Submitted for
The 3rd Conference in Innovations in Travel Modeling
May 9-12, 2010, Tempe, Arizona

How to Calibrate Transit O-D Tables from Information about Limited O-D Pairs

Shinya Kikuchi, Ph.D. PE, Professor
Nopadon Kronprasert, Graduate Student

Department of Civil and Environmental Engineering
Virginia Tech, National Capital Region
7054 Haycock Rd. Falls Church, VA 22043, USA
Telephone (703)538-8436
kikuchi@vt.edu and nopkron@vt.edu

October 2009

How to Calibrate Transit O-D Tables from Information about Limited O-D Pairs

Shinya Kikuchi and Nopadon Kronprasert
Virginia Tech, National Capital Region
7054 Haycock Rd. Falls Church, VA 22043, USA

Abstract

When constructing the O-D-table for an existing transit network, usually the analyst has two basic pieces of information to start with: one, the boarding and alighting data at each station that are collected as part of the FTA's annual transit operating database; and, two, the general knowledge about the travel pattern of the passengers for some stations pairs. This paper suggests how to utilize this knowledge in a most efficient manner when constructing the O-D table. The O-D table is estimated by maximizing the entropy while using the available knowledge as constraint. The study shows that the efficiency of deriving the O-D table depends on how to pick the available knowledge when incorporating it into the constraints of the optimization model. The O-D pairs which have a large deviation from the average value is most useful in converging the O-D values quickly. Two O-D tables from the real-world transit lines are used to support this point.

Problem

Kikuchi and Kronprasert (*1, 2*) proposed a model that constructs an O-D table using the boarding and alighting passenger counts and the analyst's knowledge about selected O-D pairs. The model is an optimization scheme in which the entropy is maximized under the constraint of the available spot knowledge about selected O-D pairs. They also showed that the speed by which the values of the O-D elements converge is sensitive to the values of the selected O-D pairs for which the spot information is available.

However, the questions remain as to (i) what type of O-D pair values should be sought as the constraints?; and (ii) how many pieces of such additional information are needed to create a sufficiently accurate O-D table? These questions relate to how to pick and prioritize the available knowledge about the O-D pair values in order to yield a reasonable O-D table consistent with the available information.

Model Structure

The Kikuchi-Kronprasert model applies two basic principles, uncertainty maximization and uncertainty minimization (*3*, *4*). The former suggests that the analyst make no commitment to the values for which no information is available. This principal is performed by maximizing the entropy. The latter principle suggests that the analyst use all the available information, including approximate values. These two principles are incorporated into a multi-objective fuzzy optimization model (*5*).

Model Formulation

The model is formulated as follows.

Objective: $Z = \text{Max } h$ (1)

Subject to three groups of constraints

Constraint 1: Entropy Maximization (Uncertainty maximization)

Entropy function, $H = -\sum_i \sum_j \left(\frac{t_{i,j}}{T} \log_2 \left(\frac{t_{i,j}}{T} \right) \right)$; for $i < j$ (2)

$S_{\text{large}}(H) \geq h$ (3)

Constraints 2: Use of the boarding and alighting data

Type 1: The boarding and alighting counts at each station, i.e., the row total and the column total. This value may be available as the exact values $b_i = \sum_{j=1}^{n} t_{i,j}$, and $a_j = \sum_{i=1}^{n} t_{i,j}$ or the approximate value $b_i \approx \sum_{j=1}^{n} t_{i,j}$, $a_j \approx \sum_{i=1}^{n} t_{i,j}$ with $S_{Bi}(b_i) \geq h; \forall i$ and $S_{Aj}(a_j) \geq h; \forall j$ (4)

Type 2: The passenger load (number of passengers on-board between any two stations.) This value may be the exact value $v_{k,k+1} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} t_{i,j}$, or the approximate value $v_{k,k+1} \approx \sum_{i=1}^{k} \sum_{j=k+1}^{n} t_{i,j}$ with $S_{Vk}(v_{k,k+1}) \geq h; \forall k$ (5)

Type 3: Spot knowledge for selected O-D pairs

This value may be the approximate values

$S_{Qk}(t_{i,j}) \geq h$ (6)

Relationships between selected O-D pairs

$S_{Qk}(t_{i,j}, t_{r,s}) \geq h$ (7)

Constraint 3: Basic parametric relationships among the elements of the O-D table

Total boarding = total alighting: $\sum_{i=1}^{n} b_i = \sum_{j=1}^{n} a_j = T$ (8)

Cumulative boarding to stop k ≥ Cumulative alighting to stop k: $\sum_{i=1}^{k} b_i \geq \sum_{j=1}^{k} a_j$ (9)

where h = degree of acceptability,
H = entropy,
T = total number of passenger trips,
$t_{i,j}$ = the number of passenger trips traveling from stops i to j, (decision variable)
b_i, a_i = the number of passengers boarding and alighting at stop i, respectively,
$v_{k,k+1}$ = the number of passengers travelling between stops k and k+1,
$S_{large}(H)$ = the membership function that represents “large” entropy,
$S_{Bi}(b_i)$ = the membership functions of approximate values of boarding counts at stop i,
$S_{Aj}(a_j)$ = the membership functions of approximate values of alighting counts at stop i,
$S_{Vk}(v_{k,k+1})$ = the membership functions of approximate values of passenger load data.

The model is solved as a multi-objective fuzzy optimization problem.

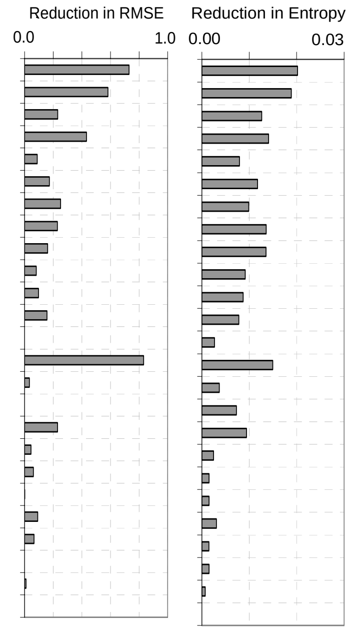
How to Select the O-D Pairs for Incorporating Spot Knowledge

The question in this paper is how to select the O-D pairs whose spot information is to be used for Constraint 2, Type 3 above. We believe that how quickly the value of root-mean-squared error (RMSE) and the entropy of the O-D table reduce is a reasonable measure for determining which O-D pairs should be picked for the constraints. We propose a two-step process: first, select the row (or column), and second, select a specific O-D pair among the O-D pairs in the row (or column). This procedure will be tested using the O-D data obtained from transit lines in Osaka and Yokohama, Japan.

<u>Select the row or the column.</u> Figure 1 shows the actual O-D table of Osaka’s LRT and also the reduction of the RMSEs and the entropy values as a result of adding the spot values for the O-D pairs in the row or in the column as Type 3, Constraint 2 above. For example, if the O-D pairs whose origin is stop 7 (i.e., 7-8, 7-9, 7-10, …, 7-26) are used for the constraints, the bar graphs on the right corresponding to stop 7 show the degree of reduction of the RMSE and the entropy. As seen, for stops 1, 2 and 14, the reduction of the RMSE and the entropy are large.

This is because the number of boarding passengers (or numbers of trips originating from) at Stops 1, 2, or 14 are very large (B_1 = 460, B_2 = 314, and B_{14} = 411). This suggests that the analyst should look for the row (or column) total which has a large number. Among the elements of the O-D table in Figure 1, those that are marked by a square have a large impact on calibrating the values of the O-D table. It is seen, in general, these marked values are found on the rows or columns whose total value, total boarding or total alighting, are large.

From Stop	To Stop																										Total Boarding
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1		23	15	110	23	46	30	50	36	17	12	11	1	35	2	2	16	0	0	1	0	4	3	1	12	10	460
2			8	36	23	36	28	63	24	12	9	8	0	30	0	4	15	0	2	1	0	2	0	1	9	3	314
3				14	7	9	20	28	18	8	7	10	2	19	4	5	7	0	0	0	0	4	0	1	4	6	173
4					9	10	19	23	27	15	17	8	1	46	3	6	26	7	9	1	0	6	4	4	13	17	271
5						1	9	7	10	6	8	7	2	16	1	3	8	4	2	2	1	3	0	0	11	7	108
6							2	19	1	6	6	6	2	37	0	1	12	1	3	2	0	4	1	0	12	13	128
7								13	24	9	16	9	3	37	1	13	17	1	2	0	0	4	1	1	8	8	167
8									5	6	20	11	1	40	0	8	11	0	2	4	0	4	0	0	3	13	128
9										0	7	2	0	35	2	9	12	1	0	2	2	6	0	3	21	12	192
10											1	0	3	32	1	5	18	2	3	2	0	2	0	1	30	10	110
11												1	2	54	2	4	14	0	5	1	0	4	2	2	25	14	130
12													0	33	1	2	16	1	2	0	1	4	2	0	7	6	75
13														10	0	0	3	1	1	0	0	1	2	0	11	9	38
14															4	36	23	11	26	8	6	17	11	2	171	96	411
15																7	17	6	13	3	2	6	11	2	44	42	153
16																	3	4	3	0	2	8	7	2	41	23	139
17																		4	8	9	7	9	10	5	15	20	87
18																			5	1	3	5	7	0	16	15	52
19																				1	1	8	3	2	20	24	59
20																					0	4	7	2	31	13	57
21																						8	7	2	103	34	154
22																							9	3	102	42	186
23																								0	51	28	113
24																									1	4	5
25																										7	7
26																											0
Total Alighting	0	23	23	160	62	102	108	203	145	79	107	74	17	449	21	105	231	44	93	41	28	113	87	40	844	518	3717



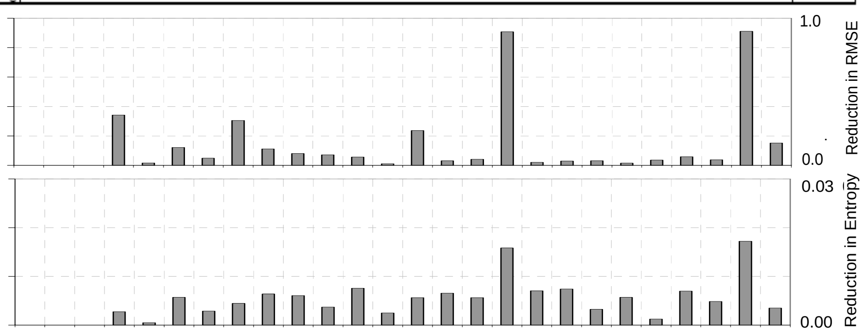
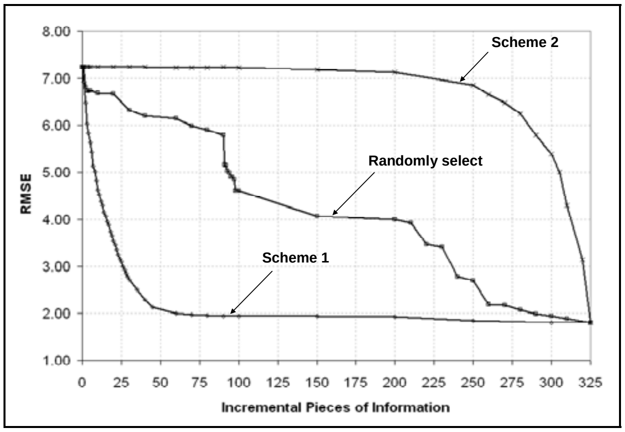


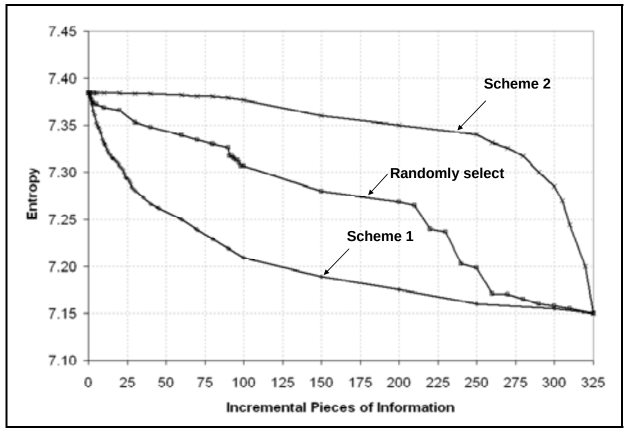
Figure 1: The actual O-D table and the reduction in RMSE and entropy when a set of O-D volumes from/to a particular stop are added

<u>Select the O-D pair from the row</u>. The next step is, among the O-D pairs in the selected row, to select the O-D pair and incorporate its approximate value into type 3 constraint 2. We test two schemes of sequentially incorporating the approximate values of the O-D pairs in the constraint. Both cases, we added the spot information (in the form of the approximate value) incrementally and examined how the values of the RMSE and the entropy change.

Scheme 1 identifies the cell that has the largest difference between the previously derived value and the actual value, and adds the approximate information in the constraint. Scheme 2 identifies the cell that has the smallest difference between the previously derived value and the actual value, and adds the approximate information about the trips. In both cases, the added approximate value is represented by the base of the membership function, ±20% of the actual O-D value.

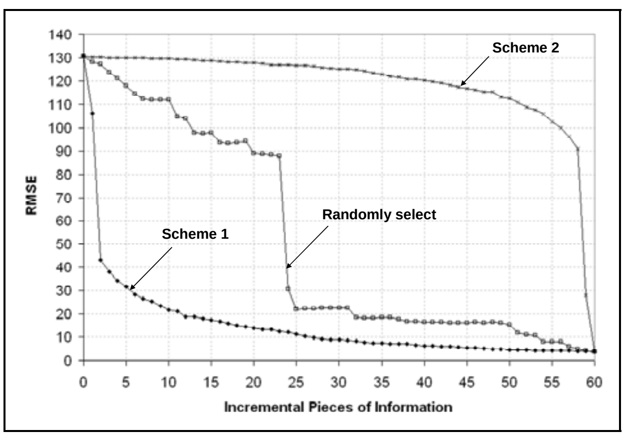
Figures 2 and 3 show the effects of the incremental addition of the approximate information under the two schemes on the values of the RMSE (Figure 2(a) for Osaka line and Figure 3(a) for Yokohama line), and the entropy (Figure 2(b) for Osaka line and Figure 3(b) for Yokohama line). It is seen that Scheme 1 is much more efficient in reducing these values than Scheme 2. This means that the way to add information is to target the cells whose value appears to be very much "off" from the actual value. Also shown in Figures 2 and 3 is the case that information is added in random; in this case, the trend line falls between Schemes 1 and 2.

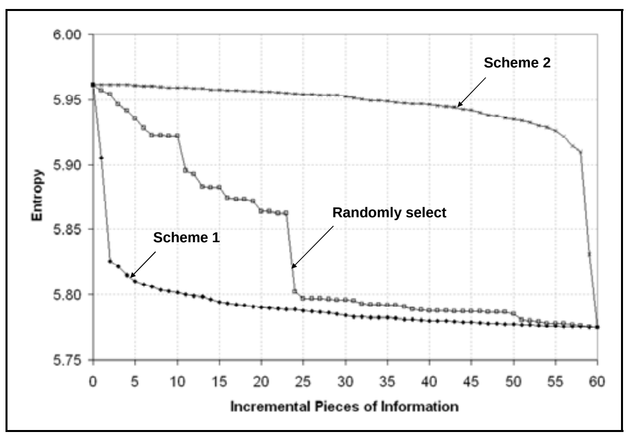




2(a) RMSE vs. amount of information **2(b) Entropy vs. amount of information**

Figure 2: Effects of additional spot information on reduction of RMSE and entropy (Hankai Line, Osaka)





3(a) RMSE vs. amount of information **3(b) Entropy vs. amount of information**

Figure 3: Effects of additional spot information on reduction of RMSE and entropy (Yokohama Seaside line, Yokohama)

Suppose that the analyst obtains information for the value of a specific cell from two different sources; one source says that the value is small, and the other source says the value is large. Use of these pieces of conflicting information for the same cell can be incorporated in the model, and a solution that reflects a compromise of the two pieces of information can be derived. It is interesting, however, to note that if the conflicting information is given in two crisp values, then solution is not feasible. In other words, fuzzy information allows finding the compromise between two conflicting pieces of information.

Conclusion

This paper presents a multi-objective fuzzy optimization model that calibrates the elements of an O-D table using limited information as the constraint. It then analyzes how the available limited information should be picked and incorporated in the constraint. It is found that (i) the more accurate information, if available, is given, the OD table approaches the real values quickly with less pieces of information; (ii) the values of the O-D pairs whose row and column show a very large total boarding or alighting should be picked. These findings should be useful when calibrating the transit O-D table with limited information.

References

1. Kikuchi, S. and N. Kronprasert. Constructing a Transit Origin-Destination Table Using the Uncertainty Maximization Concept. 88th Transportation Research Board Annual Meeting CD ROM Paper 09-2686, January 2009 (*In Press*).
2. Kikuchi, S. and N. Kronprasert. Constructing Transit Origin-Destination Tables from Fragmented Data. Accepted for Presentation at 89th Transportation Research Board Annual Meeting CD ROM Paper 10-3305, January 2010.
3. Klir, G.J. and M.J. Wierman. *Uncertainty-Based Information: Elements of Generalized Information Theory (Studies in Fuzziness and Soft Computing)*. 2nd Edition. Heidelberg: Physica-Verlag, 1999.
4. Klir, G. J. *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley and Sons, 2006.
5. Kikuchi, S., S. Mangalpally, and A. Gupta. A Method to Balance the Observed Boarding and Alighting Counts for a Transit Line. In *Transportation Research Record: Journal of the Transportation Research Board No.1971*, TRB, National Research Council, Washington D.C. 2006, pp.43-50.