A GENETIC ALGORITHM FOR THE ESTIMATION OF TRANSIT ORIGIN AND DESTINATION MATRIX USING TRAFFIC COUNT TECHNIQUES

Shinhae LEE
Researcher
Department of Transport Studies
Seoul Development Institute
Yejang Dong Chung-Ku, Seoul Korea
Tel:+82-2-726-1277
Fax:+82-2-726-1291
Email:newsun@sdi.re.kr

Seungjae LEE
Assistant Professor
Dept of Transportation Engineering
The University of Seoul
Dongdaemoon-Ku, Seoul, Korea.
Tel:+82-2-2210-2172
Fax:+82-2-2215-5097
E-mail: sjlee@uoscc.uos.ac.kr

Sooyeon MOON
Graduate Student
University of Seoul

Kangwon LIM
Professor
Seoul National University
E-mail: kwlim@snu.ac.kr

Abstract: The objective of this study is the development of the estimation technique for the transit origin and destination matrix using counted data. This problem can be described and formulated as a bilevel network design problem. The objective function is the minimization in both of the based and estimated transit origin and destination matrix, and the assigned and counted person data. In our studies, a genetic algorithm has been developed in order to obtain a global solution using the Spiess and Florian's transit assignment rule. The comparison of the estimated transit origin and destination matrices using each of the developed Genetic Algorithm and the Gradient algorithm based on the Spiess and Florian transit assignment algorithm has been conducted using a small example transit network. The results show that the GA method and the Gradient method reproduced similar assigned values as the counted values.

Key Words: Transit Origin and Destination Matrix, Traffic Counts, Gradient algorithm, Genetic algorithm, Transit assignment

1. INTRODUCTION

The origin and destination (O-D) matrix is one of the most important elements in transportation planning process. The accuracy of the O-D matrix plays key roles in the transport planning process in order to make and evaluate various transport policies. However, due to the nature of the O-D matrix, which is the desired people's or freight's movements on urban and regional space, it is very difficult and costly to estimate the O-D matrix. Traditionally, transport planners survey the O-D movements in order to estimate the O-D matrix. Even though the cost of the O-D survey requires high amounts of resources, the accuracy is relatively low. Even more, in the developing country, the transportation situation has changed very quickly and thus the transportation environment has been unstable. So the transportation planning should be frequently rectified according to the newly planned environment in order to capture the changed situation in the limited cost and time.

Fortunately, Intelligent Transport Systems have been introduced and deployed many high-tech and reliable surveillance systems. These enables to obtain and use the accurate data for the estimation and prediction of the O-D matrix. In particular, the information on each of the people's origin and destination movements between transit stations can be quite accurately obtained using the database of the ticket sales and ticket gates. In conjunction with these entrance and exit information on the ticket gates, the person trip survey database can be used for the estimation of the transit origin and destination matrix. The person trip surveyed data has been outdated because the survey has been conducted every five years, and in its nature, it requires a huge amount of costs and times.

Therefore, the objective of this study is the development of the estimation technique for the transit origin and destination matrix using counted entrance and exit data in ticket gates. The base transit origin and destination matrix can be initially estimated in the base of the expansion techniques using the person-surveyed data. Then the matrix has been updated using the entrance and exit counted data in ticket gates. This problem can be described and formulated as a bilevel network design problem. The objective function is the minimization in both of the based and estimated transit origin and destination matrix, and the assigned and counted person data. The constraints are the person movement conservation equation between origin and destination, and a transit assignment principle based on a transit assignment rule. In our studies, we assume transit assignment follows the Wardrop's first principle, and the Spiess and Florian's transit assignment rule. The nature of this problem leads to have multiple equilibria. In our studies, a genetic algorithm has been developed in order to obtain a global solution using the Spiess and Florian's transit assignment rule. The comparison of the estimated transit origin and destination matrices using each of the developed genetic algorithm and the gradient algorithm based on the Spiess and Florian transit assignment algorithm has been conducted using a small example transit network, which has 3 lines, 8 nodes and 18 links.

2. EXISTING STUDIES

Since the demand data cannot be observed directly, it must be collected by carrying out elaborate and expensive surveys, involving home or road based interviews or complicated number plate tagging schemes. By contrast, observed link volumes can be obtained easily either manually or automatically using mechanical or inductive counting devices. Thus, a considerable amount of research has been carried out to investigate the possibility of estimating or improving an origin-destination demand matrix with observed volumes on the links of the considered network. Many models have been proposed in the past such as Van Zuylen and Willumsen (1980), Van Vliet and Willumsen (1981), Nguyen (1982), Van Zuylen and Branston (1982) and Spiess (1987) among others. Most of these traditional approaches can be formulated as convex optimization problems in which the objective function corresponds to some distance function between a priori demand matrix and the resulting demand. The constraints are then used to force the assigned volumes to correspond to the observed volumes on the count post links.

2.1 Transit Assignment based on Optimal Strategy (Spiess and Florian, 1989)

The fixed cost transit assignment model based on optimal strategies is briefly described. For a more detailed description of the model, see Spiess (1984) and Spiess and Florian (1989). For easier presentation of the model, the transit network is assumed to be represented by a standard node/link type network, where a set of nodes is connected by a set of links. A travel time (or cost) and a service frequency is associated with each network link. The demand between nodes is given priori. The itineraries of the transit lines are implicitly contained in the network topology. The set of nodes not only contains the physical nodes of the underlying street or rail network, but also one additional node for each transit stop of each line. The links are subdivided into various classes, such as boarding, alighting, in-vehicle and walking links. Only boarding links imply waiting, thus have a finite frequency. All other links are served continuously. The waiting time at a node depends on the set of attractive links, i.e. the set of outgoing links which are considered for travel by the travelers by boarding the first vehicle leaving on any of these links. For any given set of attractive links a node, the combined waiting time is proportional to the combined total frequency of these links and the probability of leaving nodes on links.

Given the above relations, any strategy for reaching destination is completely defined by the corresponding subset of attractive links. The optimal strategy for reaching a destination is the one which minimizes the total expected cost. The cost of a strategy is the sum of link travel times weighted by the probability of traveling on links, and the waiting time at nodes weighted by the probability of traveling through nodes. For fixed link travel times, the assigning of the trips from all origins to destination according to the optimal strategy corresponds to solving the linear optimization problem. The problem (see Spiess, 1993) can be solved very efficiently by means of the label-setting type algorithm.

2.2 The Gradient Approach for solving the demand estimation using traffic counts (Spiess, 1990)

Spiess(1990) proposed the gradient approach in order to estimate the O-D matrix using traffic counts. It is formulated as an optimization problem so as to minimize a measure of distance between observed and assigned volumes. The simplest function of this type is the square sum of the differences, which leads to the convex minimization problem. This is subject to the pseudo-function used to indicate the volumes resulting from an assignment of the demand matrix. The particular assignment model used must correspond to a convex optimization problem, in order for the objective function to be convex. Since the matrix estimation problem as formulated in the Spiess(1990) is highly underdetermined, it usually admits an infinite number of optimal solutions, i.e. possible demand matrices which all reflect the observed volumes equally well. In the actual planning context, we expect the resulting matrix to resemble as closely as possible the initial matrix, since it contains important structural information on the origin-destination movements. Therefore, just finding one solution to the problem in Spiess(1990) is clearly not enough.

If we would have a solution algorithm which inherently finds a solution close to the starting point, we could leave the objective function as is. Fortunately, the gradient method, also called the method of steepest descent, has exactly this property that we look for. It follows always the direction of the largest yield with respect to minimizing the objective function and, thus, it also assures us not to deviate from the starting solution more than necessary. In order to implement the gradient method, we also need to provide values for the step lengths. Choosing very small values for the step length has the advantage of following more precisely the gradient path, but has the disadvantage of requiring more steps. On the other hand, when choosing too large values for the step length, the objective function can actually increase and the convergence of the algorithm would be lost. Thus, the optimal step length at a given demand can be found by solving the one-dimensional subproblem.

3. PROBLEM FORMULATION

The transit assignment based on the shortest path problem is concerned with finding the path with minimum travel time including waiting and transfer time from a source to a destination through a directed network. Let G=(N, A) be a directed network with node set N=1, ..., n and arc set $A \subset N \times N$. Each arc is denoted by a pair of nodes (i, j), where $i \in N$ and $j \in N$. If A is changed to a set of ordered pairs of distinct elements of N, then set G=(N, A) is a directed network and A is the set of ordered pairs (i, j). The ordered pairs (i, j) are referred to as arcs or links and an arc goes from node i to node j. A path from a node s to a node d is a sequence of arcs p=(s, i_1), (i_1, i_2), ..., (i_k, d) in which the initial node of each arc is the same as the terminal node of the preceding arc in the sequence. In other words, node d can be reached from node s.

A path p is open if $s \neq d$ and is closed if s = d. A cycle is a closed path p in which no nodes are repeated except s and there exists at least one arc. A network that contains no cycles is called acyclic. If each arc (i, j) has an associated time t_{ij}, then a path p has an associated distance equal to the sum of the distances of the arcs in the path. This in turn gives rise to the shortest path problem, which is to find the path with minimal distance between two nodes s and d. In a given network both a source node and destination node can be designated. These are interpreted as terminal nodes at which some activity begins and ends. In an acyclic directed network with n nodes, the source can be labeled as node 1 and the destination as node n and all other nodes can be labeled such that for any arc (i, j), i<j.

Let t_{ij} represent the travel time associated with the arc (i, j). If a set of all paths between a predetermined pair of nodes, say s=1 and d=n, is denoted as a $P(p \in P)$, then the transit assignment based on the shortest path problem can be formulated as:

$$\min z(p) = \sum_{(i,j)\in p} t_{ij} \tag{1}$$

where z is to minimize the travel time taken to traverse from a source to a destination

4. GENETIC ALGORITHM APPROACH

The GA is a method for approximate optimization simulating the process of natural evolution. In the GA, candidate solutions are represented by genetic strings. In addition, a diverse and random population of such genetic strings must be created to represent an initial population. The genetic string is a called a chromosome and the population at a given time is termed a generation. The population of chromosomes is maintained by the generations. At each generation, each chromosome is evaluated by a fitness function and the chromosomes is then selected to reproduce next generation on the basis of their relative fitness. The selected chromosomes are recombined using genetic operators such as crossover and mutation to form the new population. In each generation, the chromosomes in terms of relatively less fitness are removed, and a set of chromosomes with relatively high fitness is reproduced to maintain a population of chromosomes.

4.1. Chromosome representation

In constructing a GA for a specified problem, the first step is designing an appropriate chromosome representation, i.e. encoding, of the problem. In this research, a new approach, namely link based representation, is proposed to create an initial population by focusing attention on links consisting a network. The approach is to use the network itself instead of a path as a chromosome. Then, each component of the chromosome represents a degree of the contribution of each link for forming a path connecting an origin and a destination. A chromosome is represented as a real number string between 0 and 1 with the length of the number of links consisting a given network and is given as

$$C = (c[1]\ c[2]\ \cdots\ c[n-1]\ c[n]), \tag{2}$$

where each component c[i] (i=1,2,...,n), $0 \leq c[i] \leq 1$, is the contribution value of the link i in the network. If $c[i] \geq 0.5$, then the link i is defined as a real link with the original values(time and cost) of link i. Otherwise, the dummy link with the infinite value replaces the link i.

The chromosome C assigns the real links or the dummy links for each link in the original network. Because the link values have been changed, the chromosome C creates another network of a new structure from the original network. It is also possible to generate a unique path from the newly constructed network. Therefore this encoding method is capable of equally representing all possible paths(candidate solutions) for the original network. Also, the advantage of this string representation is the simplicity of the chromosome structure and permitting the application of conventional genetic operators.

4.2. Evaluation

The next stage is to convert a chromosome into a path, i.e. decoding, and to calculate the fitness of a chromosome. In our encoding procedure, the chromosome representation generates a new structure of the network. It is easy to generate a unique path from the network reconstructed by the chromosome.

During each generation, chromosomes are evaluated using some measure of fitness instead of an objective function. In most optimization applications, fitness is calculated based on the original objective function. Because we treat with a minimization problem, we must convert the objective value for each chromosome into the fitness value, so that a fitter chromosome has a larger fitness value. This can be simply done by the inverse of its objective value as follows:

$$Eval(p) = 1 / z(p) \tag{3}$$

where eval(p) is the fitness value for the chromosome p and z(p) is the objective value of the chromosome p.

4.3. Selection

The selection is a mechanism to reproduce next generation from current population based on the fitness function, in which a fitter chromosome has a larger chance of being selected to be reproduced into next generation. The fitness is calculated based on the original objective function, and the value of a chromosome is a scalar within certain boundaries.

Schaffer(1989) was the first one to attempt this approach and proposed the Vector Evaluated Genetic Algorithm. In the VEGA, a population was divided into exclusive n sub-populations for a problem with n objectives and each sub-population favored its own objective function. Then, a general selection scheme is used to select the parents for reproduction from each sub-population. Thus it tried to find solutions in the neighborhood of the extreme points on the optimal frontier. Some studies have been attempted to overcome the problem with VEGA and also the traditional elitist strategy for single objective optimization problem can be selection scheme to distribute the population along the optimal frontier. In this paper we use combined method of Schaffer's VEGA and the elitist strategy to give out a set of optimal solutions. In each generation solutions are updated and eventually, optimal solutions are searched through the mechanism of natural evolution such as crossover and mutation.

4.4. Genetic operators

The crossover is an exchange of a portion of the chromosomes. The operator combines the components of two parent chromosomes to form one or more offsprings. From a characteristic of chromosome structure for our problem, a new method, called the one-point weighted sum crossover, is developed. In the proposed method, the offsprings inherit the weighted sum of components of two parent chromosomes. To illustrate, consider a network with n links and the following parent chromosomes:

$$\begin{aligned} A &= (a[1]\ a[2]\ \cdots\ a[n-1]\ a[n]) \\ B &= (b[1]\ b[2]\ \cdots\ b[n-2]\ b[n]) \end{aligned} \tag{4}$$

In applying the one-point weighted sum crossover to parents A and B, we first choose a crossover site. The position is easily obtained by using a integer k(truncate n/2). If the number of links n in a given network is even, then the offsprings are created as

$$\begin{aligned} A' = a'[i] &= \alpha \times a[i]+(1-\alpha)\times b[i+k] \quad \text{if } i\leq k \text{ for all } i=1, 2, ..., n, \\ &= a[i] \quad \text{if } i>k \quad \text{for all } i=1, 2, ..., n, \\ B' = b'[i] &= \alpha \times b[i]+(1-\alpha)\times a[i+k] \quad \text{if } i\leq k \text{ for all } i=1, 2, ..., n, \\ &= b[i] \quad \text{if } i>k \text{ for all } i=1, 2, ..., n. \end{aligned} \tag{5}$$

And those for a network with an odd number n are as follows:

$$\begin{aligned} A' = a'[i] &= \alpha \times a[i]+(1-\alpha)\times b[i+k+1] \quad \text{if } i\leq k \text{ for all } i=1, 2, ..., n, \\ &= a[i] \quad \text{if } i>k \text{ for all } i=1, 2, ..., n, \\ B' = b'[i] &= \alpha \times b[i]+(1-\alpha)\times a[i+k+1] \quad \text{if } i\leq k \text{ for all } i=1, 2, ..., n, \\ &= b[i] \quad \text{if } i>k \text{ for all } i=1, 2, ..., n, \end{aligned} \tag{6}$$

where α is a random real number with values $0\leq \alpha < 1$. When $\alpha = 1$, the offsprings A' and B' correspond to the parent chromosomes A and B respectively

The mutation is an occasional random alteration process in a small probability. The operator arbitrarily alters one or more components of a chromosome. To increase the variability of the population, we arbitrarily alter all of components of a selected chromosome by some random numbers in the range between 0 and 1.

5. NUMERICAL EXAMPLE

The evaluation of the GA approach is conducted using a small example network shown in Figure 1. The GA is implemented using Turbo C++ on an IBM/PC compatible whilst the gradient method is carried out using EMME2. Tables 1, 2 and 3 give some characteristics in number of traffic zones, number of network links and lines, number of links with observed volumes. The values are given in Table 4 between observed and assigned volumes using the initial demand in Table 3 and the adjusted demands in Table 5 and 6 after the adjustment using the gradient method and the GA method. These results show both methods are very good in terms of the similar values reproducible. In particular, the structure and the degree of magnitude in the initial O/D matrix are maintained after the adjustment process has been carried out using the Gradient method and the GA method. And the assigned link flows in both methods have produced similar counted values in links. However, the GA method has produced more close values to the initial demand and the counted link flow values than those of the Gradient method.

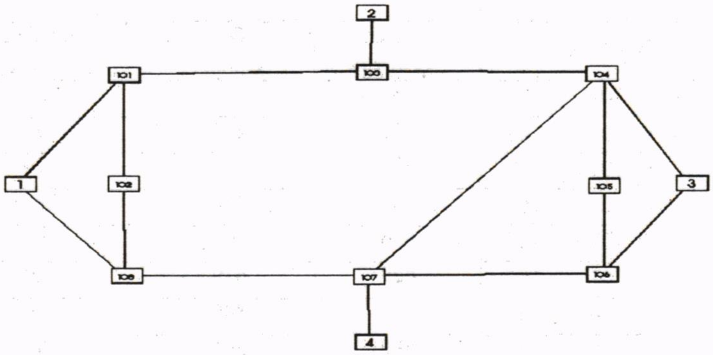

Figure 1. Example network

Table 1. The attribute of analysis network

Division	Number	Note
Zone	4	1~4
Node	8	101~108
Connector	12	-
Link	18	-

Table 2 Transit line data

Division	Headway (mi)	Capacity (seated/tot)	Line capacity (person)	Initerary (circulation)
Line 1	5	50/100	500	· 101, 103, 104
Line 2	5	50/100	500	· 106, 107, 108
Line 3	5	50/100	500	· 104, 107

Table 3. The Initial O/D Matrix

O \ D	1	2	3	4	Sum
1	-	257	374	299	930
2	318	-	288	235	841
3	455	501	-	485	1,441
4	201	299	325	-	825
Sum	974	1,057	987	1,019	4,037

Table 4. The assignment results of transit line segment

Division	Itinerary		Gradient Method	GA Method	link data (observed)
Result	line1	101 → 103	231	231	231
		103 → 104	472	472	471
		104 → 103	720	720	720
		103 → 101	286	286	286
	line2	108 → 107	606	605	606
		107 → 106	629	629	629
		106 → 107	846	846	846
		107 → 108	590	591	590
	line3	107 → 104	269	269	269
		104 → 107	211	212	212

Table 5. A revised O/D matrix after applying the gradient method

O \ D	1	2	3	4	Sum
1	-	231	325	280	837
2	286	-	260	211	757
3	398	451	-	447	1,297
4	191	268	303	-	763
Sum	875	951	889	939	3,655

Table 6. A revised O/D matrix after applying the GA method

O \ D	1	2	3	4	Sum
1	-	251	323	294	868
2	295	-	267	230	792
3	398	468	-	466	1,332
4	197	274	328	-	799
Sum	890	993	918	990	3,791

6. SUMMARY

The proposed GA approach and the Gradient algorithm provided in EMME 2 has been tested on a small size network with 8 nodes and 18 links and 3 lines of the 4 origin and destination matrix. The results showed that the GA algorithm can find the known solutions. In the test carried out, the gradient method also was found to perform very well. These results show both methods are very good in terms of the similar values reproducible. In particular, the structure and the degree of magnitude in the initial O/D matrix are maintained after the adjustment process has been carried out using the Gradient method and the GA method. And the assigned link flows in both methods have produced similar counted values in links. However, the GA method has produced more close values to the initial demand and the counted link flow values than those of the Gradient method. However, we need to conduct more experiments in terms of the size of the network and the degree of the discrepancy of the counted values and assigned values in various ways to see if these methods are applicable in real practice.

REFERENCES

a) Books and Books chapters

INRO Consultants Inc. (1989), **EMME/2** User's Manual.

b) Journal papers

Solomom, M.(1987) Algorithms for the vehicle routing and scheduling problems with time window constraints, **Operations Research 35**, 254-265.

Chen, Y and K. Tang,(1998) Minimum time paths in network with mixed time constraints, **Computers and Operations Research 25**, 793-805.

Spiess H. (1987). A maximum likelihood model for estimating origin-destination matrices. **Transpn. Res. B 21,** 395-412.

Spiess H. and Florian M. (1989). Optimal strategies: A new assignment model for transit networks **Transpn. Res. B 23**, 83-102.

Van Zuylen H.J. and Branston D.M. (1982). Consistent link flow estimation from counts. **Transpn. Res. B 16,** 473-476.

Van Zuylen H.J. and Willumsen L.G. (1980). The most likely trip matrix estimated from traffic counts. Transpn. Res. B 14, 281-293.

c) Papers presented to conferences

Grefenstette, G, R. Gopal, B. Rosmaita and D.V. Gucht,(1985) Genetic algorithms for the traveling salesman problem, in: J. Grefenstette(Ed.), **Proceedings of the First International Conference on Genetic Algorithms,** Lawrence Erlbaum, Hillsdale, NJ, pp. 160-168.

Oliver, I, D.J. Smith and J.R.C. Holland,(1987) A study of permutation crossover operators on the traveling salesman problem, in: J. Grefenstette(Ed.), **Proceedings of the Second International Conference on Genetic Algorithms**, Lawrence Erlbaum, Hillsdale, NJ, pp. 224-230

Whitley, D, T. Starkweather and D'A. Fuquay, (1989) Scheduling problems and traveling salesman: The genetic edge recombination operator, in: J. Schaffer(Ed), **Proceedings of the Third International Conference on Genetic Algorithms**, Morgan Kaufmann, Los Altos, CA, pp. 2-9.

Brenninger-Göthe M., Jörnsten K.O. and Lundgren J.T. (1988). Estimation of origin-destination matrices from traffic counts using multi-objective programming. formulations. **Publication 88-14**, Department of Mathematics, University of Linköping.

Nguyen S. (1982). Estimating origin-destination matrices from observed volumes. **Proceedings of the First Course on Transportation Planning Models of the International**

School of Transportation Planning, Amalfi, Italy, October 1982.

Van Vliet D. and Willumsen L.G. (1981). Validation of the ME2 model for estimating trip matrices from traffic counts **Proceedings of The Eights International Symposium on Transportation and Traffic Theory,** June 1981.

Willumsen L.G. (1984). Estimating time-dependent trip matrices from traffic counts. **Ninth International Symposium on Transportation and Traffic Theory**, VNU Science Press, 397-411.

Wardrop J.G. (1952). Some theoretical aspects of road traffic research. **Proc. Inst. Civil Engineers, Part II 1,** 325-378.