Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 318 - 327, 2005

CALCULATION MODEL AND ALGORITHM DESIGN FOR RAILWAY PASSENGER OD-MATRIX

Wei Xiong ZHA
Professor
Institute of Transportation and Economics
East China Jiaotong University
Shuanggang East Road, Nanchang,
330013 China
E-mail: wxzh@ecjtu.jx.cn

Jun ZHANG
Graduate Student
Institute of Transportation and Economics
East China Jiaotong University
Shuanggang East Road, Nanchang,
330013 China
E-mail: zhangjunedward@yahoo.com.cn

Abstract: This paper studies the model and algorithm of railway passenger's OD-matrix estimation, based on actual passenger train information, such as: train manifest, train capacity, route and attendance rate etc. For acquiring the railway passenger's OD-matrix matching with factual traffic flow, this paper utilizes the method of traffic flow allocation suit to railway fieldwork, aim at bias rate between theory traffic flow and factual traffic flow and make use of error correction to iteratively compute, until get a satisfying solution.

Key Words: railway transportation, passenger train, OD-matrix, model and algorithm

1. INTRODUCTION

OD-matrix (Origin-Destination matrix) estimation model, such as increase rate method, gravity model, transportation model and opportunity model, appeared in the 1950s. In past two decades, some scholars put forward many methods to estimate OD-matrix with observation value of traffic flow. According to the ways of matrix allocation, these methods can be classified as proration model and non-proration model. It is suggested that the proportion of travelers choosing one route is independent of the traffic flow in the route in proration model, see e.g. *Duan, J.Y. et al. (2000)* and *Li, J. et al. (2001)*. Being considered the effect of traffic jam in road network, it is suggested that the number of travelers in one route is dependent of the traffic flow in the route in non-proration model, see e.g. *Gao, Z.Y et al. (1990)*, *Zhou, X.Z (2001)* and *Yang, Q. et al. (2002)*. In recent years, there are some new methods to estimate OD-matrix by traffic flow. *Zhang, X.W. et al. (2000)* put forward four-dimensional search method to differentiate the traffic flow between direct flow and indirect flow. *Guo, X.C. et al. (2000)* estimated the cooperative and competitive OD-matrix by URT's joint modal split assignment model. Additionally, hybrid genetic algorithm and artificial nerve network algorithm can be applied in some road net work with simple structure.

Passengers O-D flow survey provides basic data for passengers transportation management and plan. Railway network is one of traffic networks and the OD-matrix estimation is the base to make passenger train plan. However, the factors influencing the railway passenger flows are very complex and some of them can not be quantified. In this case, even if the large-scale passengers OD survey could be conducted throughout the railway network, as costs lots of manpower, material resources and money, the result is not very accurate. However, on the other hand, in China, there are perfect statistics, especially the information on trains running, such as trains composition, the tonnages and the capacity of all kinds of carriages, the route and running kilometrage of each train, and attendance rate of each running section. The statistics are attainable at real time easily by Ticket-Selling Subsystem in Train Management Information System, which are used in railway ticket offices. It reduces the cost of data collection to calculate real-time OD-matrix utilizing the information of passenger trains in existence and it is easy to be organized and carried out. In addition, the statistical precision can be controlled easily.

2. CALCULATING TRAFFIC FLOW UTILIZING PASSENGER TRAINS INFORMATION IN EXISTENCE

Influenced by many factors, the railway passenger transportation is a dynamic system with a variety of passenger demand, and the variety is reflected consequentially in the statistics of passenger trains in existence and in the change of traffic flow. Therefore, how to collect these data becomes the foundation to estimate the OD flow.

Suppose a railway network G with k sections， which are noted as a , $a = 1,2,\cdots,k$ respectively. Additionally, there are n trains passing through this railway network during unit time (one night). A notation is used in which most quantities are expressed as vectors: l is the train passing through the railway network G , and $l = 1,2,\cdots,n$; RW_l, YW_l, RZ_l, YZ_l denote the number of cushioned berth carriages, semi-cushioned berth carriages, cushioned seat carriages, semi-cushioned seat carriages formatted in train l , respectively; $RWN_l, YWN_l, RZN_l, YZN_l$ denote the capacity of cushioned berth carriage, semi-cushioned berth carriage, cushioned seat carriage, semi-cushioned seat carriage in train l, respectively; S_l is the set of train routs, $l = 1,2,\cdots,n$; v_{al}^{rw}, v_{al}^{yw}, v_{al}^{rz}, v_{al}^{yz} denote the attendance rate of cushioned berth carriages, semi cushioned berth carriages, cushioned seat carriages, semi-cushioned seat carriages in train l on section a , respectively and $a = 1,2,\cdots,k$, $l = 1,2,\cdots,n$.

Set $$\delta_{al} = \begin{cases} 1 & \text{if } a \in S_l \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

The passenger traffic flow a is described by the following equation:

$$f_a = \sum_{l=1}^{n} (RW_l \times RWN_l \times v_{al}^{rw} + YW_l \times YWN_l \times v_{al}^{yw} + RZ_l \times RZN_l \times v_{al}^{rz} + YZ_l \times YZN_l \times v_{al}^{yz}) \delta_{al}$$

$$a = 1,2,\cdots,k \tag{2}$$

In practice, the train composition may be more complicated or there are more calculation steps, but the calculating principle is similar. For the sake of convenience, the following formula can be used:

$$f_a = \sum_{l=1}^{n} DY_l \times \bar{v}_{al} \times \delta_{al} \tag{3}$$

where $\bar{v}_{al}$ is the mean attendance rate of train l on section a ($a = 1,2,\cdots,k$, $l = 1,2,\cdots,n$); DY_l is the capacity of train l ($l = 1,2,\cdots,n$).

To a concrete railway network, all data of passenger train plans are saved in the computer database and compute automatically. The steps are described as follows:

Step0: Prepare data; read data of network structure and trains, such as train composition, the capacity of each kind of carriages, route of each train and attendance rate on each running section. Let $l = 1$, $f_a = 0$, $a = 1,2,\cdots,k$

Step1: $a = 1,2,\cdots,k$ if $a \in S_l$, then: $f_a = f_a + DY_l \times \bar{v}_{al} \times \delta_{al}$

Step2: If $l < n$, then $l = l + 1$, go to Step 1; else go to Step3

Step3: Record all the results to database and end.

3. RAILWAY PASSENGER OD ALLOCATION MODEL AND ALGORITHMS

To sum up, all theories and methods which have been used in calculating OD-matrix from traffic flow on sections can be classified as parameter estimating, matrix estimation method—such as maximum entropy, minimum information, equilibrium method and statistic estimating--and neural network method. The calculation of OD-matrix from traffic flow-volume is the inverse process of traffic allocation. Here, based on the features of running route selection in railway transportation, the passenger O-D flow on railway network can be deduce by establishing the traffic allocation model with the characters of railway.

Suppose a railway network G with k links, which are noted as a, $a = 1,2,\cdots,k$ respectively. The passenger flow-volume on section a is f_a, $a = 1,2,\cdots,k$. The set of stations on the network is denoted with $Z = (z_1, z_2, \cdots, z_n)$. $E(i,j)$ is the adjacency matrix of point-pares (z_i, z_j), $ED(i,j)$ is the length between adjacent stations on section (z_i, z_j), and $EA(i,j)$ is noted as the passing-through capacity on this section. $E(i,j)$ is described as:

$$E(i,j) = \begin{cases} 1 & z_i, z_j \text{ are adjacent stations and the section is single - track} \\ 2 & z_i, z_j \text{ are adjacent stations and the section is double - track} \\ +\infty & z_i, z_j \text{ are not adjacent stations} \end{cases} \tag{4}$$

3.1 Railway Passenger O-D Allocation Model

According to the different hypotheses based on, the traffic allocation model can be described in different formations and thereby, the different allocation results can be got. Consequently, before making an allocation model, it is necessary to make some hypotheses which are as consistent with the practice as possible. It is out of question that if the hypotheses accord with the practice better and the factors are considered more entirely, the model will be more accurate. The allocation of highway and urban traffic deals with the personal activities and because of the diverse benchmarks, passengers often select the routes subjectively with uncertain information. As the practice is, Ministry of Railways takes it for granted that the selection of the optimum route from origin to destination does not have to include any route in the network but consider the reasonable routes first all, that is, to select the shortest route, and then the shorter one, etc, which are called bypass routes.

However, the selection of bypass route has to be followed one principle that there is some restrictions on the bearable excessive distance and time on the route selected compared with the shortest route, that is, if the cost of bypass route is much more than the shortest route, this bypass route will not be accepted. Hence, we hope that compared with the shortest route, the excessive cost of bypass route be restricted within an acceptable and reasonable range.

Out of convenience, the shortest route of O-D to (z_i, z_j) is noted as $l_0(i,j)$, and other routes are noted as $l_1(i,j)$, $l_2(i,j)$,…….The mechanism is described by the following equation.

$$\frac{l_k(i,j) - l_0(i,j)}{l_0(i,j)} \le L \tag{5}$$

where: $k = 1,2,\cdots$, L is proportion factor which means the ratio of the excessive distance of bypass route to the shortest route should be less than some scale. In practice, it can be used as follows:

$$l_k(i,j) \le (1+L) \times l_0(i,j) \quad k = 1,2,\cdots \tag{6}$$

It is expected that the route passenger trains running on be the most reasonable and economical, and had best, one of the same-shortest routes. However, in practice, there are no real same-shortest routes. Therefore, generally, the routes with approximate length can be regarded as the same-shortest routes. That inspires us that by partitioning the ratio L, we can research on how to optimize the train route in the network hierarchically.

Set an ascending series $\{L_0,\ L_1, \cdots,\ L_N\}$, and $0 = L_0 < L_1 < \cdots\cdots < L_N = L$. That is, partition L, the ratio of excessive length for bypass. First, dispatch the train flow on the shortest route. Second, dispatch the left train flow if $l_k \le (1+L_1)l_0$. Namely, the ratio of excessive length to the shortest route length is less than L_1. Third, dispatch the left train flow if $l_k \le (1+L_2)l_0$. Such-and-such, end till N. L can be partitioned into arithmetical progression or geometric progression.

(a) Arithmetical Progression Method:

$$L_i = \frac{i \times L}{N} \qquad i = 1,2,\cdots,N \tag{7}$$

(b) Geometric Progression Method:

$$L_i = \alpha^i \qquad \text{where} \ : \ \alpha = \sqrt[N]{L} \qquad i = 1,2,\cdots,N \tag{8}$$

Empirically, we have taken use of the two partition methods and got satisfying results. In addition, the selection mode of geometric progression method is more consistent with our original intention of the priority of the same-shortest routes.

3.2 Algorithm Design for Railway Passenger O-D Allocation

The essential thought is as follows. First, dispatch the train flow on the shortest route and allocate the flow step by step until all the point-pairs are considered. Second, initialize the status parameters and train flow-volume according to the result. On some condition, consider the bypassing mode according to formula (7) and (8), the calculation process of which is to consider the shortest route and allocate flow step by step based on the initialized parameter and train flow-volume. Third, based on the calculation result, initialize the status parameter of network and train flow again and consider the bypassing mode by the method of the shortest route and allocating flow step-by-step. Such-and-such, end till there is no train flow to be dispatched or no optional bypassing mode. The concrete steps are as follows:

Step0: Set $N = 1, ZF = 0.0, T_1 = \Phi, T_2 = \Phi$ and input L, α and matrix $E(i,j)$, $ED(i,j)$, $EA(i,j)$;

Step 1: Evaluate the shortest route from any point to (z_i, z_j) and its length and save them into matrices P and D;

Step 2: Evaluate L based on formula (7) or (8). To all the point-pares of which the flow is non-zero and the shortest route exists, select some of them that satisfy $D_1(i,j) \le (1+L) \times D(i,j)$ to make set S_1. If $S_1 = \Phi$, then go to Step 6.

Step 3: In set S_1, select the point-pares of which the route are not changed to make set S_2. If $S_2 = \Phi$, then $N = N+1, D = D_1$, go to Step 2. Or take the point-pare (z_s, z_t) with the longest route as the object.

Step 4: Evaluate $\min l$, which is the minimum of the passing capacity on the route of point-pare (z_s, z_t).

If $f(z_s, z_t) \ge \min l$, take $f(z_s, z_t) - \min l$ as the accession flow of this section and save it into matrix F. At the same time, delete all the points and links with saturated capacity to make a new network and go to Step 5.

If $f(z_s, z_t) < \min l$, take $f(z_s, z_t)$ as the accession flow of this section and save it into matrix F. At the same time, subject the corresponding value from points and links. Let $ZF = ZF + f(z_s + z_t)$ and go to Step2.

Step 5: Calculate the shortest route and its length in new network and save them into matrix P_1, D_1 and then go to Step 2.

Step 6: Calculate the employment capacity of every point and link, evaluate the distribution of passenger flow and the train flow volume among the point-pairs, print and end.

3.3 Analyses of Algorithm

In Step0 and Step5, the shortest route between any two crunodes can be come out by Floyd's Algorithm and get the time-boundary $O(n^3)$. In Step2 and Step3, the time-boundary is $O(n^2)$ and $O(n\log(n))$ in Step4. Because the outcome of Step4 is that the passenger flow- volume is zero between the point-pairs or to delete the points or links to make a new network, at the worst, Step2、Step3、Step4、Step5 will be iterated for $2n^2+n$ times at best. Therefore, the complexity of the algorithm is $O(n^5)$.

The features of this algorithm are described as follows.

First, it makes the algorithm structure simple, perspicuous and programmable by dispatching the train flow by normal route and bypassing route hierarchically. Additionally, in actual railway network, there are a few optional bypassing routes, so partition with 5 hierarchies will meet our demand. At the same time, out of the calculating results, we can see clearly the matching status between the network and passenger flow-volume.

Second, to this algorithm, passenger flow-volume corresponds to our demand on the conveyance capacity of railway network. Consequently, the conveyance capacity of railway network and the demand met can be reflected in the calculation outcome rather than only a simple conclusion that plan is not feasible.

Third, whether dispatch train flow on the shortest route or on the bypassing route bases on the principal of being economical, reasonable and optimizing system, which corresponds to the practice. Simultaneously, for the compact data structure and clear hierarchies, the algorithm makes the train composition plan-making expedient. In additionally, with small storage while large quantity of information, it is advantageous to solve the complex problem using computer.

Forth, calculation is efficient and with a high speed. Out of the tens of cases with 5-15 crunodes, networks and passenger flow-volume, all of which are set up discretionally, it takes not less than tens of seconds to compute and the outcome is satisfying.

4. CALCULATION MODEL OF RAILWAY PASSENGER OD-MATRIX BASED ON TRAFFIC ALLOCATION

4.1 Inverse Calculation Model of Passenger OD Flow-Volume

It is an inverse process of traffic allocation to calculate passenger OD flow-volume from traffic flow-volume on sections. Out of convenience, it is described as matrices for the calculation model of passenger OD flows. The elements of OD-matrix are noted as T_r ($r = 1,2,\cdots,K$) and the traffic flows allocated on all sections are noted as $f_a(a = 1,2,\cdots,N)$. To calculate the OD flows, it is necessary to make sure that the bias between the academic allocating traffic flow-volume and actual traffic flow-volume is the least, which is described as follow:

$$\min Z = \sum_a (f_a - f'_a)^2 \tag{9}$$

Where: f_a and f'_a are academic allocating traffic flow-volume and actual traffic flow-volume on section a.

4.2 Algorithm Design for Passenger OD Flow-Volume

Based on definite allocation method, to calculate the passenger OD flows matching with actual traffic flow volume, we introduce two concepts-bias rate and modification rate-to carry out iterative inverse-calculating.

The bias rate is the contrast between academic allocating traffic flow-volume and actual traffic flow-volume on section a for the kth time of iteration calculation, which is denoted by D_a^k, i.e.:

$$D_a^k = f'_a / f_a^k \ , a = 1,2,\cdots,N \tag{10}$$

Where: f_a^k is academic allocating traffic flow-volume on section a for the kth time of iteration calculation. S_r^l is the section set of rth OD point pair on the lth route, which is described as:

$$\delta_{a,r}^l = \begin{cases} 1 & \text{if } a \in S_r^l \\ 0 & \text{otherwise} \end{cases} \quad a = 1,2,\cdots,N \ , r = 1,2,\cdots,K \ , l = 1,2,\cdots,M \tag{11}$$

$t_{r,l}^k$ is the allocated flow-volume of the rth OD point pare on the lth route for the kth time of iteration calculation.

$$P_{a,r}^k = \sum_{l=1}^{M} t_{r,l}^k \delta_{a,r}^l \qquad a = 1,2,\cdots,N \ , r = 1,2,\cdots,K \tag{12}$$

E_r^k is the modification rate of the rth OD point pare for the kth time of iteration calculation, and is described by the following equation:

$$E_r^k = \sum_{a=1}^{N} P_{a,r} D_a^k / \sum_{a=1}^{N} P_{a,r} \tag{13}$$

The calculation steps of the inverse-deduction model are as follows:

Step0: Give an initial OD-matrix T_r^0 and the precision requirement ε and set $k = 0$;
Step1: Evaluate f_a^k by traffic allocation model;
Step2: Evaluate bias rate $D_a^k = f'_a / f_a^k$, $a = 1,2,\cdots,N$;

Step3: If the bias rate on every section satisfies

$$\sum_{a=1}^{N} \left|1 - D_a^k\right| < \varepsilon \ , \tag{14}$$

go to Step1;

Step4: Calculate modification rate

$$E_r^k = \sum_{a=1}^{N} P_{a,r}^k D_a^k / \sum_{a=1}^{N} P_{a,r}^k \ , r = 1,2,\cdots,K \tag{15}$$

and

$$T_r^{k+1} = E_r^k T_r^k \ , r = 1,2,\cdots,K \text{ . Set } k = k+1 \text{ and go to Step1} \tag{16}$$

Step5: Output calculation outcome of T_r^k , $r = 1,2,\cdots,K$.

4.3 Algorithm Analyses

Diverse initial OD-matrixes will influence the outcome of passenger OD flows deduction, so a good initial matrix is helpful to carry out the algorithm. On the basis of the known passenger train information, a satisfied initial value of passenger OD flows can be achieved.

If there are L pairs of origin and destination trains between the rth point-pair, the train capacity of the lth pair of trains is DY_r^l and the attendance rate along the running route is $\overline{v}_r^l$, the initial value is described as follow:

$$T_r^0 = \sum_{l=1}^{L} DY_r^l \overline{v}_r^l \tag{17}$$

If there is no OD point-pair of origin and destination trains, the initial value can be set any non-negative.

5. COMPUTER REALIZATION OF RAILWAY PASSENGER OD INVERSE CALCULATION

Owing to the large-scale railway network and complex train running information, utilizing present information of passenger trains, we have to take use of computer to calculate the passenger OD flows. The following chart is the flow of estimating railway passenger OD-matrix based on traffic allocation.

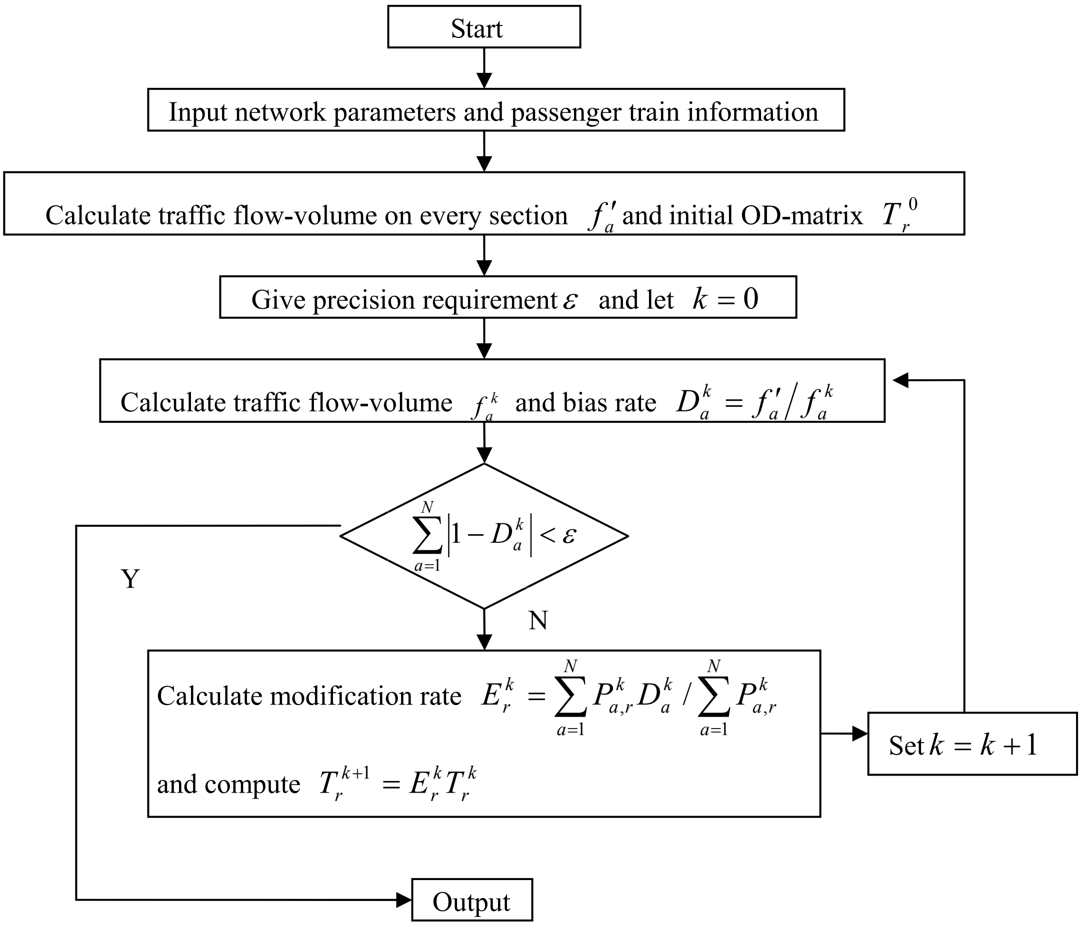

Figure1. Flow Chart of Inverse Calculation of Passenger O-D Flow

6. SUMMARY

The estimation of railway passenger OD-matrix is of importance to railway traffic management and plan, which provides the basic data for passenger train plan. On the condition of market economy, it is a principle to make passenger train plan by flow-volume. However, taking use of OD flows survey frequently is neither feasible nor economical. Additionally, utilizing mathematics model to estimate passenger OD matrix, many researches are focus on highway traffic, especially urban traffic instead of railway network. The theory and method put forward in this paper have been applied in the development of Passenger Train Plan Decision System in Zhengzhou Railroad of China. The system developed has been applied by the railroad for one year and the result is satisfying.

ACKNOWLEDGEMENTS

This research was funded by the Railways Ministry of China under Program of Science and

Technology Development. We appreciate the valuable research assistance of Jian Li, Mao-sheng Zhong and Ji-ru Wang for a great deal of work on the test of algorithms throughout the research.

REFERENCES

Guo, X. C and Lv, S. (2000) Study of URT'S Joint Model Split Assignment Model on Cooperative and Competitive OD Matrix, **China Journal of Highway and Transport, Vol. 13, No. 1,** 91-94.

Li, J. and Peng, G.X. (2001) A Study on Estimating OD Trips From Observed Link Flows-Based on the Method of Iterative Procedure to Calculate OD on the Multi-path Assignment Model, **Journal of Traffic and Transportation Engineering, Vol. 1, No. 2**,78-82.

Gao, Z.Y. and Shi, B.F. (2002) A New Algorithm for Estimation of Origin—Destination Demands from Link Traffic Flows, **Transportation Systems Engineering and Information Technology, Vol.2, No.1,** 30-41.

Zhang, W.X., Hu, L.G. and Guo, X.Y. (2000) Four-dimensional Search Method-A solution for OD Traffic Flow, **Systems Engineering**, **Vol.18, No.5**, 40-43.

Duan, J.Y., Miao, L.X. and Jiang, J.J. (2000) Application of Entropy Maximization Model for Estimating OD Trip Matrices, **Journal of Tsinghua University, Vol.40, No.6,** 123-126.

Zhou, X.Z. (2001) Combinatorial Model and Algorithm Involving OD Distribution and Stochastic User Equilibrium Assignment, **Journal of Systems Engineering, Vol.16, No.3,** 197-201.

Yang, Q. and Wang, W. (2002) OD Matrix Capacity Calculation Method, **Journal of Highway and Transportation Research and Development, Vol.19, No.2, 9-13.**