A FUZZY LOGIC INFERENCE APPROACH FOR THE ESTIMATION OF THE PASSENGERS FLOW DEMAND

Aránzazu Berbey Alvarez, Rony Caballero George
Universidad Tecnológica de Panamá
aranzazu.berbey@utp.ac.pa, rony.caballero@utp.ac.pa

Juan de Dios Sanz Bobi, Ramón Galán López
Universidad Politécnica de Madrid
juandedios.sanz@upm.es, ramon.galan@upm.es,

Keywords: Fuzzy logic inference, passenger flow modelling, maximum entropy theory, matrix method of area trips O-D.

Abstract: This paper presents a new approach that designs the flow of passengers in mass transportation systems in presence of uncertainties. One of the techniques used for the prediction of passenger demand is the origin-destination matrices. However, this method is limited to urban areas and rarely to explicit stations. Otherwise, the gravity models based on friction functions can be another alternative; however, it is difficult to fit into practical achievements. Another solution might be the application of artificial intelligence techniques so as to include some intuitive knowledge provided by an expert to predict the flow demand of passengers' trips in explicit stations. This paper proposes to combine a matrix of origin-destination trips of travel zones, with the intuitive knowledge, applying a fuzzy logic inference approach.

1 INTRODUCTION

The passenger flow modelling has been tackled by several authors (Kiluchi et al, 1999) (Aldian et al, 2003) (Watson and Prevedouros, 2006) (Cheng et al, 2009) (Murat, 2010) (Xie et al, 2010). Some of them have applied linear programming methods considering small samples (Murat, 2010). It has also been applied the maximum entropy theory (Xie et al, 2010) and results have been compared between different methods (Watson and Prevedouros, 2006). However, it has been shown that these methods have limitations in certain scenarios. In this case, the use of fuzzy logic has proven to be a promising tool because it can integrate the railway planning experts' experience in multiple scenarios (Aldian et al, 2003) (Cheng et al, 2009). This paper proposes to integrate the effectiveness of the methods based on origin-destination matrices, with the experience of experts in railway planning using an artificial intelligence support based on fuzzy logic.

2 AREA ANALYSIS

The matrix method of area trips O-D is one of the most used methods to design the movement of passengers. This method divides the urban environment in areas of interests that generate and attract trips. However, due to economic and practical reasons, the size of the interests' zones is usually large in order to apply directly to urban planning models. Therefore, it is interesting in many cases to divide these macro areas into smaller areas, allowing a better analysis of passenger flow.

3 SUBZONE DESIGN

Considering the attraction vectors $D_T \in \mathbb{R}^{1\times n_m}$ and the generation vectors $O_T \in \mathbb{R}^{n_m\times 1}$, each element of the attraction vectors D_{TJ} and generation O_{TI} can be subdivided into n_J subzones, becoming two new vectors $D \in \mathbb{R}^{1\times n_s}$ and $O \in \mathbb{R}^{n_s \times 1}$, where $n_s = \sum_{J=1}^{n_m} n_J$.

Therefore, the macro areas matrix $T \in \mathbb{R}^{n_m \times n_m}$ is subdivided at the same time, obtaining sub-zone matrix $M \in \mathbb{R}^{n_s \times n_s}$. Consequently, it can be shown that every element of the attraction and the generation vectors, as well as the macro area and the subzone matrices must meet with:

$$D_{TJ} = \sum_{j=1+m_B(J)}^{m_E(J)} D_j \tag{1}$$

$$O_{TI} = \sum_{i=1+m_B(I)}^{m_E(I)} O_i \tag{2}$$

$$T_{IJ} = \sum_{i=1+m_B(I)}^{m_E(I)} \sum_{j=1+m_B(J)}^{m_E(J)} M_{ij} \tag{3}$$

$$A_j = \sum_{i=1}^{n_s} M_{ij} \ \ \forall j = 1 \cdots n_s \tag{4}$$

$$B_i = \sum_{j=1}^{n_s} M_{ij} \ \ \forall i = 1 \cdots n_s \tag{5}$$

where

$$m_B(x) = \sum_{K=1}^{x} n_{K-1} \tag{6}$$

$$m_E(x) = \sum_{K=1}^{x} n_K \tag{7}$$

$$n_0 = 0 \tag{8}$$

To obtain a subzone model, a system is governed by the principle of maximum entropy would be applied. However, this model can be improved by incorporating information by an expert, even if this is inaccurate; this supposition is done with the aim to improve the approximation, and it means to apply the principle of maximum entropy. The information can be included using techniques based on artificial intelligence and more specifically based on fuzzy logic.

We have to apply the following steps to set up the subzone model:

a) $\boldsymbol{T_{IJ}}$, $\boldsymbol{O_{TI}}$ and $\boldsymbol{D_{TJ}}$ have little uncertainty within the same planning horizon.

b) Every element of the vector $\boldsymbol{O_i}$ and $\boldsymbol{D_j}$, corresponding to a subzone $\boldsymbol{i}$ or $\boldsymbol{j}$, that is part of the macro area $\boldsymbol{I}$ or $\boldsymbol{J}$ respectively, can be represented as a function of O_{TI} or $\boldsymbol{D_{TJ}}$ vectors of the macro area in which it is contained, and by a potential function with exponents $\boldsymbol{C_{DJ}}$ or $\boldsymbol{C_{Oi}}$.

$$D_j = \frac{2^{\frac{1}{\ln 2} C_{Dj}}}{n_J \sum_{k=1+m_B(J)}^{m_E(J)} 2^{\frac{1}{\ln 2} C_{Dk}}} D_{TJ} \tag{9}$$

$$O_i = \frac{2^{\frac{1}{\ln 2} C_{Oi}}}{n_I \sum_{k=1+m_B(J)}^{m_E(J)} 2^{\frac{1}{\ln 2} C_{Ok}}} O_{TI} \tag{10}$$

c) Every element of the trip matrix $\boldsymbol{M_{ij}}$ corresponding to the subzones $\boldsymbol{i}$, $\boldsymbol{j}$ and that is part of the macro area $\boldsymbol{I}$, $\boldsymbol{J}$ respectively, can be represented as a function of $\boldsymbol{T_{IJ}}$ and by a potential function with exponent $\boldsymbol{C_{Mij}}$.

$$M_{ij} = \frac{2^{\frac{1}{\ln 2} C_{Mij}}}{n_I n_J \sum_{r=1+m_B(I)}^{m_E(I)} \sum_{s=1+m_B(J)}^{m_E(J)} 2^{\frac{1}{\ln 2} C_{Mrs}}} T_{IJ} \tag{11}$$

d) $\boldsymbol{C_{Dj}}$, $\boldsymbol{C_{Oi}}$ and $\boldsymbol{C_{Mij}}$ come from a fuzzy inference engine based on the experience of an expert.

It is important to note that the exponents $\boldsymbol{C_{Dj}}$ and $\boldsymbol{C_{Oi}}$ are related to the relative level of importance of the station in a given planning horizon. If these exponents are zero, the estimation becomes the estimation by maximum entropy, whereas if you have a negative or positive number it corresponds to a station with low or high demand respectively.

On the other hand the exponent $\boldsymbol{C_{Mij}}$ establishes the relative level of importance of the passenger flow between two stations. Here, the estimation using the maximum entropy also corresponds for a value equal to zero, and a negative or positive number corresponds to a flow of low or high demand respectively.

4 FUZZY INFERENCE ENGINE

The exponents C_{Dj}, C_{Oi} and C_{Mij} can be designed using the help of fuzzy logic. In this particular case it seems reasonable to use triangular membership functions (Figure 1). These exponents can be estimated not only considering the proximity of the station by major population centres, such as: residential areas, industrial estates, hospitals, schools and shopping centres, but also with the presence of transport interchanges.

VL	Very Low
L	Low
M	Medium
H	High
VH	Very High

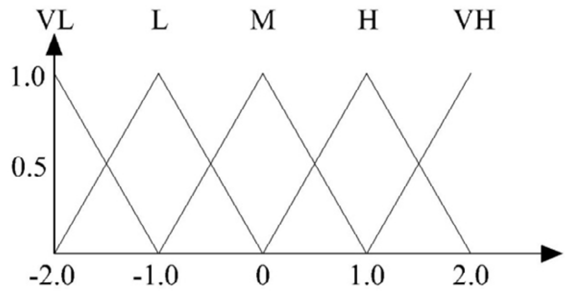

Figure 1: Starting fully logic functions for C_{Dj}, C_{Oi} y C_{Mij}.

5 PROPOSED ARGORITHM

a) Estimate the exponents C_{Dj}, C_{Oi} and C_{Mij} basing on the inaccurate available information.

b) Estimate the trip attraction and generation vectors $\boldsymbol{D}$ and $\boldsymbol{O}$, for the stations using,

$$D_j = \frac{2^{\frac{1}{\ln 2} C_{Dj}}}{n_J \sum_{k=1+m_B(J)}^{m_E(J)} 2^{\frac{1}{\ln 2} C_{Dk}}} D_{TJ} \qquad (12)$$

$$O_i = \frac{2^{\frac{1}{\ln 2} C_{Oi}}}{n_I \sum_{k=1+m_B(J)}^{m_E(J)} 2^{\frac{1}{\ln 2} C_{Ok}}} O_{TI} \qquad (13)$$

c) Estimate the trip matrix $\boldsymbol{M_{ij}}$, using,

$$M_{ij} = \frac{2^{\frac{1}{\ln 2} C_{Mij}}}{n_I n_J \sum_{r=1+m_B(I)}^{m_E(I)} \sum_{s=1+m_B(J)}^{m_E(J)} 2^{\frac{1}{\ln 2} C_{Mrs}}} T_{IJ} \qquad (14)$$

d) Thus it is checked if $O_i = \sum_{j=1}^{n_s} M_{ij} \ \forall i = 1 \cdots n_s$.

Consequently, if $\left| \sum_{j=1}^{n_s} M_{ij} - O_i \right| < \varepsilon$ is met for a given value, it is accepted. Meanwhile, if not, the values of the matrix elements are corrected with

$$M_{ij} \Leftarrow \frac{O_i M_{ij}}{\sum_{j=1}^{n_s} M_{ij}} \qquad (15)$$

e) It is checked if $D_j = \sum_{i=1}^{n_s} M_{ij} \ \forall j = 1 \cdots n_s$.

Consequently, if $\left| \sum_{i=1}^{n_s} M_{ij} - D_j \right| < \varepsilon$ is met, it is accepted. Meanwhile, if not, the values of the matrix elements are corrected with

$$M_{ij} \Leftarrow \frac{D_j M_{ij}}{\sum_{i=1}^{n_s} M_{ij}} \qquad (16)$$

f) It is checked if $T_{IJ} = \sum_{i=1+m_B(I)}^{m_E(I)} \sum_{j=1+m_B(J)}^{m_E(J)} M_{ij}$ for all i, j, I, J. Consequently if, $\left| \sum_{i=1+m_B(I)}^{m_E(I)} \sum_{j=1+m_B(J)}^{m_E(J)} M_{ijm} - T_{IJ} \right| < \varepsilon_m$ is met, it is accepted. Meanwhile, if not, the values of the matrix elements are corrected with

$$M_{ij} \Leftarrow \frac{T_{IJ} M_{ij}}{\sum_{i=1+m_B(I)}^{m_E(I)} \sum_{j=1+m_B(J)}^{m_E(J)} M_{ij}} \qquad (17)$$

g) Go back to step d, if tolerances are not met.

6 RESULTS FOR SIMULATED DATA

The following case illustrates how the proposed algorithm works. Let's consider the railway network topology shown in Figure 2. In this case it is known from previous studies that the mobility preferences among three regions so-called A, B and C, but without giving any specific details concerning the connection preferences between specific stations. The region A has the stations E1, E2 and E3; the region B has the stations E4, E5, E6 and E7; and the region C has only two stations E8 and E9.

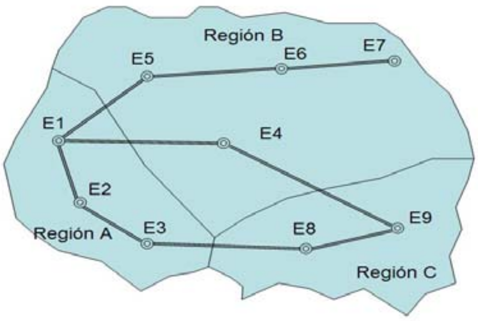

Figure 2: Red topology.

O-D matrix between these regions is known and it is described below:

Table 1: O-D matrix between the regions.

	A	B	C	
A	9397	5282	5213	19892
B	25118	8272	5065	38455
C	22570	8732	1134	32436
	57085	22286	11412	90783

We have to consider the estimated subzones exponents' vectors

$$C_D = \begin{bmatrix} A & A & M & A & M & M & M & MA & B \end{bmatrix},$$
$$C_o = \begin{bmatrix} A & M & M & A & M & M & A & MA & B \end{bmatrix}$$

and the exponents matrix,

$$C_D = \begin{bmatrix} & A & M & A & M & M & M & MA & MB \\ A & & B & M & M & B & M & B & MB \\ A & A & & M & B & B & B & M & MB \\ M & M & B & & M & M & M & M & B \\ M & M & B & M & & M & M & M & MB \\ M & M & B & A & M & & M & M & MB \\ A & A & M & MA & A & A & & A & B \\ A & A & M & A & A & M & A & & MA \\ B & B & MB & MB & MB & MB & MB & M & \end{bmatrix}$$

After, applying the proposed algorithm, we obtain the following matrix OD,

Table 2: Resulting O-D matrix using the proposed algorithm.

0	2945.4	1142.1	1667.2	433.1	575.7	486.1	4203.5	85.9
2051.3	0	381.6	557.1	289.4	192.4	324.9	351.2	28.7
1467.9	1408.6	0	398.7	103.6	137.7	116.2	502.6	41.1
2036.9	1954.6	757.9	0	295.3	392.6	331.5	872.9	142.7
1978.7	1898.8	736.3	552.2	0	381.3	322	847.9	69.3
1853.7	1778.8	689.8	1034.6	268.7	0	301.7	794.4	64.9
4903.1	4705	1824.4	2736.4	710.8	944.9	0	2101.1	171.8
8445.1	8104	3142.4	3292.4	1710.5	1136.9	1920	0	783.6
1234.6	1184.7	459.4	240.6	125	166.2	140.3	350.4	0

Which is compared to the original test matrix

Table 3: Original test matrix.

0	2585	1155	920	543	460	485	3247	211
1769	0	653	510	322	306	353	538	151
1704	1531	0	519	302	267	295	901	165
2282	2194	1012	0	465	401	440	844	231
1972	1935	832	687	0	441	496	828	179
2049	2064	946	778	519	0	598	789	212
3953	4127	1752	1554	1009	884	0	1557	425
7308	7405	3766	3018	1617	1398	1556	0	805
1447	1440	1204	343	263	264	263	329	0

We can see that the results for the dominant values of the matrix are approximated; obviously the errors are due to the uncertain information coming from the expert, but still much lower than using the approximation for maximum entropy.

Table 4: Resulting O-D matrix using the maximum entropy method.

0	1181.2	1228.4	526.2	353.4	438.1	351.4	2011.1	540.3
2153.9	0	1183.9	507.2	681.1	422.2	677.4	484.6	520.8
1774.2	1875.5	0	417.8	280.5	347.8	279	798.3	857.9
1915.8	2025.2	2106.1	0	593.9	736.3	590.7	522.6	1123.2
2069.3	2187.5	2274.9	477.7	0	795.3	638	564.4	606.6
2001.9	2116.2	2200.7	924.2	620.6	0	617.2	546	586.8
1970.7	2083.2	2166.4	909.8	610.9	757.4	0	537.5	577.7
3243.1	3428.3	3565.2	1129.6	1517.1	940.4	1508.8	0	885.8
3907.4	4130.5	4295.4	680.5	913.9	1133	908.9	248.2	0

Making a comparison of the percentage of the errors of both methods shows that the method of the proposed algorithm presents a certain amount of errors lower than the highest entropy method. Hereunder is the matrix that arises from the proposed algorithm.

Table 5: Percentage errors using the proposed algorithm.

	-13,94	1,12	-81,22	20,24	-25,15	-0,23	-29,46	59,29
-15,96		41,56	-9,24	10,12	37,12	7,96	34,72	80,99
13,86	7,99		23,18	65,70	48,43	60,61	44,22	75,09
10,74	10,91	25,11		36,49	2,09	24,66	-3,42	38,23
-0,34	1,87	11,50	19,62		13,54	35,08	-2,40	61,28
9,53	13,82	27,08	-32,98	48,23		49,55	-0,68	69,39
-24,03	-14,01	-4,13	-76,09	29,55	-6,89		-34,95	59,58
-15,56	-9,44	16,56	-9,09	-5,78	18,68	-23,39		2,66
14,68	17,73	61,84	29,85	52,47	37,05	46,65	-6,50	

Meanwhile the matrix that arises from the highest entropy method provides higher percentage errors in some cells, even 100 to 200 percent.

Table 6: Percentage errors using the maximum entropy method.

	54,31	-6,35	42,80	34,92	4,76	27,55	38,06	-156,07
-21,76		-81,30	0,55	-111,52	-37,97	-91,90	9,93	-244,90
-4,12	-22,50		19,50	7,12	-30,26	5,42	11,40	-419,94
16,05	7,69	-108,11		-27,72	-83,62	-34,25	38,08	-386,23
-4,93	-13,05	-173,43	30,47		-80,34	-28,63	31,84	-238,88
2,30	-2,53	-132,63	-18,79	-19,58		-3,21	30,80	-176,79
50,15	49,52	-23,65	41,45	39,45	14,32		65,48	-35,93
55,62	53,70	5,33	62,57	6,18	32,73	3,03		-10,04
-170,03	-186,84	-256,76	-98,40	-247,49	-329,17	-245,59	24,56	

7 CONCLUSIONS

In this paper the potential and effectiveness of this new methodology has been proven. Clearly, the obtained matrix has an error of 10% compared to the original matrix. However, these results are reasonable considering the level of uncertainty coming from the information provided by the expert.

As a guideline for future research, it is interesting to improve the fuzzy inference engine, since the used in this test was quite simple. Here, other sources of information can be incorporated without limiting the information just to one expert.

ACKNOWLEDGEMENTS

The authors of this paper want to express their gratitude to the National Secretary of Science and Technology (SENACYT) of the government of the Republic of Panama for funding this study through the R & D project “Methodologies and Performance Indicators of Railway Transport Systems” adjudged by a call for the promotion of R&D activities (MDEPRB09-001). Additionally, they want to thank the Research Centre on Railway Technologies (CITEF) of Universidad Politécnica de Madrid (Spain).

REFERENCES

Cheng, Y.H. y Yang, Li-An, 2009. *A Fuzzy Petri Nets approach for railway traffic control in case of abnormality: Evidence from Taiwan railway system*. Expert Systems with Applications 36 (2009) 8040–8048.

Lindaren, R. y Tantiyanugulchai, S., 2003. *Microscopic Simulation of Traffic at a Suburban Interchange*.

Watson, J. R. y Prevedouros, P. D., 2006. *Derivation of Origin–Destination Distributions from Traffic Counts Implications for Freeway Simulation*. Transportation Research Record: Journal of the Transportation Research Board,No. 1964, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 260–269.

S. Kikuchi y D. Miljkovic., 1999. *Method To Preprocess Observed Traffic Data for Consistency Application of Fuzzy Optimization Concept*. Transportation Research Record 1679 Paper No. 99-0129 73.

Aldian, A y Taylor, M., 2003. *Fuzzy multicriteria analysis for inter-city travel demand modelling*. Journal of the Eastern Asia Society for Transportation Studies, Vol. 5, October, 2003.

C. H. Murat., 2010. *Sample size needed for calibrating trip distribution and behaviour of the gravity model*. Journal of Transport Geography 18,183–190.

C. Xie, K.M. Kockelman y S.T. Waller., 2010. *A maximum entropy method for subnetwork origin-destination 1 trip matrix estimation*. The 89th Annual Meeting of the Transportation Research Board, January 2010 in Washington, DC, and for publication in Transportation Research Record. 2010.