A PRACTICAL O-D MATRIX ESTIMATION MODEL BASED ON FUZZY SET THEORY FOR LARGE CITIES

Yousef Shafahi and Reza Faturechi
Department of Civil Engineering
Sharif University of Technology
Azadi Avenue, Tehran, Iran
E-mail: shafahi@sharif.edu

KEYWORDS
O-D matrix estimation, Traffic counts, Fuzzy theory.

ABSTRACT

Estimation of the origin-destination trip demand matrix (O-D) plays a key role in travel analysis and transportation planning and operations. Many researchers have developed different O-D matrix estimation methods using traffic counts, which allow simple data collection as opposed to the costly traditional direct estimation methods based on home and roadside interviews.

In this paper, a new fuzzy O-D matrix estimation model (FOMSEM) is proposed to estimate the O-D matrix from traffic count. A gradient-based algorithm, containing a fuzzy rule based approach to control the estimated O-D matrix changes, is proposed to solve FOMSEM.

Since link data only represents a snapshot situation, resulting in inconsistency of data and poor quality of the estimated O-Ds, the proposed method considers link data as fuzzy values that vary within a certain bandwidth. An equilibrium based fuzzy assignment method is proposed to assign the estimated O-D matrix, which causes the assigned volumes to be fuzzy numbers. The shortest path algorithm of the proposed method is similar to the Floyd-Warshall algorithm, and we call it the Fuzzy Floyd-Warshall Algorithm (FFWA). We introduce a new fuzzy comparing index to compare and estimate the distance between the assigned and observed link volumes and the model is formulated based on this index. FOMSEM is implemented in Mashhad city in Iran. Real data obtained from Mashhad Comprehension Transportation Study (MCTS) are used in this study and results are presented to show high capability of FOMSEM to estimate O-D matrix in large networks.

INTRODUCTION

Obtaining the origin-destination (O-D) matrix by conventional methods takes a considerable amount of time, money, and manpower, while gathering traffic volume data for the links of the transportation network is easy. Recently, a variety of analytic models has been developed to establish O-D trip matrices based on traffic counts along with other information.

Although O-D matrix estimation models from traffic counts are different in formulation, they are similar in that their implementation is extremely difficult in large-scale networks. Another important problem is the inconsistency among link-count data that results in poor quality of the estimated O-D's and even non-convergence of the solution algorithm. Data inconsistency occurs due to diverse reasons, such as counting error and the derivier's behaviour. Some researchers tried to solve this problem by using a set of "Fuzzy Weights" for each piece of inconsistent data (Xu and Chan 1993). 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.

The traffic assignment methods utilized in the above O-D estimation models, such as equilibrium and proportional based, consider the user's route choice as a crisp and precise process. Some research has demonstrated that the common equilibrium based assignment methods cannot correctly reproduce real traffic volumes. It is indicated that real assigned traffic flow might not certainly converge to the user-equilibrium. In other words, travel times used in routes have salient differences from the lowest route travel time (Jan and Ridwan 1994). As a result, it should be realized that common equilibrium and stochastic assignment models do a poor job of accounting for imprecision and uncertainty in the user's route choice behavior. Recently, some assignment models based on fuzzy theory have been proposed, but these models have not been widely used in the O-D estimation problem. Some researchers proposed a fuzzy inference based assignment algorithm in the O-D matrix estimation problem (Harikishan and Partha 1998). They applied an entropy model previously developed in the upper level of the O-D estimation problem.

In this paper, we use a new fuzzy approach to propose equilibrium based O-D matrix estimation model. We implement this model in Mashhad City, one large city in Iran. This model has a bi-level structure, and the fuzzy approach is applied in both the upper and lower levels.
THE PROPOSED FUZZY O-D MATRIX ESTIMATION MODEL (FODMEM)

Here, we propose the fuzzy O-D matrix estimation model (FODMEM). The model is formulated as an optimization problem. The objective function tries to minimize the fuzzy distance between the assigned and observed link volumes based on a new fuzzy index $I_i^t$:

$$
\min Z(q) = \frac{1}{2} \sum_{a \times i} \left( 1 - I_i^t(\tilde{x}_a) \right)^2 \\
\text{s.t.} \quad \tilde{x} = \text{assign}(q) 
$$

(1)

(2)

$q$ : Trip demand matrix

$\text{assign}(q)$: Fuzzy traffic assignment

$\tilde{x}$ : The observed links set

The observed link volume is defined as a fuzzy value:

$$
\tilde{x} = (\tilde{x}^L, \tilde{x}, \tilde{x}^R)
$$

(3)

where $\tilde{x}$ is the crisp value of observed link volume, and $\tilde{x}^L = \tilde{x}(1-\beta_i)$ and $\tilde{x}^R = \tilde{x}(1+\beta_i)$ are the lower and upper boundaries. $\beta_i$ and $\beta_i$ are experimental values and represent the degree of imprecision in link counts, which can be obtained based on expert knowledge or observation of daily traffic variation of each link in a specific time period.

The assigned fuzzy link volumes $\tilde{x}$ is formulated as below:

$$
\tilde{x} = (x^L, x, x^R)
$$

(4)

where $\tilde{x}$ is the fuzzy link volume, $x$ is the assigned link volume, and $x^L = x(1-\alpha_i)$ and $x^R = x(1+\alpha_i)$ are the lower and upper boundaries. The parameter $\alpha_i$ is the link degree of saturation, and $\alpha_i$ is an experimental value. These parameters can be specified through different approaches, such as interviewing, using expert knowledge or simulation. $I_i^t(\tilde{x}_a)$ is the proposed fuzzy comparison index between the fuzzy observed link volume $\tilde{x}_a$ and the fuzzy assigned link volume $\tilde{x}_a$ in link $a$. This index is computed as below:

$$
I_i^t(\tilde{x}_a) = \frac{x_a^R - \tilde{x}_a^L}{(x_a^R - x_a^L) + (x_a^R - x_a^L)}, \quad -\infty \leq I_i^t \leq \infty 
$$

(5)

Where:

$I_i^t(\tilde{x}_a) = I_i^t(\tilde{x}_a) = 1 \iff \tilde{x}_a = \tilde{x}_a$

Figure 1 illustrates the schematic diagram of this index for two fuzzy number "M" and "N", which is defined as the height of cross point of $[M, \infty)$ and $(-\infty, N]$:

$$
\mu_{[M, \infty]}(x) = \frac{n - x}{n^R - n}, \forall x 
$$

(6)

$$
\mu_{(-\infty, N]}(x) = \frac{x - m^L}{m^R - m^L}, \forall x 
$$

(7)

Figures 1: The fuzzy comparison index $I_i^t(M)$

The lower level of problem $\text{assign}(q)$ includes a new fuzzy traffic assignment model (Shafahi and Ramezani 2007), resulting in the assigned fuzzy link volumes $\tilde{x}$. Shafahi and Ramezani assumed that the user's decision-making process in route choice is based on their perceived travel time (PTT), which is presented as a triangle fuzzy number:

$$
\tilde{t} = (t^L, t^C, t^R)
$$

(8)

where $\tilde{t}$ is the fuzzy link travel time, $t^C$ is the assigned link travel time, and $t^L$ and $t^R$ are the lower and upper boundaries. The route choice decision-making process in this model is based on the fuzzy ranking method (Dubois and Prade 1983). Shafahi and Ramezani chose the index $I_3$ as the better ranking criterion in the selection of the best path and assigning traffic (Shafahi and Ramezani 2007). The computation formula of the index $I_3$ for comparing two fuzzy link travel times $\tilde{t}_1$ and $\tilde{t}_2$ is equal to:

$$
I_3(\tilde{t}_1) = \frac{t_2^R - t_1^C}{t_1^R + t_2^R + (t_1^L + t_2^L)}
$$

(9)

When the index $I_3$ values for two fuzzy numbers are equal, these numbers have no differences. Shafahi and Ramezani demonstrated that two fuzzy numbers $\tilde{t}_1$ and $\tilde{t}_2$ are equal if:
\[ I_3(\tilde{t}_i) = I_3(\tilde{t}_j) = 0.5 \iff \tilde{t}_i = \tilde{t}_j \] (10)

We improve Shafahi and Ramezani's fuzzy Shortest Path Algorithm (FSPA), which is similar to the Dijkstra algorithm and propose a more efficient and faster approach based on the Floyd-Warshall shortest path algorithm (FWA). This algorithm simultaneously finds the shortest path between all nodes. In the approach, similar to the FSPA algorithm, we use the index \( I_3 \) described above to compare link PTTs and to choose the shortest path, and call it "Fuzzy Floyd-Warshall Algorithm (FFWA)". To run FFWA:

**S**: The set of all nodes,
\[ \tilde{d}(i, j) \]: The fuzzy distance (fuzzy travel time) of the shortest path between two nodes \( i, j \in S \),
\[ \tilde{t}(i, j) \]: The fuzzy length of the link between two nodes \( i \) and \( j \). The link length between \( i \) and \( i \) is defined to be 0, and the length of two nodes without any direct links between them is assumed to be \( \infty \),
\[ \text{min} \]: The function computes the minimum of the fuzzy lengths according to index \( I_3 \).

FFWA is as follows:

**Step 1- Initialization**
Set: \( \tilde{d}(i, j) = \tilde{t}(i, j) \), for all \( i, j \).

**Step 2- Induction**
For all \( k \in S \)
For all \( i, j \)
Set: \( \tilde{d}(i, j) = \min[\tilde{d}(i, j), \tilde{d}(i, k) + \tilde{d}(k, j)] \)

Fuzzy Equilibrium Principle: If "M" and "N" are assumed to be two used paths between a distinct origin and destination and "K" is an unused path between that origin and destination, then the following equations hold (Shafahi and Ramezani 2007):
\[ I_3(M) = I_3(N) = 0.5 \]
\[ I_3(K) \leq 0.5 \] (11)

(12)

The fuzzy assignment algorithm is an incremental assignment, so that it can be called Fuzzy Incremental Traffic Assignment Algorithm (FITAA). Assuming that the travel time function of the link "a" is \( t_a(x_a) \), the trip demand for origin-destination \( rs \) is \( q_{rs} \), and \( N \) is the repetition number of the incremental algorithm, then the assignment algorithm is:

**Step 0 (Initialization) - Given** \( N \), \( \alpha_i \), \( \alpha_j \), and \( \delta \) (\( \delta \) =a small value, say 0.001), set:
\[ n = 1, t_a = t(0), t_a^{N} = t(0) + \delta, \ t_a^{N} = t(0) - \delta \]

**Step 1-** For all O-D pairs \( rs \), find the shortest paths and assign \( q_{rs} \) to the shortest paths.

Step 2- Update link travel times according to the volumes assigned in the previous step:
\[ t_a = t_a(x_a), \ t_a^{N} = t_a([1 + \alpha_r x_a], \ t_a^{N} = t_a([1 - \alpha_r x_a]) \]
\( x_a \) = Assigned volume to link "a"

**Step 3 (stop criterion) - if** \( n = N \) **stop, otherwise set** \( n = n + 1 \) and go to step 1.

Ramezani proved the convexity of the fuzzy assignment model (Ramezani 2007), so equation (1) includes a convex problem that ensures the uniqueness of the problem solution in the lower level. However, there is an infinite number of possible matrices that might result in the observed volumes, so similar to conventional O-D estimation models, we consider an original O-D matrix \( \hat{q} \) from past demand studies to guide the output solutions and expect the resulting matrix to closely resemble the original matrix. We use the gradient method as an efficient way to minimize deviation from the starting solution. Thus, we can obtain the most reliable O-D matrix among all possible matrices, which is near the original demand matrix and reproduces the observed link volumes (Spiesz 1990). Here, we improve upon this method for our fuzzy approach.

Based on the gradient method, the relative change in demand \( q \) can be written as:
\[ q_{rs}^{l+1} = \left[ q_{rs}^{l} - \lambda^{l} \left( \frac{\partial z(q)}{\partial q_{rs}} \right) \right]_{l=1,2,3...} \] (13)

where \( q_{rs}^{l} \) and \( q_{rs}^{l+1} \) denote the original matrix and the adjusted demand matrix in step \( l \) for origin-destination pairs \( rs \). This approach implies that the O-D pairs without trip demand would not be affected by the adjustment and zeros will be preserved. The parameter \( \lambda^{l} \) is the step length and has to be chosen small enough to ensure that the path followed by \( q_{rs}^{l} \) is sufficiently close to the true gradient path. The parameters \( \frac{\partial z(q)}{\partial q_{rs}} \) and \( \lambda^{l} \) have to be computed in the adjustment process. The relationship between link volumes and path flows in the fuzzy equilibrium assignment is expressed as below:
\[ \hat{\chi}_a = \sum_{rs} \sum_{k} \hat{t}_{rs} \delta^{rs}_{a,k}, a \in A \] (14)

\( k \in K_r \) is the set of paths used for each origin-destination \( rs \). An expression for the gradient \( \frac{\partial z(q)}{\partial q_{rs}} \) is easily determined if the path probabilities are used instead of path flows, as shown:
Because $x_a$ and $x'_a$ are independent of $\lambda$, the optimal step length can be found by computing the zero of the derivative. The derivative is again obtained by applying the chain rule:

\[
\frac{\partial z(\lambda)}{\partial \lambda} = \frac{\partial z(\lambda)}{\partial \tilde{x}_a} \times \frac{\partial \tilde{x}_a}{\partial \lambda} = 0
\]

assuming $\alpha_a = \beta_a$.

\[
\lambda^* = \sum_{a=1}^{\lambda} \left( \frac{x'_a (\lambda)}{\lambda^2} \right) \leq 1
\]

subject to:

\[
\hat{q}_{rs} > 0 \text{ with } rs \in RS, \text{ for all } \lambda \frac{\partial z(\lambda)}{\partial q_{rs}} \leq 1
\]

By using equations (17) and (24), we can solve for all unknowns in (13).

**An Expert Based Fuzzy Control Rule**

Here, we propose an initiative fuzzy rule based approach to control the O-D matrix changes and to enter expert knowledge in the estimation process. First, the original O-D matrix elements are clustered into tree clusters: $c_1$, $c_2$, and $c_3$ as "Small", "Medium", and "Large" O-D demand values, using C-means fuzzy clustering method (Zimmermann 1996), and then we define the maximum allowable change of demand values assigned to each cluster using a fuzzy rule base. This method results in controlling and limiting the estimated O-D matrix, preventing unreasonable changes during algorithm iterations.

By testing different membership functions for elements of clusters, we choose the "Gaussian Function" as the best fitting one. In the next step, we define three experimental fuzzy sets: $ac_1$ (Large), $ac_2$ (Medium), and $ac_3$ (Small) as maximum allowable changes to O-D demand values, which are shown in figure 2. There are no restrictions considered to the minimum allowable changes. Therefore, the membership functions presented in figure 2 are in a non-symmetric formation.
The fuzzy membership functions are:

\[ \mu_{cs}(x) = \begin{cases} 3 - x & , 2 \leq x \leq 3 \\ 0 & , otherwise \end{cases} \]  

(26)

\[ \mu_{cm}(x) = \begin{cases} 4 - 2x & , 1.5 \leq x \leq 2 \\ 0 & , otherwise \end{cases} \]  

(27)

\[ \mu_{cl}(x) = \begin{cases} 6 - 4x & , 1.25 \leq x \leq 1.5 \\ 0 & , otherwise \end{cases} \]  

(28)

Finally, we define the fuzzy rule base including three fuzzy rules as follows:

If \( \hat{q}_{rs} \) is "Small", then \( \frac{q_{rs}}{q_{rs \_{max}}} \) is "Large"  

(29)

If \( \hat{q}_{rs} \) is "Med", then \( \frac{q_{rs}}{q_{rs \_{max}}} \) is "Med"  

(30)

If \( \hat{q}_{rs} \) is "Large", then \( \frac{q_{rs}}{q_{rs \_{max}}} \) is "Small"  

(31)

We assume that as the demand increases (in developed zones), the relevant allowable change decreases and as the demand decreases (in developing zones), the relevant allowable change increases.

In the final step, the results are combined and defuzzified and the maximum allowable change of O-D demand values \( \sigma_{rs} \) is computed as a crisp number. We use the center of gravity (COG) method as the most popular defuzzification method (Zimmermann 1996), which is computed by the following equation:

\[ \left( \frac{q_{rs}}{q_{rs \_{max}}} \right)^{\text{COG}} = \sigma_{rs} = \frac{\sum_{i=1}^{g} b_i \int_{x_i}^{x_{i+1}} \mu_{ac_i}(x)dx}{\sum_{i=1}^{g} \int_{x_i}^{x_{i+1}} \mu_{ac_i}(x)dx} \]  

(32)

where \( R \) is the number of rules (here \( R = 3 \)), \( b_i \) is the center of area of the membership function of \( ac_i \) for the \( i^{\text{th}} \) rule, and \( \int_{x_i}^{x_{i+1}} \mu_{ac_i}(x)dx \) denotes the area under \( \mu_{ac_i}(x) \). We add the constraint below to the O-D estimation model:

\[ \hat{q}_{rs} > 0 \quad \text{with} \quad rs \in RS \quad \text{for all} \quad \frac{q_{rs}}{q_{rs \_{max}}} \leq \sigma_{rs} \]  

(33)

**Solution Algorithm**

Finally, the solution algorithm can be proposed as follows:

Step 0 (Initialization) - Given \( \beta_i \) and \( \beta_j \) (the degree of link count imprecision), set:

\[ l = 0 \quad q_{rs} = q_{rs}^l = \hat{q}_{rs} \]

Step 1 - Assign the demand matrix \( q_{rs}^l \) by the proposed fuzzy assignment method to find the center of the assigned fuzzy volume \( x_{rs} \).

Step 2 - Find used paths \( K_{rs} \) and the center of fuzzy path flows for every \( rs \in RS \). Compute the value of parameters \( \frac{\partial z(q)}{\partial q_{rs}} \) and \( \lambda \) as below:

\[ \frac{\partial z(q)}{\partial q_{rs}} = \sum_{k=K_{rs}} \sum_{a,b} \delta_{a,b} \frac{\tilde{x}_s(\beta_s + \alpha_s)(x_s - \hat{x}_s)}{\beta_s \hat{x}_s + \alpha_s x_s} \]

\[ \lambda' = \frac{\sum_{a,b} \left( \frac{x_s}{\hat{x}_s} \right)^2}{\sum_{a,b} \left( \frac{x_s}{\hat{x}_s} \right)^2} \]

Step 3 - If constraint (25) or the demand maximum allowable change condition (33) is not satisfied, the revised value of step length is:

\[ \lambda' = \frac{\lambda_1 + \lambda_2}{2} \]

for which:

\[ \kappa_i = \min \left\{ \frac{1 - \sigma_{rs}}{\frac{\partial z(q)}{\partial q_{rs}}} \right\} \quad q_{rs} > 0, \frac{\partial z(q)}{\partial q_{rs}} > 0 \]

\[ \kappa_i = \max \left\{ \frac{1}{\frac{\partial z(q)}{\partial q_{rs}}} \right\} q_{rs} > 0, \frac{\partial z(q)}{\partial q_{rs}} < 0 \]

Step 4 - Determine the new values of \( q_{rs} \) using the equation below:
$$q_{n+1}^l = q_{n}^l \left[ 1 - \lambda \left( \frac{\partial z(q)}{\partial q_n^l} \right) \right]$$

Step 5 - Let $l = l + 1$ and $q_n^l = q_n^l$

Step 6 (stop criterion) - If $\left| -I^l_i \left( \frac{\partial z}{\partial x_n} \right) \right| < \varepsilon$, stop, else go to step 1.

**APPLICATION OF FODMEM IN MASHHAD CITY**

The FODMEM has been successfully tested by using various examples with different characteristics. An user-friendly software “FODMES” (Fuzzy O-D Matrix Estimation Software) has been produced to use the method for different examples easily.

In order to numerically illustrate the effectiveness of FODMEM in a real world networks, we implemented FODMEM for Mashhad city network, a large city in Iran. Real O-D matrix data obtained from Mashhad Comprehension Transportation Study (SCTS) in 1996 are used as original matrix in this. Mashhad city included 2.0 million people. The city is divided into 141 traffic zones and the traffic network includes 935 nodes, 2538 links and 7157 O-D pairs with non zero demand. An Origin-Destination survey with house interview was conducted in 1996. Data were gathered from 4% of the households and were validated by observation from different screen lines in the study area. In addition, traffic volumes of 119 links of Mashhad network are surveyed and are used as observed link volumes. Figure 3 illustrates Mashhad City network and observed links.

![Observed Links](image)

**Figure 3: Mashhad City network and observed links**

The general links travel time follows the BPR function

$$t(x) = t_o \left[ 1 + 0.15 \left( \frac{x}{C} \right)^4 \right]$$

where $x$, $t_o$ and $C$ denote the traffic volume, free flow travel time and practical capacity of links. All links in Mashhad network are divided into 8 classes and average values of BPR function parameters $t_o$ and $C$ are estimated by volume-travel time observation from several links in each class as demonstrated in table 1.

<table>
<thead>
<tr>
<th>Link class</th>
<th>Values of BPR Function Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t_o$ (min)</td>
</tr>
<tr>
<td>Local</td>
<td>1.50</td>
</tr>
<tr>
<td>Collector</td>
<td>1.33</td>
</tr>
<tr>
<td>Minor arterial- two way (CBD area)</td>
<td>1.20</td>
</tr>
<tr>
<td>Minor arterial- two way (Non CBD)</td>
<td>1.20</td>
</tr>
<tr>
<td>Minor arterial-one way</td>
<td>1.10</td>
</tr>
<tr>
<td>Principal arterial (Class II)</td>
<td>1.00</td>
</tr>
<tr>
<td>Principal arterial (Class I)</td>
<td>1.00</td>
</tr>
<tr>
<td>Freeway</td>
<td>0.80</td>
</tr>
</tbody>
</table>

We consider 25 iterations as the stop criterion of FODMEM and 15 iterations for FITA. The parameters $\alpha_r$, $\alpha_l$, $\beta_r$, and $\beta_l$ are assumed to be 0.2. Figure 4 shows the dispersal diagram of the observed and estimated link volumes and figure 5 shows the dispersal diagram of the estimated and original O-D matrix results of the estimated link volumes in the observed links. The $R^2$-squared values between the observed and estimated values (0.99 for link volume and 0.89 for O-D values) indicate that the model successfully reproduces the observed link volumes and ensures the high vicinity of the estimated O-D matrix to the original case.

![Dispersal Diagram](image)

**Figure 4: Dispersal diagram of the observed and estimated link volume**
SUMMARY AND CONCLUSION

This paper attempts to develop a fuzzy O-D estimation model FODMEM using a gradient-based algorithm to utilize more of the estimated demand matrix. The observed and assigned link volumes are defined as fuzzy numbers. This definition applies fuzzy sets to the uncertainty embedded in the driver’s route choice behaviour and solves the inconsistency in link counts. A new fuzzy comparing index is proposed to evaluate link counts and assigned volumes, and to measure the fuzzy distance between them and the O-D estimation model is constructed based on this index. A gradient-based iterative algorithm is applied to solve the model. In the lower level of this model, we use an equilibrium based fuzzy assignment model (Shafahi and Ramezani 2007). We improve this model by substituting a new fuzzy shortest path algorithm FFPA based on the Floyd-Warshall algorithm, instead of using the FSPA. In order to limit unreasonable changes to the original O-D matrix, the elements of this matrix are clustered into three groups using the fuzzy clustering method and a fuzzy rule is applied to every O-D group. The proposed fuzzy rule based approach facilitates the use of expert knowledge in the model and finds a more reliable solution. Finally, Mashhad city is chosen as our case study. The results show that FODMEM can successfully find the O-D demand matrix, while reproducing the observed link volumes and the original and estimated O-D matrix closely resemble each other.

REFERENCES


AUTHOR BIOGRAPHIES

YOUSEF SHAFAHI was born in Kazeroon, Iran. He got his BS from Shiraz university, Shiraz, Iran; MS from Isfahan University of Technology, Isfahan, Iran and Ph.D. from University of Maryland at College Park, MD, USA. He is now a faculty of Civil Engineering Department of Sharif University of Technology, Tehran, Iran. His e-mail address is shafahi@sharif.edu.

REZA FATURECHI was born in Oromiyeh, Iran and went to Sharif University of Technology, where he studied Civil Engineering and obtained his B.S. degree in 2006. He obtained his M.Sc. degree in Transportation Engineering in 2008. He is now leading a research group in the field of Traffic Engineering and Transportation Planning at ARG Company. His e-mail address is reza.faturechi@yahoo.com.