Efficient Implementation of Bilevel Programming Model for Origin-Destination Estimation

(OD交通量の推定におけるツーレベルモデルの効率的計算法の開発)

Lin Cheng[*] Yasunori Iida[**] Nobuhiro Uno[***] Wei Wang

1. Introduction

The origin-destination (OD) trip matrix is a fundamental input for most problems of traffic planning and management. In practice, the true OD matrix is seldom, if ever, available and various methods can be used for its estimation. Many researchers have been considering the estimating of origin-destination trip matrix from traffic flows. Common methods include entropy maximizing (Van Zuylen and Willumsen, 1980), maximum likelihood (Spiess, 1987), generalized least squares (Cascetta, 1984;) and Baysian inference estimation techniques (Maher, 1983). These models have in general the form:

$$\min_{\mathbf{x},\mathbf{t}} \; F_1(\mathbf{t},\bar{\mathbf{t}}) + F_2(\mathbf{x},\bar{\mathbf{x}}) \qquad (1a)$$

subject to $\quad \mathbf{x} = P\cdot\mathbf{t} \qquad (1b)$

where

1) $\bar{\mathbf{t}} = (\Lambda, \bar{t}_w, \Lambda)^T$ is column vector of the target OD matrix represented as, $\bar{t}_w$ = the trips between OD pair $w \in W$, and W is the set of origin-destination pairs.

2) $\bar{\mathbf{x}} = (\Lambda, \bar{x}_a, \Lambda)^T$ is column vector of observed link flows, $\bar{x}_a$ is the flow observed on link, $a \in \bar{A}$, and $\bar{A}$ is a subset of links in the network with observed flows.

3) $\mathbf{t}$ and $\mathbf{x}$ are vectors representing the OD matrix and corresponding link flows to be estimated, respectively.

4) $F_1(\mathbf{t}, \bar{\mathbf{t}})$ and $F_2(\mathbf{x}, \bar{\mathbf{x}})$ are functions of the generalized distance measurement, or errors between $\mathbf{t}$ and $\bar{\mathbf{t}}$, and $\mathbf{x}$ and $\bar{\mathbf{x}}$, respectively. For example, the entropy function and Euclidean distance function are frequently employed.

4) $P = [p_{aw}]$ is referred to as the assignment proportion matrix, which describes the relationship between the estimated variables of link flow and OD matrix. It is well known that the flow passing through each network link is a linear combination of the elements of the OD matrix with coefficients comprised between 0 and 1.

The assignment matrix's element p_{aw}, called link choice proportion, denotes the fraction of trips of the OD pair w contributing to the flow of link a. Obviously the elements of matrix P is not known exactly, but they must be obtained from one of the various assignment models. Some of these models assume that the assignment matrix is independent of OD matrix and therefore no route choice mechanism is incorporated (Cascrtta, 1984), or say no congestion phenomenon is considered (Bell, 1991); others assume that it depends on the OD matrix in the deterministic way of user equilibrium (Yang *et al*, 1992).

Most existing models belonging to family (1) hold the fixed assignment matrix in the sense that the assignment matrix P is predetermined given. Estimation models with a fixed assignment matrix have a computational advantage because this formulation can be cast into the form of single convex optimization problems. In particular, the statistical characteristics of the estimates can be obtained analytically, allowing for a better understanding of the improvement produced with the use of traffic flow-based information.

The assumption of a fixed assignment matrix, however, has inherent shortcomings (Yang *et al*, 1991, 1992). Because the demand matrix is estimated from observed link flows with fixed route choice proportions, and the demand matrix is, in general, assigned to the network with user equilibrium, there is an inconsistency in using one set of route choice proportions to obtain a trip matrix from link flows, and another to obtain the link flow distribution by assigning the trip matrix to the network. In a network with realistic congestion, the shortcoming becomes more apparent. Furthermore, the errors arising from the predetermined link use proportions and traffic flow observations may lead to an inconsistent system of

Keyword: bilevel model, user equilibrium, least square, path-based algorithm

*Regular member of JSCE, Associate Professor, Transportation College, Southeast University, SiPaiLou 2, Nanjing, China, T: 025-379-2233, F: 025-3794102
tei5126@yahoo.co.jp
**Fellow of JSCE, Ph.D., Professor, Department of Civil Engineering, Kyoto University
***Regular Member of JSCE, Ph.D., Associate Professor, Department of Civil Engineering, Kyoto University

equations (1b). These deficiencies have motivated researchers to attempt to estimate OD matrices and route choice proportions simultaneously.

Network equilibrium-based methods include the entropy-maximizing model constrained by user equilibrium (Fisk, 1988), and the calibration of the combined distribution/assignment model with link traffic flow data (Fisk and Boyce, 1983). Because of the existence of measurement errors, as well as temporal fluctuations, observed link flows may not completely satisfy the user equilibrium conditions, even if they are consistent (Yang *et al*, 1991). In practical situations, the assumptions underlying the equilibrium-based model formulations should be relaxed. Consequently, a model that allows for uncertainty and/or non-equilibrium in traffic flows becomes necessary.

A bilevel model was proposed by Yang *et al* (1992), for OD matrix estimation in a congested network case. The model allows for uncertainties in both the target OD matrix and traffic flow data while incorporating user equilibrium conditions. The upper-level problem seeks to minimize the sum of distance measurements shown in (1a), and the lower-level problem represents a user optimal assignment (1b) ensuring that the estimated OD matrix and corresponding link flows satisfy the user equilibrium conditions. A heuristic algorithm for solving bilevel programming problem is proposed and a simple numerical simulation experiments are implemented. However, their algorithm has a significant shortcoming, since a particular path flow pattern has to be chosen using the Frank-Wolfe algorithm. As known, the Frank-Wolfe algorithm is inherently link-based, and tend to be inefficient to obtain the link flow components of each origin-destination pair.

In this paper, a path-based algorithm for user equilibrium model is adopted for facilitating solution to the OD matrix estimation problem from link traffic flows. The algorithm may efficiently calculate and store path flow pattern, and easily transfer the information of each origin-destination demand contributing to link flow pattern. Such implementation initiates the application of bilevel models to the OD estimation problem of large-scale transportation network.

2. Bilevel Model for Estimating OD Matrices

A bilevel programming approach can be used as a general technique for OD matrix estimation in congested networks. The bilevel problem is also called a Stackelberg game with an upper-level vector of decision variables **t** for the leader, and a lower-level vector of decision variables **x** for the follower. It is assumed that the leader firstly selects **t** in accordance with his constraints to minimize the leader's objective function, while taking into account the reaction of the follower. In light of this decision, the follower selects an **x** according to the corresponding constraints to minimize the follower's objective function.

2.1 The Upper-Level Formulation

The generalized least squares (GLS) estimation method can be extended network case with uncertain data in the form of a bilevel programming problem. The GLS method has the advantage that it permits the weighted combination of survey and traffic flow data. It does not require flows for all links, nor does it require flows to be error free.

Assuming that the target matrix $\bar{\mathbf{t}}$, is obtained from sample surveys or model estimates, and $\bar{\mathbf{x}}$ is the observed link flows. The relationship between the observed and the estimated variables can be illustrated by the random system: $\bar{\mathbf{t}} = \mathbf{t} + \xi$, $\bar{\mathbf{x}} = \mathbf{x} + \boldsymbol{\eta}$, where ξ and η are vectors of random error terms reflecting uncertainty in the survey. It is often assumed that ξ and η have zero mean. $E(\xi) = \mathbf{0}$, $E(\eta) = \mathbf{0}$. The GLS estimator that belongs to family (1) can be obtained by minimizing the weighted sums of squared distance between observed and estimated variable values.

$$\min_{\mathbf{t}} (\bar{\mathbf{t}} - \mathbf{t})\mathrm{U}^{-1}(\bar{\mathbf{t}} - \mathbf{t}) + (\bar{\mathbf{x}} - \mathbf{x})\mathrm{V}^{-1}(\bar{\mathbf{x}} - \mathbf{x}) \qquad (2)$$

where U and V are matrices, interpreted as the variance-covariance matrices of random error terms, ξ and η, respectively.

If we assume that the linear assignment map (2) is one that correctly represents the trip makers' route choice behavior, and further assume that the target matrix is an unbiased estimate and the bias of measurement error in the traffic flows is negligible ($E(\xi)=0$, $E(\eta)=0$), the solution to the GLS can be shown to be the best linear unbiased estimate of the true OD matrix (Cascetta, 1984).

Note that, when only a subset of links on the network are observed (i.e. $\overline{A} \subseteq A$), then the second term of the upper-level object function must be defined only for the links with observed flows.

2.2 The Lower-Level Formulation

We assume that user equilibrium assignment is a perfect representation of trip makers' route choice behavior. Namely, trip makers are assumed to choose routes in a user optimal manner, consistent with Wardrop's first equilibrium principle. It is well known that the user equilibrium problem are equivalent to the following non linear programming.

$$\text{minimize} \quad z(\mathbf{x}) = \sum_a \int_0^{x_a} t_a(x) \cdot dx \tag{3a}$$

$$\text{subject to} \sum\nolimits_k f_k^w = t^w \quad \forall w \in W \tag{3b}$$

$$x_a = \sum\nolimits_w \sum\nolimits_k f_k^w \cdot \delta_{ak}^w \quad \forall a \in A \tag{3c}$$

$$f_k^w \geq 0 \qquad \forall\, k \in K^w, w \in W \tag{3d}$$

where

f_k^w = flow on path $k \in K^w$

K^w = path set of OD pair $w \in W$

W = set of OD pair

δ_{ak}^w = 1 if path k uses link a, otherwise 0

A = link set in the network

$t_a(x)$ = travel cost function

We will adopt the path-based algorithm to solve the standard UE problem (3a)-(3d) to find link flow pattern and path flow pattern simultaneously. The available feasible path flow pattern will easily produce the corresponding assignment proportion matrix $P = [p_{aw}]$. Assignment matrix will play a significant role in the algorithmic efficiency of the bilevel optimization. Through the assignment proportion, the upper level will get the feedback of the follower's reaction to the leader's decision.

2.3 Information Exchange between Two Levels

Because of the positive definiteness of the quadratic form in (2), the estimator **t**, can be obtained by equating to zero the first partial derivatives of (2) with respect to **t**. The equation of the first-order condition becomes:

$$2(\mathbf{t} - \bar{\mathbf{t}})^T U^{-1} + 2(\mathbf{x} - \bar{\mathbf{x}})^T V^{-1} \nabla_{\mathbf{t}} \mathbf{x}(\mathbf{t}) = 0 \tag{4a}$$

where, $\nabla_{\mathbf{t}}\ \mathbf{x}(\mathbf{t})$ is the derivatives of link flow with respect to OD demand t, which can be actually obtained using sensitivity analysis method, but this is computationally burdensome. Since $\mathbf{x}(\mathbf{t})=P\mathbf{t}$, $\nabla_{\mathbf{t}}\ \mathbf{x}(\mathbf{t})$ might be approximated by the corresponding assignment matrix P. The first order condition of upper level problem is consequently simplified and becomes

$$(\mathbf{t} - \bar{\mathbf{t}})^T U^{-1} + (P\mathbf{t} - \bar{\mathbf{x}})^T V^{-1} P = 0 \tag{4b}$$

and then, the solution is finally expressed as

$$\mathbf{t} = (U^{-1} + P^T V^{-1} P)^{-1} (U^{-1}\bar{\mathbf{t}} + P^T V^{-1} \bar{\mathbf{x}}) \tag{4c}$$

Estimating the OD matrix requires the knowledge of the assignment matrix for solving the upper level problem. In this paper, the assignment matrix P relies on the OD matrix in the way of deterministic user equilibrium. For each solution of the upper level problem, the lower level decision variables **x** will be updated according to the assignment matrix $\mathbf{x} = P\mathbf{t}$. This indicates that the estimation problem holds Stackelberg leader-follower structure, i.e. the route choice mechanism is explicitly considered in the bilevel model.

The assignment matrix P is the result of the reaction of follower to the leader's decision, and can be obtained from the final path flow pattern. The element of P is calculated by

$$p_{aw} = \sum_k f_k^w \delta_{ak}^w \Big/ t_w \tag{5}$$

Yang *et al* (1992) use the Frank-Wolfe algorithm to get one particular path flow pattern, by storing the ultimate unique shortest path. The calculation of the assignment matrix is considerably tough and incomplete since the Frank-Wolfe algorithm is inherently link-based. On the contrary, this paper adopts the path-based Newton method (Cheng *et al*, 2002), which employs column generation technique, to produce multi-shortest paths for each OD pair. The multi-shortest paths share origin-destination demand, according to Wardrop's first principle, and readily produce the assignment matrix. This implementation makes bilevel model possible for the large-scale traffic network OD estimation

Although the path flow pattern is not unique in user equilibrium, any particular path flow pattern, as demonstrated in our example, could be chosen to determine the assignment proportion matrix P, and have no effect on the final estimate of OD matrix. Actually, the value of the upper level objective function, which is defined by vectors **t** and **x**, is independent of the values of path flows, hence the global optimal solution should be invariant to the variety of route choice proportions.

3. Solution Procedure

Bilevel programming problems are generally difficult to solve because evaluation of the upper-level objective function requires solving the lower-level optimization problem. Furthermore, since the lower-level problem is in effect a nonlinear constraint, the problem is a non convex programming problem. However, since the upper-level objective function of our problems (2) is strongly convex in variables **t** and **x**, a local minimum is likely to be a global minimum. This could be particularly likely when the target OD matrix, used as a starting solution, is closed to the true one.

This paper presents a state-of-the-art algorithm for transferring the follower's reaction to the leader. Accordingly, the leader adjusts his strategy in response to the latest action of the follower. This algorithm is more suited to do with large practical traffic network because it is path-based and assignment matrix necessary for information exchange is readily available. Moreover, the path-based Newton method, which precisely portrays route choice behaviors, might accelerate the convergence of the bilevel model. The calculation procedure iterates between the upper-level (least square estimation) problem and lower-level (user equilibrium) problem. The general scheme has the similar form (Yang *et al*, 1992):

Step0. Initialize $P^{(0)} = [p_{aw}^{(0)}]$; $n := 0$

Step1. Find $\mathbf{t}^{(n+1)}$ using $P^{(n+1)}$

Step2. Update $P^{(k+1)} = [p_{aw}^{(k+1)}]$ using $\mathbf{t}^{(k+1)}$

Step3. If convergence criterion is met, then stop ; else let $n := n+1$ and go to Step 1

The difference between bilevel models in this paper and existing papers (e.g. Yang *et al*, 1992) is that the assignment matrix, characterizing route choice behavior, is directly generated in the path-based Newton method.

4. Conclusion

OD matrix estimation has been considered by incorporating user equilibrium assignment into trip matrix estimation in the form of a bilevel programming problem. The contribution of this paper is relatively modest in that we extended the work of previous researchers to show efficient implementation in the solution to the bilevel model. Instead of the Frank-Wolfe algorithm, the path-based Newton method readily saves path flow information and OD demand is reasonably shared among the corresponding shortest paths saved. This makes implementation better than that of Yang *et al* in efficiency and convergence. This approach appears highly attractive, particularly since the assignment matrix in the other link-based approaches is not guaranteed. The new implementation could be adapted to practical traffic network inclusive of more nodes and links.

5. Reference

Bell. Estimation of origin-destination matrices by constrained generalized least squares. *Trans. Res.* 25B, 1991, 13-22.

Cascetta. Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. *Trans. Res.* 18B, 1984, 289-299.

Cheng, Iida, Uno. Application of Quasi-Newton Methods to Capacitated UE Assignment, 土木計画学研究論文集、Vol. 19, No 3, 2002

Fisk. On combining maximum entropy trip matrix estimation with user optimal assignment. *Trans. Res.* 22B, 1988, 66-79.

Fisk. Trip matrix estimation from link traffic counts. *Trans. Res.* 23B, 1989, 331-356

Fisk, Boyce. A note on trip matrix estimation from link traffic count data. *Trans. Res.* 17B, 1983, 245-250.

Maher. Inferences on trip matrices from observations on link volumes. *Trans. Res.* 17B, 1983.

Spiess. A maximum likelihood model for estimating origin destination matrix. *TransRes*. 21B,1987, 395-412

Van Zuylen, Willumsen. The most likely trip matrix estimated from traffic counts. *Trans. Res.* 14B, 1980, 281-293.

Yang, Iida, Sasaki. An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. *Trans. Res.* 25B, 1991, 351-363

Yang, Asakura, Iida, Sasaki. A bilevel programming model for estimating origin destination matrix from counts with congestion effects. 土木学会論文集 No 440/Ⅳ-16, 1992, 117-124