paper mainly focuses on how to efficiently model the passenger flow assignment problem for a metro network with AFC data and timetable.

1.1. Timetable and AFC Transaction Data
1.1.1. Timetable Information

The metro timetable contains the set of all train trips with arrival and departure times per station and per train number. Figure 1is an example of the timetable for a metro line in the Shenzhen metro system .Since the metro system is operated based on the timetable, a passengers travel time between the origin and destination stations is subject to not only the chosen route but also the timetable.

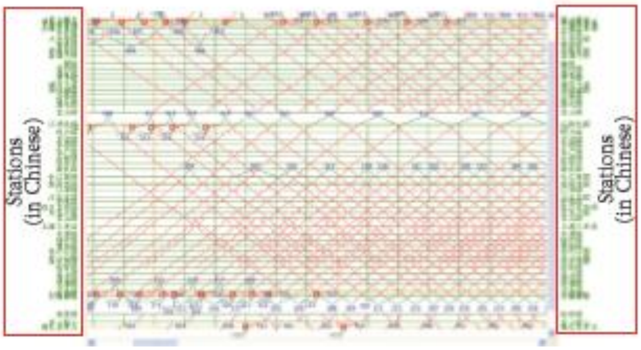


Figure 1: Example of timetable

1.1.2 .AFC transaction Data

The assignment address in this paper obviously requires AFC transaction data. The ID number of a smart card holder is recorded each time the holder passes the entry or exit gates, and the corresponding transaction record indicates a unlinked trip. These smart card transaction records provide information on ID numbers, the date ,departure station ,passage time at an entry gate, arrival station ,and passage time at an exit gate. The entry and exit times are recorded in the exact number of seconds, based on which observed travel times can be obtained. Example AFC data are shown in Table 1.Our initial analysis (Figure 2) of the observed travel times indicates that the routes connecting a given O-D pair are related to their observed travel times, although there is also travel time uncertainty the route level.

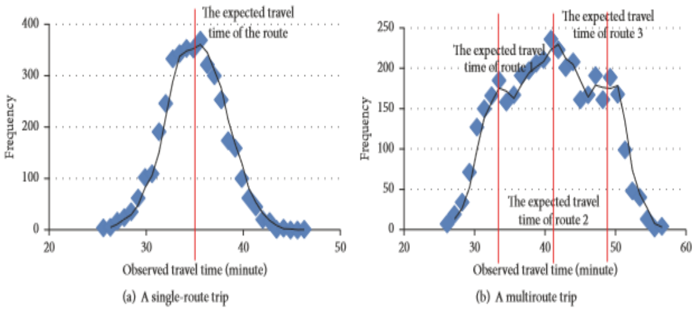


Figure 2: Frequency distribution of the observed travel times extracted from AFC data.

1.2. Problem Description

The observed travel time is relevant to the passenger travel process, Figure 3 shows a typical travel process of a metro passenger .It consists of entry walking ,waiting for train, traveling in vehicle, transfer walking, waiting for transfer trains if necessary, and exit walking. Correspondingly, the observed travel time(OTT) of a passenger includes entry walking time(ENT),waiting time on platforms(WT),in vehicle time(IVT),exit walking time(EWT),transfer walking time(IWT),and another waiting time(WT) if there is a transfer .Moreover ,we define that CIT is the check-in time at the origin station recorded by AFC data ,COT is the check-out time at the destination station ,BOT is the actual time point that the passenger boards ,on the train ,and GOT is the actual time point that the passenger gets off from the train. Obviously, both BOT and GOT are related to the timetable. Thus, the interval between CIT and BOT is the sum of ENT and WT, and the interval between COT and GOT is EWT. Based on the abovementioned, the pure travel time(PTT), which is relevant to the timetable and an important notion that we defined in this paper, can be calculated from the interval between BOT and GOT.

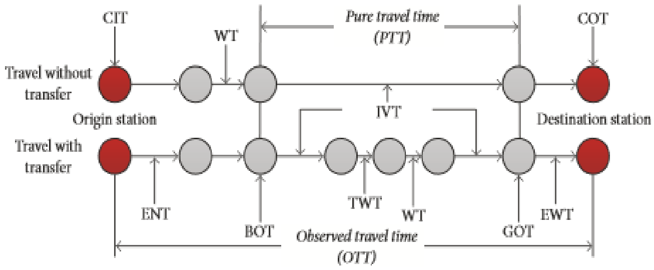


Figure 3: Typical travel process of a metro passenger

As mentioned in section 1.1 OTTs derived from AFC data are relevant to the route choices, and there may be a wide variation of OTTs for a given O-D pair, especially in a large scale network. In extreme cases, the origin WT can affect a passengers OTT to such a great extent that there is no determinate relationship between the OTT and possible routes. For example, if the OTTs between two routes vary only by 3 minutes while the interval between services is 9 minutes, it is difficult to assign an OTT from AFC data to one of the routes as on average 4.5 minutes of WTs result from the random CITs (figure 4).

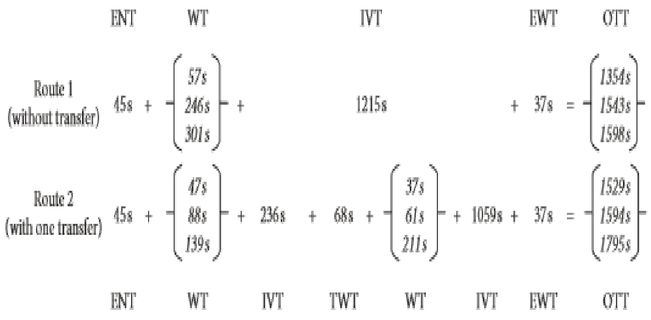


Figure 4: Illustration of random WTs influence on OTTs.

The relationship between the possible routes for a given O-D pairs and corresponding PTTs which delete ENTs, EWTs, and origin WTs from OTTs (Figure 5). How to abstract these OTTs/PTTs based on AFC data and then complete the passenger flow assignment with them?

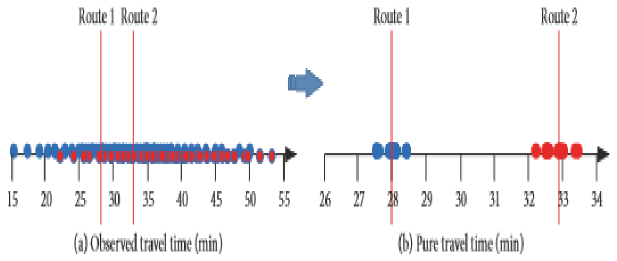


Figure 5: Distributions of observed travel times (OTTs) and pure travel times (PTTS)

The objective is to propose a methodology to assign passenger flow on a metro network based on travel time (OTTs/PTTs) abstracted from AFC data. The following approaches are used:

1 .We propose a transit assignment model using revealed information including AFC data and realized time table of metro system .

2. We introduce a novel clustering approach to conduct the assignment .It is only based on the distance between data points and can detect non spherical clusters and automatically the correct number of clusters**.**

3. We find that PTT is better than OTT when being used for clustering it can reduce variation of travel times for O-D pairs to a great extent**.**

4. We also provide an approach, accompanying the initial application to categorical O-D pairs on the Shenzhen metro network, for determining the theoretical maximum accuracy.

II- LITERATURE REVIEW

Passenger flow is required to make and co –ordinate operational plans for a metro system. Conventionally, models to solve passenger flow assignment problems can be classified according to whether war drops principle is flowed [5].One model is the non equilibrium assignment, and the other is the equilibrium assignment model [6]. Moreover, it is assumed tht the process of passenger's choice has some random characteristics because of imperfect knowledge of travel time, individual differences ,measurement errors ,and so[5-7].therefore, confronted with today's metro systems, the result from passenger's route choices can be described more appropriately by the stochastic user equilibrium (SUE) with time and space constraints, which is the passengers can choose *j*th train to arrive at their destination or transfer station. The proved by some simulation experiments [5,8]and full-scale case tests[4]. Up to date, those models to solve a SUE problem are becoming increasingly complex due to the many diverse parameter types. Through review were presented in some of the literature [2, 3,9,10].

In recent years, automatically collected fare data such as smart card data have been used by transit service providers to analyze passengers demand and system performance. these data have been used for O-D matrices estimation[11,12], travel behavior analysis[15], operational management, public transist planning [16-18], the stud0ies on the use smart card data can be grouped into 3 categories: strategic(long term planning), tactical(service adjustment and network development), and operational(ridership statistics and performance indicators).

Chan developed two applications based on Oyster card data in the London Underground d : one of these estimated an O-D flow matrix ,while other constructed rail service reliability metrics .This is the first attempt at measuring service delivery quality using elapsed travel time . X u et al.try to estimate metro passengers route choice behavior using smart card data and proposes a new model for passenger flow assignment based on an AFC system environment .However, the problem of calibrating the vast number of parameters in behavior functions such as arrival /departure distributions still exist. Zhu et al. present a method for calibrating metro assignment models using

AFC data .Their calibration approach uses a genetic algorithm-based framework with nonparametric statistical techniques.

The existing studies on transit assignment models with AFC data are either too simple or too computationally costly and should be improved.

III- METHODOLOGY

3.1. OTT and PTT Abstracting Approach.

Since CITs and COTs are recorded in the AFC data, it is convenient to obtain OTTs .This section focuses on abstracting PTTs from AFC data. We first give some basic definitions on the train timetable. The train timetable illustrates the relationship between space and time of train operation. The main information it contains are trains' arrival and departure times at each station. Denote the set of metro lines as L = {1,2, ...,l,...,N}andSl ={1,2,...,i,...,M}as the set of stations in line l; Sl,i means station i in line l. Then the arrival time Ajl, and departure time $Dj\ l$, of jth train at station Sl, can be described as Sl,(Ajl,i,$D\ j\ l$,i). Therefore, its train path is defined as the collection {$\forall i \in l$ | Sl,(Ajl,i,$D\ j\ l$,i)}. For each line, each station, and each train, the train timetable can be represented by T = {$\forall j$,, | $S\ l$,i(Ajl,i,$D\ j\ l$,i)}. Moreover, an AFC data record can be described as OD (ID, CIT, COT, entry_st_no , exit_st_no), where enter_st_no and *exit_st_no* represent enter station ID and exit station ID, respectively.

3.1.1. Determination of BOT

Let be its CIT from AFC data, and let Sl, be its enter_st_no. Searching every train which runs through the station S l,i in order, the train that stops at station S l ,i at time t can be determined by locating j such that

$D^{j-1}{}_{l},\leq t\leq D^{j}{}_{l,i.}$ (1)

The passengers can choose jth train to arrive at their destination or transfer station. The search process is illustrated in Figure6. Therefore, the ATT can be set as

BOT←$D^{j}{}_{l,j'}$ (2)

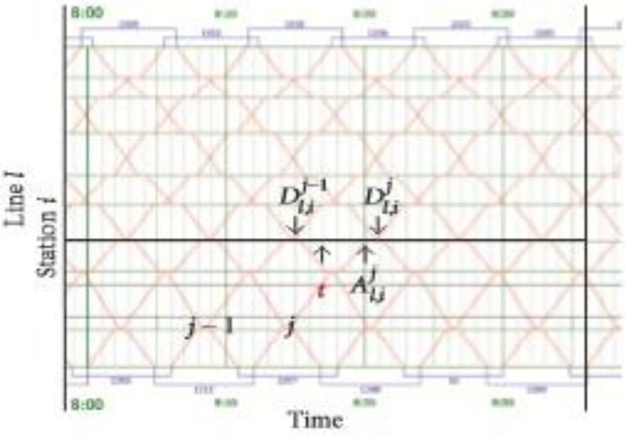
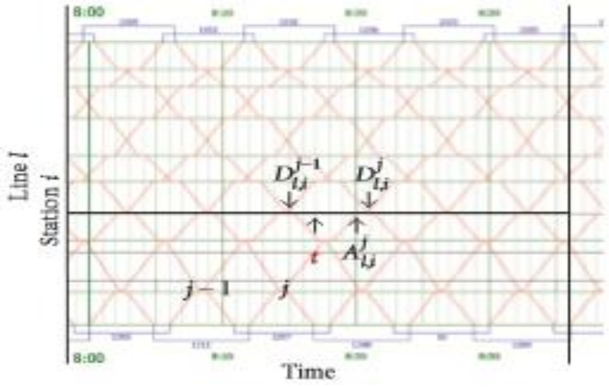


Figure 6: Illustration of how to get BOT

3.1.2. Determination of GOT:

Similarly, let t be its COT from AFC data , and letSl,ibe its exit_st_no, as shown in Figure7. Then search every train which stops at the station $S_{l,'}$ in reverse order. Passengers getting off the jth train can be obtained from the condition

$A^{j}{}_{l,\mathrm{I}'}\leq t\leq A^{j+1}{}_{l,\mathrm{I}}$ 3



,

Figure 7: Illustration of how to get GOT

Passengers will check out from station once they get off trains, it is simpler for containing no waiting time comparing passengers' entry. Thus, GOT is equal to$A^{j}{}_{l,,i,'}$. It should be noted that there is a minimum interval threshold between CIT and BOT as well as GOT and COT, because walking or waiting will also need time.

Therefore, smart card data can be trimmed as follows**:** From OD(ID ,CIT ,COT, enter_st_no,exit_st_no) to OD(ID ,BOT ,GOT ,enter_st_no,exit_st_no)

3.1.3. Determination of PTT.

After the AFC data record is trimmed from OD(ID, CIT,COT,enter_st_no,exit_st_no) to OD (ID ,BOT ,GOT, enter_st_no,exit_st_no),PTT can be expressed as

PTT = GOT−BOT. (4)

3.2.A Novel Clustering Approach.

Since the AFC transaction data can be used to estimate passengers' route choices, it is possible to use it for passenger flow assignment. To achieve this, we have applied cluster analysis techniques. Unlike the existing assignment model, the cluster analysis technique in this paper clusters trips between a given O-D pair based on PTTs/OTTs derived from the AFC data. It then assumes that similar PTTs/OTTs are linked to the same route. Cluster centers for a given O-D pair are considered the expected travel times (ETTs) of the feasible routes, and PTT/OTT is assigned to th e corresponding cluster center.

Several clustering strategies have been proposed, including the k-means method [26], the k-medoids method [27], distribution-base d algorithms [28], density-based algorithms [29],and the mean-shift method. However, a novel clustering approach was recently proposed by Laio and Rodriguez[30]. We have used this method for the following reasons

1) The k-means and k-medoids methods cannot detect non spherical clusters, because a data point is always assigned to the nearest center .The OTTs for a given O-D pair consist of non spherical clusters

2) Distribution-based algorithms attempt to reproduce the observed data points using a mix of predefined probability distribution functions. The accuracy of such methods depends on how well the trial probability represents the data.

(3) Density-based algorithms choose an appropriate threshold which may be nontrivial, though clusters with an arbitrary shape can be easily detected by approaches based on the local density of data points.

(4) The mean-shift method only works for data defined by a set of coordinates and is computationally costly, although it does allow for non spherical clusters and does not require a non trivial threshold.

(5) The clustering approach proposed by Laio and Rodriguez [30] is superior, because it is only based on the distance between data points, it can detect non spherical clusters ,and it automatically determines the appropriate number of clusters.

The adopted clustering approach is based on the idea that cluster centers are characterizing d by a higher density than their neighbor s and by a relatively large distance from points with higher densities. For each data point i, we compute two quantities: its local density ρ_i and its distance δi from points of higher density. Both these quantities depend only on the distances between data points, which are assumed only on the distances n d_{ij} between data points, which are assumed to satisfy the triangular inequality. The local density ρ_{I} of data point i is defined as

$$\boldsymbol{\rho^{i}= \sum\psi\,(d_{ij} - {}^{\mathrm{D}}{}_{\mathrm{c}}),} \qquad \textbf{(5)}$$

Where $\psi\ (x) = 1\mathrm{if}x<0\mathrm{and}(x) = 0\mathrm{otherwise}.dc$ is a cutoff distance, and ρi is the number of points that are closer than dc to point i. The algorithm is only sensitive to the relative magnitudes of ρi values for different points. This implies that, for large data sets, the results of the analysis are robust with respect to the choice of d_c. δ_i is determined by computing the minimum distance between point i and any other point with higher density. That is,

$$\delta i = \min_{j:\rho j>\rho i}(dij). \qquad (6)$$

For the point with the highest density, we conventionally take $\delta_i = \max(dij)$. Note that δ_i is much larger than the typical nearest neighbor distance only for points that are local or global maxima in the density. Thus, cluster centers are recognized as points for which the value of δ_i is anomalously large(as shown in Figure8).

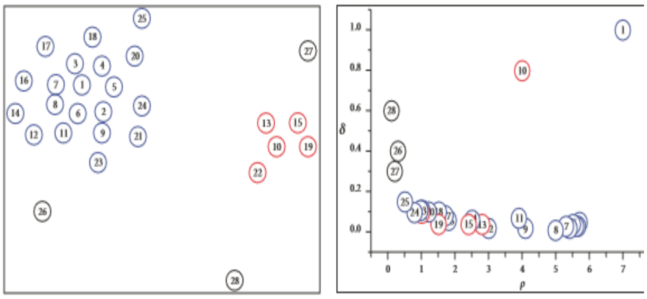


Figure 8: The algorithm in two dimensions.

After the cluster centers have been found, each remaining point is assigned to the same cluster as its nearest neighbor of higher density. The cluster assignment is performed in a single step,in contrast with other clustering algorithms where an objective is optimized iteratively

IV- INITIAL APPLICATION AND ANALYSIS TO CATEGORICAL O-D PAIRS ON THE SHENZHEN METRO NETWORK

4.1. Passenger Flow Assignment for the O-D Pairs with Single Route.

Although the proposed approach aims to assign passenger flows to the routes between a given O-D pair, those O-D pairs with single route should be identified first of all. There are two types of O-D pairs with a single route:

(1) O-D pairs with a unique physical route on the network.

(2) O-D pairs that have only one feasible route when we consider the travel cost threshold, although there is more than one physical route on the network.

In both of the abovementioned cases, all the passengers for the O-D pair are assigned to only one route. And the procedure is similar to All or Nothing Assignment Model.

Taking the Shenzhen metro as an example, the feasible route set for a given O-D pair is generated using a two step route generation method. First, the *k*th-shortest path algorithm is applied and a universal route set is generated based on the physical topology of the metro network .Second, the universal set is filtered by judging the rationality of alternative routes based on the difference in the travel costs of the alternative and shortest route. This narrows the feasible route set.

The initial statistics of the Shanghai metro network demonstrates that there is a large percentage of O-D pairs with a single route(35.98%interms of O-D pairs and 60.15% in terms of trips).

4.2. Passenger Flow Assignment for the O-D Pairs with Multi routes

4.2.1. Estimating Passenger Route Choices with the Clustering Technique and Determinate PTTs.

Except those O-D pairs with single route, there are a large number of O-D pairs with multiple routes, for which passenger route choices can be estimated using the determinate PTTs and proposed clustering technique. Consider an example O-D pair with two feasible routes on the Shenzhen metro network. The distribution of OTTs and the corresponding probability density function are shown in Figure 9(a). Using the abstracting approach proposed in Section 3.1, OTTs can be further fined to PTTs shown in Figure 9(b). We computed two quantities for each point of PTTs in this example data: its local density (ρ_i) and its distance from points with higher densities (δ_i), with the corresponding decision graph being shown in Figure 9(c). We can see that two points (blue and red) have large δ values and a size able density. These two points Correspond to cluster centers, which represent the expected PTTs of two routes between the O-D pair. After determining the two centers, each point is assigned to a cluster, which is used to calculate route choice probabilities for the O-D pair(Figure9(d)).

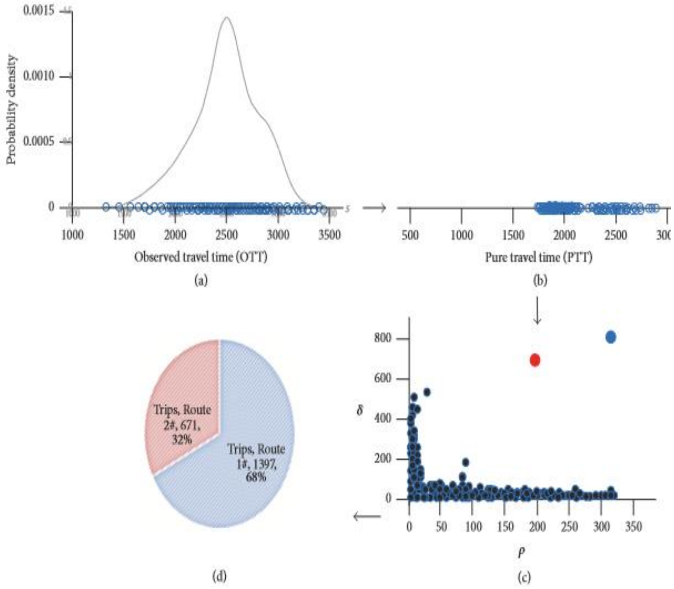


Figure 9: Cluster analysis of the pure travel times for an O-D pair with feasible route.

The test O-D pairs discussed in this section are those with determinate PTTs for which the passenger route choices can be estimated accurately to a great extent and consequently a precise passenger flow assignment result can be obtained. Takingthe Shenzhen metro network as an example, our initial calculations and analyses for all of the O-D pairs on the network showed that there are 42611 O-D pairs (35.39% in terms of O-D pairs, 22.22% in terms of trips)falling into this category of O-D pairs.

However, there are also other categories of O-D pairs that may not be suitable for the estimation of passenger route choice using PTTs .In these cases ,a passenger’s travel behavior is so complex that it is difficult to determine the passenger’s PTT. For example, if both the upstream and downstream are feasible directions for the origin station of an O-D pair to the destination (Figure 10(a)), or the origin station of an O-D pair is a transfer station(Figure10(b)),we cannot judge which train a passenger boards on in reality and consequently the corresponding PTT is not determinate .The following section discusses how to estimate these categories of O-D pairs. this category of O-D pairs.

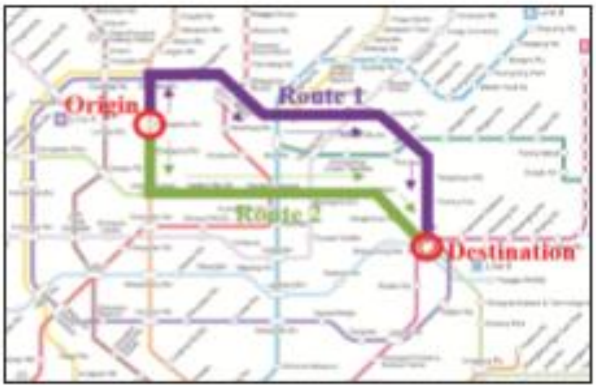


Figure 10: Illustrations of cases where we cannot judge which train a passenger boarded on and corresponding PPT is not determinate.

4.2.2. Estimating Passenger Route Choices with the Clustering Technique and Indeterminate PTTs.

For the category of OD pairs with multiple routes and indeterminate PTTs, we use the clustering technique and OTTs to estimate passenger route choices based on which the passenger flow assignment is completed .Our initial calculations and analyses for all the O-D pairs on the Shenzhen networks how that there are 34472 O-D pairs (28.63% in terms of O-D pairs, 17.63%in terms of trips)falling into these categories of O-D pairs.

Of course, the result from this assignment for the abovementioned O-D pairs may not be accurate due to a possible wide range variation of OTTs. However, among these categories of O-D pairs, there is still a kind of O-D pairs for which the corresponding assignment result can be precise to a great extent. It is because the expected travel times of routes for a given O-D pair falling into this kind of O-D pairs are obviously different from each other, and consequently the corresponding OTTs can be clustered into the routes accurately. For the Shenzhen Metro network, the corresponding percentage is 5.06% in terms of O-D pairs, as well as 7.14% in terms of trips**.**

Moreover, there are some O-D pairs for which we cannot give accurate route choice estimations**.** Such O-D pairs include those with similar expected tr its different connecting routes and with small flows from several to several dozen passenger .In the case of these O-D pairs ,the route choices of passengers are stochastic to a great extent. For the Shenzhen metro network, the corresponding percentage is 19.46% in terms of O-D pairs, as well as 3.87% in terms of trips**.**

V- DISCUSSIONS AND CONCLUSIONS

5.1. Extended Discussions to the Proposed Approach.

From the analysis in the previous sections, the proposed approach in this paper can efficiently estimate metro passenger route choices using a novel clustering technique and processed AFC data (PTTs/OTTs) and consequently provide appropriate passenger flow assignments on a metro network. Furthermore, the approach implies the potential of measuring its minimum and maximum accuracy; the minimum practice by classifying all the O-D pairs into several categories. Taking the Shenzhen metro network as an example, as shown inTable2, we can measure the minimum and maximum accuracy of the approach as follows.

(1) O-D pairs with single route :the passenger flow assignment using the proposed approach is accurate for this category of O-D pairs because there is only one feasible route between a given O-D pair and a passenger’s route choice is unique. For the Shenzhen metro network, the corresponding percentage is 35.98% in terms of O-D pairs, as well as 60.15% in terms of trips. There is a large percentage of OD pairs for which the estimated route choices are always correct, regardless of the assignment model. This is a n interesting characteristic of a metro network compared with an urban road network.

(2) O-D pairs with multiple routes and determinate PTTs: the passenger flow assignment using our approach is accurate for this category of O-D pairs because the variation of travel times for a route is narrowed to a smaller range by using PTTs instead of OTTs. For the Shenzhen metro network, the corresponding percentage is m and maximum accuracy can be approached in 35.39%in terms of O-D pairs, as well as 22.22%in terms of trips.

(3) O-D pairs with indeterminate PTTs but obviously different expected travel times for routes: for the route choices between an O-D pair in this category whose route expected travel times are obviously different from each other, the proposed approach can also give an accurate

assignment .For the Shenzhen metro network, the corresponding percentage is 5.06% in terms of O-D pairs,as well as7.14%in terms of trips.

(4) Some special O-D pairs: we cannot give accurate passenger flow assignment for them. In the case of these O-D pairs, the route choices of passengers are stochastic to a great extent. For the Shenzhen metro network, the corresponding percentage is 19.46% in terms of O-D pairs, as wellas3.87%in terms of trips.

(5) Others: except the above categories of O-D pairs ,the remainder is those O-D pairs for which the proposed approach cannot guarantee giving an accurate assignment but may have the potential of approaching the actual route choices in theory. In summary, based on the above discussions for different categories of O-D pairs, the minimum and maximum accuracy of the proposed approach with the clustering technique and AFC data can be measured in practice. Taking the Shenzhen metro network as an example, the proposed approach is accurate for 94.10% of trips, cannot be accurate for 5.28% of trips, and may be accurate for 0.62% of trips. And the total accuracy range is 75.43%∼79.54%in terms of O-D pairs with 89.51%∼96.13%in terms of trips.

5.2. Concluding Remarks.

A metro system is operated based on the timetable. Developments in the application of ADC systems such as AFC systems have made the collection of detailed passenger trip data in a metro network possible. In this paper, we aim to propose an efficient approach to assign passenger flows on a metro network k combing AFC data and time table. The advantages of the proposed approach include the following:

(1) A posteriori transit assignment model, which uses *reVealed* information including AFC data and timetable of metro systems rather than a priori knowledge, wa s proposed.

(2) A novel clustering approach was introduced to conduct the assignment .It is only based on the distance between data points and can detect non spherical clusters and automatically the correct number of clusters.

(3) It was found that PTT is better than OTT when being used for clustering, because it can reduce the variation of travel times for O-D pairs to a great extent.

(4) Accompanying the initial application to categorical O-D pairs on the Shenzhen metro network, an interesting approach was also provided for determining the theoretical maximum accuracy of our proposed assignment model.

However, some additional issues still need to be addressed. For example, several unusual phenomena during peak periods such as "failing to board on" should be accounted for in the assignment process, and the computational efficiency of the approach should be further improved considering the massive amounts of AFC data and time table data. All the above mentioned is the prospective working the future.

Overall, this study provides a promising approach that can efficiently assign passenger flows on a metro network not only in the common case but also in the case of sudden changes in the time table or disruptions in the metro system

REFERENCES

[1] *J. G .Jin, K. M. Teo, and L. J. Sun, "Disruption response planning for an urban mass rapid transit network," in Proceedings of the 92nd TRB AnnualMeeting,Washington,DC,USA,2013.*

[2] *Y. Sheffi, Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods, Prentice Hall, Inc., EnglewoodCliffs,NJ,USA,1985.*

[3] *M. G. H. Bell and Y. Iida , Transportation Network Analysis, John Wiley&Sons,Chichester,UK,1997.*

[4] *H. Kato, Y. Kaneko, and M. Inoue, "Comparative analysis of transit assignment :evidence from urban railway system in the Tokyo Metropolitan Area," Transportation, vol. 37, no. 5, pp. 775–799,2010.*

[5] *Y. Liu, J. Bunker, and L. Ferreira, "Transit user ′ s route-choice modeling in transit assignment: a review," Transport Reviews, vol.30,no.6,pp.753–769,2010.*

[6] *T. E. Smith, C. -C. Hsu , and Y.-L. Hsu, "Stochastic user equilibrium model with implicit travel time budget constraint," Transportation Research Record ,no.2085,pp.95–103,2008.*

[7] *M. Ben-Akiva and S. R. Lerman , Discret e Choice Analysis: Theory and Application to Travel Demand, MIT Press, Cambridge, Mass,USA,1985.*

[8] *W. Zhu, Research on the model and algorithm of mass passenger flow distribution in network for urban rail transit [Ph.D .thesis]TongjiUniversity,Shenzhen,China,2011.*

[9] *R. Thomas, Traffic Assignment Techniques ,Aldershot :The Academic PublishingGroup,1991.*

[10] *E. Cascetta, Transportation Systems Analysis : Models and Applications, Springer Science & Business Media, New York, NY, USA,2009.*

[11] *M. A. Munizaga and C. Palma, "Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile," Transportation Research Part C: Emerging Technologies, vol.24,pp.9–18,2012.*

[12]*M. Munizaga, F Devillaine, C.Navarrete,andD.Silva,"Validatingtravelbehaviorestimated fromsmartcarddata,"TransportationResearchPartC:EmerginglyTechnologies,vol.44,pp.70–79,2014.*

[13]*C. Morency, M. Tr´ epanier, and B. Agard, "Measuring transit use variability with smart-card data,"TransportPolicy,vol. 14,no.3, pp.193–203,2007.*

[14] *C. Seaborn, J. Attanucci, and N. H. M. Wilson, "Analyzing multimodal public transport journeys in London with smart 10 Discrete Dynamics in Nature and Society smartcard fare payment data," Transportation Research Record, no. 2121,pp.55–62,2009.*

[15] *M. Bagchi and P. R. White, "The potential of public transport smart card data," Transport Policy, vol. 12, no. 5, pp. 464–474, 2005.*

[16] *M. Lehtonen, M. Rosenberg, J.Rasanen, and A.Sirkia, "Utilization of the smart card payment system (SCPS) data in public transport planning and statistics," in Proceedings of the 9th World Congress on Intelligent Transport Systems, Chicago, Ill, USA,2002.*

[17] *M. Utsunomiya, J. Attanucci, and N.Wilson," Potential uses of transit smart card registration and transaction data to improve transit planning," Transportation Research Record: Journal of the Transportation Research Board 1971, Transportation Research Board of the National Academies, Washington, DC, USA,2006.*

[18] *Z. Guoand N.Wilson, "Transfer behavior and transfer planning in public transport systems: a Case Of the London underground,"in Proceedings of the 11 th International Conference on Advanced Systems foPublic Transport,HongKong,2009.*

[19] *M.-P. Pelletier, M. Tr´ epanier, and C. Morency, "Smart card data use in public transit: a literature review, "Transportation Research Part C: Emerging Technologies, vol. 19, no. 4, pp. 557– 568,2011.*

[20] *J. Chan, Rail transit OD matrix estimation and journey time reliability metrics using automated fare data [M.S. thesis],Massachusetts Institute of Technology ,Cambridge ,Mass ,USA, 2007.*

[21] *T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers' train choice behavior with smart card Transportation Engineering, Tongji University,Shanghai,China,2011.*

[23] *Y .SunandR.Xu ,"Rail transit travel time reliability analysis and passenger route choice behavior estimation using automated fare collection data," Transportation Research Record: Journal of the Transportation Research Board 2633, Transportation Research Board of the National Academies, Washington, DC, USA,2012.*

[24]*F. Zhouand R.-H. Xu, "Model of passenger flow assignment for Urban rail transit based on entry and exit time constraints ,"Transportation Research Record,no.2284,pp.57–61,2012.*

[25] *W. Zhu, H. Hu, and Z. Huang, "Calibrating rail transit assignment models with genetic algorithm and automated fare collection data, "Computer-Aided Civil and Infrastructure Engineering, vol.29,no.7, pp.518–530,2014.*

[26] *J. Mac Queen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability ,Berkeley, Calif,USA,1967.*

[27] *L. Kaufman and P. J. Rousseeuw , Finding Groups in Data: An Introduction to Cluster Analysis, Wiley-Inter science ,New York, NY,USA,2009.*

[28] G. J. McLachlan and T. Krishnan, The EM Algorithm and Extensions,Wiley-Interscience,NewYork,NY,USA,2007.

[29] *M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise ,"in Proceedings of the ACM2 and International Conference On Knowledge Discovery and Data Mining ,E. Simoudis, J. Han, and U. Fayyad, Eds., AAAI Press, Menlo Park, Calif,USA,1996.*

[30] *A. Laio and A. Rodriguez, "Clustering by fast search and find of densitypeaks,"Science,vol.344,no.6191,*