

Estimation of Origin-Destination Matrix in Fuzzy Environment

Abdushukoor K M¹ and Sushama C M

Department of Mathematics, National Institute of Technology Calicut, India.

¹Corresponding author: kmsbukoor@gmail.com

Received 5 January 2019; accepted 12 February 2019

Abstract. An Origin-Destination (O-D) matrix provides information about the number of travel trips made between different O-D pairs located within a road network for a given region. The reliability of the estimated O-D matrix from traffic counts depends much on the quality and quantity of the available input data. Most of the proposed estimation models in literature require information about prior trip table between O-D pairs. This paper develops a new methodology for estimating Origin-Destination matrix by extensive use of fuzzy set theory in order to incorporate the uncertainty embedded in driver's route choice behavior and inconsistencies in traffic counts. Computational results on sample network from the literature are presented to demonstrate the methodology developed. Error measures are also computed for establishing the quality of O-D matrix.

Keywords: Transportation network, trip generation, traffic counting location, triangular fuzzy numbers, integer program

AMS Mathematics Subject Classification (2010): 90B20, 90C10, 03E72

1. Introduction

Estimation of the Origin-Destination (O-D) trip matrix plays an important role in describing travel analysis and transportation planning in a region. Traditional methods of estimating O-D matrix include road side interviews, household surveys and license plate surveys, which are all costly and time consuming. In recent decades, there is a substantial interest is developed in the field of estimating O-D matrix from traffic counts. The fundamental problem related to O-D matrix estimation is the unavailability of prior O-D Matrix, and the traffic counts from links alone are insufficient to obtain the required O-D matrix.

Yang and Zhou [2] conducted a comprehensive investigation based Maximal Possible Relative Error(MPRE) to determine the optimal number and locations of traffic counting stations and suggested 'Four location rules', O-D covering rule, Maximal flow fraction rule, Maximal flow interception rule and Link independence rule. The problem is formulated as an integer programming problem, where the O-D covering rule and link independence rule are considered as constraints and the total flow observed is taken as the objective function to be maximized. The major drawback for O-D covering approach is that it considers only one path (probably shortest path) between each O-D pair. If no path flow information or turning probabilities are available, Yang et al. [3] introduced

Abdushukoor K M and Sushama C M

screen-line based approach for finding optimal traffic counting locations, which is based on O-D 'separation rule'. But this approach may not be realistic for practical applications due to budgetary constraints of the traffic management agencies. Larsson et al. [11] established that the most commonly used approach for detector allocation, maximizing the coverage of O-D pair seems to be unfavorable for the quality of the estimated O-D matrix whereas Maximum flow coverage detector allocation method gives surprisingly good result in most of the experiments.

An important problem related O-D matrix estimation process is the inconsistency among traffic count data, which results in poor quality of the estimated O-D matrix. Data inconsistency occurs for diverse reasons, such as counting error and driver behavior. Xu and Chan[12] tried to solve this problem by using a set of 'Fuzzy Weights' for each piece of inconsistent data. They considered the link counts as imprecise values, so they proposed to survey link volumes more than once and computed fuzzy weights for every set of link counts.

Traffic counts from different locations has different degree of influence to the O-D matrix estimation. Hence it is significant to develop a systematic selection of optimum links from the network. Abdushukoor and Sushama [5] introduced a fuzzy graph approach for selecting optimal number of traffic counting stations and its locations in road networks. This Model uses a screen-line based approach to cover all paths having a particular flow or higher, in order to attain maximum flow coverage. As the input data on prior link flows may not be very reliable in the present context, it seems quite appropriate to use the concept of fuzzy graph theory[8, 9] to determine the above mentioned flow. So the weights are assigned to links (arcs) based on prior link flow data and accordingly weights of the paths are found out. α -strong paths as well as α -strength reducing set of arcs for each O-D pair is also determined. Based on this concept, two integer programming models are developed to determine

- 1) the number of arcs in any minimum α -strength reducing set of arcs of the O-D pairs
- 2) strong weight and identification of the corresponding α -strength reducing set of arcs of the pair.

The traveler's selection of route between the Origin-Destination pair is based on many factors: travel time/cost, average travel speed, previous experience, type and condition of road, etc. It is difficult to predict the route, because traveler may choose a route that has minimum number of signalized intersections or other reasons irrespective of the actual travel cost. Multipath optimization models may give better result than that of single path (shortest path) models [12], because it is clear the fact that there are paths other than shortest may be traversed by the travelers. This paper discussed the multipath optimization model, for the purpose of showing the reliability of estimation the entire link count volume is assumed to be known. We discuss the challenges of estimating O-D matrix with incomplete and imprecise data, a new fully fuzzy model from traffic counts is introduced. The proposed model is potentially useful when the prior O-D matrix is unavailable, requires the amount of trips generated at origins (source) and terminates at the destinations(sinks). It also requires the traffic counts from major links (optimal links).

The rest of the paper is organized as follows: The basic concepts about fuzzy set theory, triangular fuzzy numbers and its arithmetic are discussed in section 2, formulation of O-D matrix problem in section 3, the proposed fully fuzzy programming model is in

section 4, the method of solving is presented in section 5, numerical illustration with error measures are in section 6 and the concluding remarks are given in section 7.

2. Triangular fuzzy number and its arithmetic

Definition 2.1. A fuzzy number $\tilde{A} = (a^L, a, a^U)$ is said to be a triangular fuzzy number [7] if its membership function is

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b, \\ \frac{x-c}{b-c}, & b \leq x \leq c, \\ 0, & \text{otherwise} \end{cases}$$

Let $\tilde{A} = (a^L, a, a^U)$ and $\tilde{B} = (b^L, b, b^U)$ be two triangular fuzzy numbers, then [6]

- (i) Addition: $\tilde{A} \oplus \tilde{B} = (a^L, a, a^U) \oplus (b^L, b, b^U) = (a^L + b^L, a + b, a^U + b^U)$,
- (ii) Subtraction: $\tilde{A} - \tilde{B} = (a^L, a, a^U) - (b^L, b, b^U) = (a^L - b^L, a - b, a^U - b^U)$,
- (iii) Equality: $\tilde{A} = \tilde{B} \Leftrightarrow a^L = b^L, a = b, a^U = b^U$,
- (iv) Non-negativity: $\tilde{A} = (a^L, a, a^U) \geq 0$ if and only if $a^L \geq 0$,
- (v) Multiplication: Let $\tilde{A} = (a, b, c)$ be any triangular fuzzy number and $\tilde{B} = (x, y, z)$ be any non-negative triangular fuzzy number, then

$$\tilde{A} \otimes \tilde{B} = \begin{cases} (ax, by, cz), & a \geq 0 \\ (az, by, cz), & a < 0, c \geq 0 \\ (az, by, cx), & c < 0 \end{cases}$$

Definition 2.2. A ranking function is a function [10] $R : F(R) \mapsto R$, where $F(R)$ is a set of fuzzy numbers defined on set of real numbers, which maps each fuzzy number into the real line, where a natural order exists. Let $\tilde{A} = (a^L, a, a^U)$ be a triangular fuzzy number, then $R(\tilde{A}) = \frac{a^L + 2a + a^U}{4}$.

3. Formation of OD matrix estimation model

In order to introduce the model to predict the OD matrix and link flows, we consider an urban road network $G(N, A)$. Let N denotes set of nodes which represents the actual traffic intersections or zones which generates/terminates the trips, A denotes the corresponding directed arcs/links between pairs of nodes. Let O denotes the given origin nodes (source) in N and D denotes the given destination nodes (sink) in N . Let OD be the set of possible origin-destination pair so that the trip table to be estimated.

The scope of this OD matrix estimation model is that it does not requires prior OD matrix (most of the real life situation, the prior OD matrix will be unavailable). The model requires the amount of trips generated at origins and terminates at the destinations, observed link counts from major arcs (optimal traffic counting stations using screen-line approach) and the free-flow travel cost on links. Let A_v represents the set of links in the

Abdushukoor K M and Sushama C M

traffic network having traffic counting stations(link counts are known) A_u represents the set of links for which the link counts are unknown with $A = A_v \cup A_u$. Let f_a denotes the observed traffic counts for $a \in A_v$. Let O_i represent the total number of trips produced at origin O_i and D_j represent the total number of trips arrives at destination j .

To formulate the model, the following notations are also used. Let n_{ij} denote the number of possible distinct paths between the OD pair (i, j) . P_{ij}^k represent the k^{th} least-cost path between the pair (i, j) and $(P_{ij}^k)_a = 1$, if the k^{th} path between OD pair (i, j) contains the arc a , and $= 0$, otherwise. Let y_{ij}^k denote the path flow through the path k from the origin node i to the destination node j for $k = 1, 2, \dots, n_{ij}$.

The following are the balance equations to be fulfilled:

Balance equation for trip generation:

$$\sum_{j \in D} \sum_{k=1}^{n_{ij}} y_{ij}^k = O_i, \forall i \in O \quad (1)$$

where O_i denotes the total number of trips generated at origin node i .

Balance equation for trip attraction:

$$\sum_{i \in O} \sum_{k=1}^{n_{ij}} y_{ij}^k = D_j, \forall j \in D \quad (2)$$

where D_j denotes the total number of trips attracted to the destination node j .

The OD matrix estimation from equilibrium assignment is based on Wardrop's principle (Wardrop - 1952) under the assumption that when OD flow is assigned, it will reproduce the observed link flows.

Balance equation for observed link counts:

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a y_{ij}^k = f_a, \forall a \in A_v \quad (3)$$

where f_a is the observed link counts for the link $a \in A_v$ and n_{ij} denote the number of possible distinct paths between the OD pair (i, j) .

$$y_{ij}^k \geq 0, \forall (i, j) \in OD \text{ and } k = 1, 2, \dots, n_{ij}. \quad (4)$$

The perceived travel cost on k^{th} route between the OD pair is the sum of travel costs of its intermediate links. For the link $a \in A$, the travel cost C_a is proportional to the flow f_a on link a , which can be obtained using the formula suggested by Bureau of Public Roads (BPR, 1964):

$$c_a(f_a) = c_a^F [1 + 0.15 \left(\frac{f_a}{u_a} \right)^4], a \in A \quad (5)$$

where c_a^F is the free flow travel cost on link $a \in A$ and u_a is the capacity of link $a \in A$. Let (c_{ij}^k) denote the travel cost on route k from the origin i to the destination j for each $k = 1, 2, \dots, n_{ij}$ and

$$c_{ij}^k = \sum_{a \in A} (p_{ij}^k)_a c_a(f_a), (i, j) \in (OD), k = 1, 2, \dots, n_{ij} \quad (6)$$

The objective function for the multi-path trip assignment model involves the modified route cost subject to the equality constraints (1) to (4).

Estimation of Origin-Destination Matrix in Fuzzy Environment

$$\text{Minimise } \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} c_{ij}^k y_{ij}^k \quad (7)$$

All the above conditions are fulfilled by solving the following linear programming problem in which the path flows are the decision variables:

Minimize

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} c_{ij}^k y_{ij}^k \quad (8)$$

subjected to

$$\begin{aligned} \sum_{j \in D} \sum_{k=1}^{n_{ij}} y_{ij}^k &= O_i, \forall i \in O \\ \sum_{i \in O} \sum_{k=1}^{n_{ij}} y_{ij}^k &= D_j, \forall j \in D \\ \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (P_{ij}^k)_a y_{ij}^k &= f_a, \forall a \in A_v \\ y_{ij}^k &\geq 0, \forall (i,j) \in OD \text{ and } k = 1, 2, \dots, n_{ij}. \end{aligned} \quad (9)$$

4. Estimation of OD matrix in fuzzy environment

The OD matrix estimation models are formulated based on the availability of input data. The inconsistency among the number of trips generated/destined at various zones and link traffic counts data results in poor quality of the estimated OD matrix and cause of the non-convergence of solution algorithm.

In this paper, we introduce a fully fuzzy linear model to estimate the Origin-Destination matrix that incorporate the vagueness/impreciseness in the variables and parameters involved. To represent the imprecise nature of the O_i , D_j , c_{ij}^k and f_a values, we use triangular fuzzy numbers denoted by \tilde{O}_i , \tilde{D}_j , \tilde{c}_{ij}^k and \tilde{f}_a respectively. Each of these fuzzy numbers are expressed as $\tilde{x} = (x^L, x, x^U)$, where $x^L = x(1-\alpha_r)$ and $x^U = x(1+\alpha_r)$, represents the lower and upper boundaries and α_r is the experimental value represent the impression in traffic counts. All the conditions mentioned earlier are imposed with fuzzy variables and fuzzy numbers. The resulting fully fuzzy linear programming problem (FFLPP) in which the path flows are the fuzzy decision variables are given below:

Minimize

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k \tilde{y}_{ij}^k \quad (10)$$

subjected to

$$\sum_{j \in D} \sum_{k=1}^{n_{ij}} \tilde{y}_{ij}^k = \tilde{O}_i, \forall i \in O$$

$$\begin{aligned} \sum_{i \in O} \sum_{k=1}^{n_{ij}} \tilde{y}_{ij}^k &= \tilde{D}_j, \forall j \in D \\ \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a \tilde{y}_{ij}^k &= \tilde{f}_a, \forall a \in A_v \\ \tilde{y}_{ij}^k &\geq 0, \forall (i,j) \in OD \text{ and } k = 1, 2, \dots, n_{ij}. \end{aligned}$$

5. Solution procedure for FFLPP

Now we begin with transforming the FFLPP (10) into equivalent crisp (conventional LP model. Kumar et al. [1] proposed a fuzzy linear programming method for finding fuzzy optimal solution of the following type FFLPP:

$$\text{Minimise (or Maximise)} (\tilde{C}^T \otimes \tilde{X}) \quad (11)$$

$$\text{subject to } \tilde{A} \otimes \tilde{X} = \tilde{b}$$

\tilde{X} is a non-negative fuzzy number.

Represent the parameter $\tilde{c}_{ij}^k = (c_{ij}^{kL}, c_{ij}^k, c_{ij}^{kU})$, $\tilde{O}_i = (O_i^L, O_i, O_i^U)$, $\tilde{D}_i = (D_i^L, D_i, D_i^U)$, $\tilde{f}_a = (f_a^L, f_a, f_a^U)$ and the non-negative fuzzy path flow variable $\tilde{y}_{ij}^k = (y_{ij}^{kL}, y_{ij}^k, y_{ij}^{kU})$. Using arithmetic operations and by converting, without loss of generality the equations (10) can be rewrite into crisp linear programming problem as follows:

Minimize

$$R \left(\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (c_{ij}^{kL}, c_{ij}^k, c_{ij}^{kU}) \otimes (y_{ij}^{kL}, y_{ij}^k, y_{ij}^{kU}) \right) \quad (12)$$

subject to

$$\begin{aligned} \sum_{j \in D} \sum_{k=1}^{n_{ij}} y_{ij}^{kL} &= O_i^L, \forall i \in O & \sum_{i \in O} \sum_{k=1}^{n_{ij}} y_{ij}^{kU} &= D_j^U, \forall j \in D \\ \sum_{j \in D} \sum_{k=1}^{n_{ij}} y_{ij}^k &= O_i, \forall i \in O & \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a y_{ij}^k &= f_a^L, \forall a \in A_v \\ \sum_{j \in D} \sum_{k=1}^{n_{ij}} y_{ij}^{kU} &= O_i^U, \forall i \in O & \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a y_{ij}^k &= f_a, \forall a \in A_v \\ \sum_{i \in O} \sum_{k=1}^{n_{ij}} y_{ij}^{kL} &= D_j^L, \forall j \in D & \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a y_{ij}^k &= f_a^U, \forall a \in A_v \\ \sum_{i \in O} \sum_{k=1}^{n_{ij}} y_{ij}^k &= D_j, \forall j \in D & & \\ y_{ij}^k - y_{ij}^{kL} &\geq 0, y_{ij}^{kU} - y_{ij}^k &\geq 0, \forall (i,j) \in OD \text{ and } k = 1, 2, \dots, n_{ij}. & \end{aligned} \quad (13)$$

Note that after solving this crisp linear problem, equation (10) gives the fuzzy optimal value for the objective function.

Estimation of Origin-Destination Matrix in Fuzzy Environment

Finally the required OD trip tables entries \tilde{T}_{ij} and link flows \tilde{f}_a are related to the path-decomposed flows \tilde{y}_{ij}^k via relationship:

$$\tilde{T}_{ij} = \sum_{k=1}^{n_{ij}} \tilde{y}_{ij}^k, \text{ for each } (i, j) \in OD. \quad (14)$$

$$\tilde{f}_a = \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a \tilde{y}_{ij}^k, \forall a \in A, \quad (15)$$

To assess the quality of estimated OD flows and link flows Sherali et al., [4] introduced two statistics to measure the closeness are the Root Mean Square (RMSE) and the Mean Absolute Error (MAE) are given by equations (16) and (17) respectively.

$$RMSE = \sqrt{\frac{\sum_{a \in A} (f_a - \bar{f}_a)^2}{n}}, \quad (16)$$

$$MAE = \frac{\sum_{a \in A} |f_a - \bar{f}_a|}{n} \quad (17)$$

6. Test problem

To demonstrate the model, consider a hypothetical network depicted in figure 1, the network consists of 4 zones (2 origins and 2 destinations) 9 nodes (1 to 9), 4 Origin-Destination pairs and 14 directed links (1 to 14) with free flow travel time (fftt) and capacity are assumed to be known.

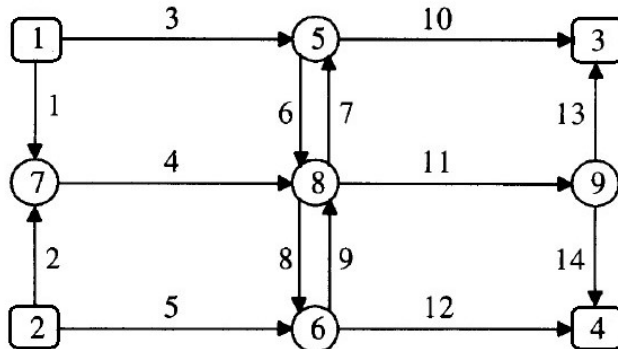


Figure 1: A small directed network for numerical analysis.

Link index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
fftt	27	86	23	19	38	51	51	43	43	71	10	13	83	17
Capacity	750	750	750	1000	750	1000	1000	1000	1000	750	1000	750	750	750

Table 1: Data for the network example

Abdushukoor K M and Sushama C M

As a first step identification of possible paths (there are 16 paths) between OD pairs and prioritized based on route cost. The remaining input data required is the supply/demand at zones $(\tilde{O}_1, \tilde{O}_2, \tilde{T}_1, \tilde{T}_2) = ((730, 750, 770), (340, 350, 360), (245, 250, 255), (825, 850, 875))$ and traffic counts from major arcs $A_v = \{3, 11, 12\} = ((375, 385, 395), (485, 500, 515), (420, 435, 450))$. The multi-path flow model considers all the 16 paths between OD pairs. The total system cost produced by FFLP model is (83673.4, 90816.93, 98234.47).

O-D Pair	Path	Estimated path flow between the OD pairs
1-3	3-10	(165, 165, 165)
1-3	3-6-11-13	(80, 85, 90)
1-4	1-4-11-14	(355, 365, 375)
1-4	3-6-11-14	(50, 50, 50)
1-4	3-6-8-12	(80, 85, 90)
2-3	5-9-11-13	(0, 0, 0)
2-4	5-12	(340, 350, 360)

Table 2: Estimated path flow between the origin-destination pairs.

O-D Pair	Estimated OD flow between the pairs
T13	(245, 250, 255)
T14	(485, 500, 515)
T23	(0, 0, 0)
T24	(340, 350, 360)

Table 3: Estimated origin-destination flow.

The required OD trip tables entries \tilde{T}_{ij} and link flows \tilde{f}_a can be obtained via the relationship (14) to (15). For the purpose of showing the reliability of estimation, the entire link volume is assumed to be known

$$(f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}, f_{13}, f_{14}) = (365, 40, 385, 405, 310, 245, 25, 285, 165, 160, 500, 435, 85, 415)$$

Measures of closeness	All the link counts are considered
RMSE(%)	6.5780
MAE(%)	3.5064

Table 4: Range of estimated error measures

7. Conclusion

In this paper, we have presented a new methodology for estimating Origin-Destination matrix by extensive use of fuzzy set theory to in order to incorporate the uncertainty embedded in driver's route choice behavior and inconsistencies in traffic counts. The advantage of this proposed model is that both the estimated OD flow and link flows are obtained in terms of triangular fuzzy numbers. Finally, for evaluating the reliability, the error measures MPRE and MAE of estimated link flow are calculated, which can be

Estimation of Origin-Destination Matrix in Fuzzy Environment

useful to find the optimal number of traffic counting stations and its locations in such a way as to minimize the relative errors.

REFERENCES

1. A.Kumar, J.Kaur and P.Singh, A new method for solving fully fuzzy linear programming problems, *Applied Mathematical Modelling*, 35(2) (2011) 817–823.
2. H.Yang and J.Zhou, Optimal traffic counting locations for origin–destination matrix estimation, *Transportation Research Part B: Methodological*, 32(2) (1998)109–126.
3. H.Yang, C.Yang and L.Gan, Models and algorithms for the screen line-based traffic-counting location problems, *Computers & Operations Research*, 33(3) (2006) 836–858.
4. H.D.Sherali and R.Sivanandan, A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes, *Transportation Research Part B: Methodological*, 28(3) (1994) 213–233.
5. K.M.Abdushukoor and C.M.Sushama, A fuzzy graph approach for selecting optimal traffic counting locations in road networks, *International Journal of Pure and Applied Mathematics*, 109(9) (2016) 90–98.
6. M.M.Gupta, *Introduction to fuzzy arithmetic: Theory and applications*. New York, NY: Van Nostrand Reinhold Company, 1985.
7. Md. Yasin Ali and K.R.Chowdhury, Solution of fuzzy non-linear equations over triangularfuzzy number using modified secant algorithm, *Annals of Pure and Applied Mathematics*, 12(1) (2016) 41-47.
8. O.T.Manjusha and M.S.Sunitha, Total domination in fuzzy graphs using strong arcs, *Annals of Pure and Applied Mathematics*, 9(1) (2015) 23-33.
9. S.Jana and B.Das, Profit maximization solid transportation problem with Gaussian type-2 fuzzy environments, *Annals of Pure and Applied Mathematics*, 12(1) (2016) 41-47.
10. T.-S.Liou and M.-J.Wang, Ranking fuzzy numbers with integral value, *Fuzzy Sets and Systems*, 50(3) (1992) 247–255.
11. T.Larsson, J.T.Lundgren and A.Peterson, Allocation of link flow detectors for origin-destination matrix estimation - A comparative study, *Computer-Aided Civil and Infrastructure Engineering*, 25(2) (2010) 116–131.
12. W.Xu and Y.Chan, Estimating an origin-destination matrix with fuzzy weights: Part i: Methodology, *Transportation Planning and Technology*, 17(2) (1993) 127–144.