A Method for Spatiotemporal Estimation of Traffic States Using ETC 2.0 Data

School of Engineering, Hokkaido University	○Student Member	Kaifeng Duan
DOCON CO., LTD.	Nonmember	Keita Zushii
DOCON CO., LTD.	Member	Osamu Nagaoka
School of Engineering, Hokkaido University	Student Member	Yuto Kawamura
Graduate school of Engineering, Hokkaido University	Member	Ryuichi Tani
Faculty of Engineering, Hokkaido University	Member	Kenetsu Uchida

1. Introduction

To estimate the traffic conditions of the entire road network, link traffic flow and link travel time are collected by traffic detectors and probe cars, respectively.

Several studies have been conducted to estimate unobserved traffic conditions in networks. Tani and Uchida (2020) constructed a maximum likelihood estimation model with equilibrium constraints based on a traffic assignment model that considers the uncertainty of the road network and proposed a method to estimate the joint distribution of link traffic flow and delay time for the entire network for multiple time periods.

Kawamura et al. (2021) proposed a method for estimating the posterior distribution of traffic conditions. The probability distribution of traffic conditions is estimated from the maximum likelihood estimation model constructed by Tani and Uchida (2020) as the prior distribution, and then Bayesian estimation using traffic flow data obtained from traffic detectors and travel time obtained from probe data as the likelihood. The proposed method is based on the Bayesian estimation of the posterior distribution of the traffic condition.

However, because of the high cost of installing and operating traffic detectors, it is not realistic to install traffic detectors covering the entire road network due to financial constraints. Therefore, traffic detectors are installed in a limited number of links throughout the road network. Since the location of the detectors cannot be changed, and it is impossible to install traffic detectors throughout the network, incomplete and highly repetitive data is obtained. However, when using traffic observation data for traffic condition estimation, complete data can be required.

In order to solve the problems mentioned above that are caused by sparsely installed traffic detectors, this study usesECT2.0 data for spatiotemporal estimation of the traffic state of the road network. Compared with traditional traffic detectors, ETC 2.0 can collect information in the road network randomly instead of only targeting fixed locations. This will make our model more realistic. In a prior study, the maximum likelihood estimation method was used to estimate the spatiotemporal state of the road traffic network, and its feasibility was verified on a small test network. When solving the likelihood maximization problem for traffic state estimation, an optimization problem under equilibrium constraints is difficult to find the search direction of the solution. This can require significant computational time when scaling up for large networks.

The paper is organized as follows. Chapter 2 describes the details of the proposed model. Finally, Chapter 3 provides a summary of the study and future prospects.

2. Model

Notations

The notations below are adopted in this study to express traffic flow characteristics in a road network. Random variables are expressed in capital letters, and lowercase letters are used to denote the expected values of the corresponding random variables.

A	Set of links in network $A = A_{wo} \cup A_w$
A_{wo}	Set of links without toll in network
A_w	Set of links with toll in network
I	Set of OD pairs
J	Set of paths

Q	Total traffic demand
Q_i	Traffic demand between OD pair i
T	Link delay time
C_a	Traffic capacity on link a
p_i	Ratio of traffic demand between OD pair i to the total traffic demand
p_{ij}	Probability of selecting path j for OD pair i
Ξ_{ij}	Travel time on path j between OD pair i
F_{ij}	Traffic flow on path j between OD pair i
l_a	Length of link a
P_a	Toll of link a (JPY/KM)
τ	Value of time
V_a	Traffic flow on link a
t_a^0	Free flow travel time on link a
α, β	Parameters of BPR function
c_{ij}	Generalized path travel cost on path j between OD pair i
μ_i	Minimum path travel cost between OD pair i
$\boldsymbol{F}, \boldsymbol{\mu}, \boldsymbol{\Xi}, \boldsymbol{Q}, \boldsymbol{C}$	Vectors of path traffic flows, minimum path travel costs, path travel times, OD traffic demands and generalized travel costs
$\boldsymbol{f}, \boldsymbol{c}, \boldsymbol{q}$	Mean vectors of path traffic flows, generalized travel costs and OD traffic demands

(1) Traffic flow

Consider a transportation network $G(N, A)$ with a set of nodes N and a set of links A, together with a set of OD pairs I. The total demand Q following a log-normal distribution is

$$Q \sim LN(\mu_Q, \sigma_Q^2) \quad (1)$$

where μ_Q and σ_Q^2 are the parameters of the log-normal distribution. The traffic demand Q_i for OD pair i is expressed as

$$Q_i = p_i \cdot Q \ \forall i \in I \quad (2)$$

here the ratio p_i for OD pair i holds $\sum_{i \in I} p_i = 1.0$. OD traffic demand Q_i follows the following log-normal distribution.

$$Q_i \sim LN(\mu_{Q_i}, \sigma_{Q_i}^2) \ \forall i \in I \quad (3)$$

Its parameters are given as follows.

$$\mu_{Q_i} = \mu + \ln(p_i) \ \forall i \in I \quad (4)$$

$$\sigma_{Q_i}^2 = \sigma^2 \ \forall i \in I \quad (5)$$

The vector of mean OD traffic demand is represented as follows:

$$\boldsymbol{\mu}_Q = (\mu_{Q_1}, \dots, \mu_{Q_i}, \dots, \mu_{Q_{|i|}}) \forall i \in I \quad (6)$$

In this study, the variation in OD traffic demand is considered constant.

$$\sigma_{Q_i}^2 = costant \quad (7)$$

The path flow F_{ij} is calculated as the product of the OD demand Q_i and p_{ij} which is the probability of selecting route j between OD pair i.

$$F_{ij} = p_{ij} \cdot Q_i = p_{ij} \cdot p_i \cdot Q \quad \forall i \in I, \forall j \in J \quad (8)$$

The path flow follows the following log-normal distribution.

$$F_{ij} \sim LN\left(\mu_{F_{ij}}, \sigma_{F_{ij}}^2\right) \ \forall i \in I, \forall j \in J \quad (9)$$

where

$$\mu_{F_{ij}} = \mu + \ln(p_{ij} \cdot p_i) \ \forall i \in I, \forall j \in J \quad (10)$$

$$\sigma_{F_{ij}}^2 = \sigma^2 \ \forall i \in I, \forall j \in J \quad (11)$$

The link flow is expressed as the sum of path flow that pass through the link.

$$V_a = \sum_{i \in I} \sum_{j \in J} F_{ij} \cdot \delta_{aj} = \sum_{i \in I} \sum_{j \in J} p_{ij} \cdot p_i \cdot Q \cdot \delta_{aj} \quad \forall a \in A, \forall i \in I, \forall j \in J \quad (12)$$

Its parameters are given as follows.

$$V_a \sim LN(\mu_{V_a}, \sigma_{V_a}^2) \ \forall a \in A \quad (13)$$

$$\mu_{V_a} = \mu + \ln(p_{ij} \cdot p_i) \ \forall a \in A \quad (14)$$

$$\sigma_{V_a}^2 = \sigma^2 \ \forall a \in A \quad (15)$$

(2) Travel Time

In this study, the link travel time is defined by the BPR function. Using the previously defined link flow and the link capacities $C_a \sim LN(\mu_{C_a}, \sigma_{C_a}^2)$ $(\forall a \in A)$, the link travel time follows the following shifted log-normal distribution.

$$t_a(V_a, C_a) = t_a^0 \cdot \left(1 + \alpha \left(\frac{V_a}{C_a}\right)^{\beta}\right) \quad \forall a \in A \quad (16)$$

For the set of links without toll, link travel time is expressed as follows.

$$t_a(V_a, C_a) = t_a^0 + \frac{t_a^0 \cdot \alpha_a}{C^{\beta_a}} \cdot (V_a)^{\beta_a} = t_a^0 + \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a} \quad \forall a \in A_{wo} \quad (17)$$

$$\gamma_a = t_a^0 \cdot \alpha_a \ \forall a \in A_{wo} \quad (18)$$

Considering the links that need to be charged like expressway, the expressway toll can be calculated as the part of travel time as shown by (17).

$$t_a(V_a, C_a) + \frac{l_a \cdot P_a}{\tau} = t_a^0 + \frac{l_a \cdot P_a}{\tau} + \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a}$$

$$\forall a \in A_w \quad (19)$$

The link travel time shown above can be divided into a deterministic term (free flow travel time) and a random term (link delay time).

The random term is affected by changes in link flow and link capacity. Since the link flow and link capacity follow mutually independent log-normal distributions, the link delay time also follows a log-normal distribution. If the link flow is used as the traffic state to be estimated simultaneously with the travel time, it is appropriate to use the link delay time as the traffic state to be estimated in order to facilitate the calculation of the likelihood. The link delay time follows the following log-normal distribution

$$T_a' = \gamma_a \cdot \left(\frac{V_a}{C_a}\right)^{\beta_a} \sim LN\left(\mu_{T_a'}, \sigma_{T_a'}^2\right) \ \forall a \in A \quad (20)$$

where

$$\mu_{T_a'} = \ln(\gamma_a) + \beta_a \cdot \ln\left(p_{ij} \cdot p_i\right) + \beta_a \cdot \left(\mu - \mu_{C_a}\right)$$

$$\forall a \in A, \forall\, i \in I, \forall j \in J \quad (21)$$

$$\sigma_{T_a'}^2 = (\beta_a)^2 \cdot \left(\sigma^2 + \sigma_{C_a}^2\right) \ \forall a \in A \quad (22)$$

The link travel time follows a shifted log-normal distribution. Thereby, the link travel time is

$$T_a = t_a^0 + T_a' \ \ \forall a \in A \quad (23)$$

Then, the mean and variance of link travel time are respectively given by

$$E[T_a] = t_a^0 + E(T_a')$$

$$= t_a^0 + \exp\left(\mu_{T_a'} + \frac{1}{2}\sigma_{T_a'}^2\right) \forall a \in A \quad (24)$$

$$var[T_a] = var[T_a']$$

$$= \exp\left(2\mu_{T_a'} + \sigma_{T_a'}^2\right)\left(\exp\left(\sigma_{T_a'}^2\right) - 1\right)$$

$$\forall a \in A \quad (25)$$

The path travel time is expressed as the sum of the link travel times as follows

$$\Xi_{ij} = \sum_{a \in A} T_a \cdot \delta_{aj}$$

$$\forall i \in I, \forall j \in J \quad (26)$$

where δ_{aj} is the variable that equals 1 if path j uses link and 0 otherwise. Here the mean and variance of path travel time are calculated as follows,

$$E\left[\Xi_{ij}\right] = \sum_{a \in A} E[T_a] \cdot \delta_{aj}$$

$$\forall i \in I, \forall j \in J \quad (27)$$

$$var\left[\Xi_{ij}\right] = \sum_{a \in A} var[T_a] \cdot \delta_{aj}$$

$$\forall i \in I, \forall j \in J \quad (28)$$

The vector of path travel times is expressed as follows.

$$\Xi = \left(\Xi_{11}, \dots, \Xi_{ij}, \dots, \Xi_{|i||j|}\right)^T \ \forall i \in I, \forall j \in J \quad (29)$$

(3) User Equilibrium Assignment Model

The vector of generalized path travel costs $\boldsymbol{C}$ is expressed by

$$\boldsymbol{C} = E[\Xi] + \theta \cdot var[\Xi] \ \forall i \in I, \forall j \in J \quad (30)$$

where θ is a parameter used to indicate the risk-averse route choice behavior of drivers. The mean vector of generalized path travel costs is

$$\boldsymbol{c} = \boldsymbol{E}[\boldsymbol{C}] = \left(c_{11}, \dots, c_{ij}, \dots, c_{|i||j|}\right)^T$$

$$\forall i \in I, \forall j \in J \quad (31)$$

The vectors of path traffic flows, OD traffic demands, and the mean vectors of path traffic flows and OD traffic demands are respectively represented by

$$\boldsymbol{F} = \left(F_{11}, \dots, F_{ij}, \dots, F_{|i||j|}\right)^T \ \forall i \in I, \forall j \in J \quad (32)$$

$$\boldsymbol{f} = \boldsymbol{E}[\boldsymbol{F}] = \left(f_{11}, \dots, f_{ij}, \dots, f_{|i||j|}\right)^T$$

$$\forall i \in I, \forall j \in J \quad (33)$$

$$\boldsymbol{Q} = \left(Q_1, \dots, Q_i, \dots, Q_{|i|}\right)^T \ \forall i \in I, \forall j \in J \quad (34)$$

$$\boldsymbol{q} = E[\boldsymbol{Q}] = \left(q_1, \dots, q_i, \dots, q_{|i|}\right)^T$$

$$\forall i \in I, \forall j \in J \quad (35)$$

When mean vector of path flows is $\boldsymbol{f} \geq 0$, minimum path travel cost between OD pair i is expressed as follows.

$$\mu_i = min\{c_{ij}, \forall i \in I, \forall j \in J\} \quad (36)$$

Here, the vector of minimum path travel costs is shown blow.

$$\boldsymbol{\mu} = \left(\mu_1, \dots, \mu_i, \dots, \mu_{|i|}\right)^T \forall i \in I \quad (37)$$

For any path $j \in J$ between OD pair $i \in I$, the mean vector of path flows $\boldsymbol{f}$ is in Wardropian equilibrium (User Equilibrium) shown below

$$f_{ij} > 0 \rightarrow c_{ij} = \mu_i \ \forall i \in I, \forall j \in J \quad (38)$$

$$f_{ij} = 0 \rightarrow c_{ij} \geq \mu_i \ \forall i \in I, \forall j \in J \quad (39)$$

The equilibrium condition for User Equilibrium (UE) assignment for a given OD pair is that the generalized costs of the used routes are all equal and are smaller than or equal to the costs of the unused routes.

Given this equilibrium condition, the user equilibrium condition can be expressed as the following variational inequality problem:

$$\sum_{i \in I} \sum_{j \in J} c_{ij}(f_{ij} - f_{ij}{}^*) \geq 0 \quad (40)$$

$$\sum_{i \in I} \sum_{j \in J} f_{ij} = q_i \quad (41)$$

$$\boldsymbol{f} \geq 0 \ \forall i \in I, \forall j \in J \quad (42)$$

In addition, (40) can be expressed as follows.

$$\boldsymbol{c} \cdot (\boldsymbol{f} - \boldsymbol{f}^*) \geq 0 \quad (43)$$

(4) Maximum Likelihood Estimation Model

Vector of link delay time $\boldsymbol{T}' = \left(T_1', \dots, T_a', \dots, T_{|A|}'\right)$ $(\forall a \in A)$ follows the following multivariate log-normal distribution

$$\boldsymbol{T}' \sim MVLN\left(\boldsymbol{\mu}_{T_a'}, \boldsymbol{\Sigma}_{T_a'}\right) \quad (44)$$

where

$$\boldsymbol{\mu}_{T_a'} = \left(\mu_{T_1'} \ \mu_{T_2'} \dots \mu_{T_{|A|}'}\right)^T \quad (45)$$

$$\boldsymbol{\Sigma}_{T_a'} = \begin{pmatrix} \sigma_{T_1',T_1'} & \cdots & \sigma_{T_1',T_{|A|}'} \\ \vdots & \ddots & \vdots \\ \sigma_{T_{|A|}',T_1'} & \cdots & \sigma_{T_{|A|}',T_{|A|}'} \end{pmatrix} \quad (46)$$

A method for estimating traffic conditions by maximum likelihood estimation model is based on the method proposed in Tani and Uchida (2020).

Given G days of traffic observation data, define the observed state vector $\boldsymbol{m}_g$ and the observed value vector $\widehat{\boldsymbol{d}}_g$ for the observation on day $g \in \{1, \dots, G\}$.

$\widehat{\boldsymbol{d}}_g$ is a vector representing the traffic conditions observed across the road network on day $g \in \{1, \dots, G\}$, defined as

$$\widehat{\boldsymbol{d}}_g = \left(\hat{t}_1 \dots \ \hat{t}_a \dots \hat{t}_{|A|} \right) \quad (47)$$

where $\hat{t}_a$ is the logarithm of the observed delay time of link a. For instance, the link delays observed on day $g \in \{1, \dots, G\}$ in a simple network which is comprised of three links are 10 min at link1 and 5min at link 2, then the $\widehat{\boldsymbol{d}}_g$ is expressed as follows.

$$\widehat{\boldsymbol{d}}_g = (\ln(10) \ \ \ln(5) \ -\infty) \quad (48)$$

If there is no observed value or observed value is less than or equal to zero, the corresponding value is assumed to be $-\infty$. Due to the fact that the link delay time is defined as the increase in travel time relative to free flow travel time. The link delay time cannot be defined if the link travel time is less than or equal to the free flow travel time.

Observed state vector $\boldsymbol{m}_g$ shown by (49) reflects the traffic state on links, and its entry $m_{a,g}$ equals 1 if the link delay time is observed on link a and 0 otherwise.

$$\boldsymbol{m}_g = \left[m_{a,g}, \forall a \in A\right] \quad (49)$$

where

$$m_{a,g} = \begin{cases} 1, & if \ \hat{t}_a \neq -\infty \\ 0, & Otherwise \end{cases} \quad (50)$$

The $\boldsymbol{m}_g$ corresponding to (48) is:

$$\boldsymbol{m}_g = (1\ 1\ 0) \quad (51)$$

$\boldsymbol{M}_g$ is defined as the diagonal matrix of $\boldsymbol{m}_g$ from which each row containing only zero entries is removed. This indicates that only data from observed links are used to calculate the likelihood. For example, $\boldsymbol{m}_g = (1\ 1\ 0)$, as shown in (51), $\boldsymbol{M}_g$ can be expressed as:

$$\boldsymbol{M}_g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (52)$$

If the number of data observed on day $g \in \{1, \dots, G\}$ is $n(g)$, the likelihood function for estimating the network traffic conditions is defined as:

$$L\left(\boldsymbol{\mu}_Q \middle| \widehat{\boldsymbol{d}}_g, \boldsymbol{M}_g \ \forall g \in G, \sigma_Q\right) = \prod_{g \in G} \frac{\exp\left(-\frac{1}{2}(\widehat{\boldsymbol{d}}_g - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\widehat{\boldsymbol{d}}_g - \boldsymbol{\mu}_g)\right)}{2\pi^{\frac{n(g)}{2}} [\boldsymbol{\Sigma}_g]^{\frac{1}{2}}} \quad (53)$$

where

$$\boldsymbol{\Sigma}_g = \boldsymbol{M}_g \boldsymbol{\Sigma}_{T_a'} \boldsymbol{M}_g^T \quad (54)$$

$$\boldsymbol{\mu}_g = \boldsymbol{M}_g \boldsymbol{\mu}_{T_a'} \quad (55)$$

$$\widehat{\boldsymbol{d}}_g = \boldsymbol{M}_g \widehat{\boldsymbol{d}}_g^T \quad (56)$$

3. Conclusions and Future Tasks

For large road networks, existence of unobserved links in large road networks makes it difficult to estimate the traffic condition of the entire network. Under the assumption that some links have no observation data, a method for estimating traffic conditions using ETC2.0 data is developed in this study, and also providing a theoretical foundation for future applications to estimate actual large road networks.

In the future, numerical computations will be performed on the large road network in Asahikawa, Hokkaido, in order to validate the model provided in this study.

References

1) 原祐輔, 花岡洋平, 桑原雅夫 : 道路ネットワーク内の関係性に着目した長期観測プローブデータによるプローブ未観測リンクの交通状態補間, 交通工学論文集, Vol. 2, No. 1, pp. 1-10.
2) 峪 龍一, 内田 賢悦 : プローブデータと感知器データを組み合わせた交通状態の時空間推定, 第 62 回土木計画学研究発表会 , 2020.
3) 川村 雄斗, 峪 龍一, 内田 賢悦 : 不完全交通データを活用した道路ネットワーク交通状態のリアルタイム更新, 第63回土木計画学研究発表会, 2021.
4) Hara, Y., Suzuki, J. and Kuwahara, M.: Network-wide traffic state estimation using a mixture Gaussian graphical model and graphical lasso. Transportation Research Part C: Emerging Technologies, Vol. 86, pp. 622-638, 2018.