A domain-based 3-D route choice modeling based on sparse observations through Wi-Fi

Kanako Izawa[1], Eiji Hato[2]

Department of Civil Engineering, The University of Tokyo, Japan
[1]E-mail: izawa@bin.t.u-tokyo.ac.jp
[2]E-mail: hato@bin.t.u-tokyo.ac.jp

Keywords

Route choice model; Space choice model; Implicit network; Domain of data relevance(DDR); Spatial incidence matrices; Sparse data; Wi-Fi data; GPS data

1 Introduction

In order to investigate whether policies are effective for urban planning, it is necessary to have a methodology that can analyze detailed behavior of pedestrians in cities. Although the probe type observation method using GPS and mobile communication equipment has developed (Hato, 2010), it is difficult to accurately analyze such detailed behavior since GPS has problems with indoor space observation and population representation.

In recent years, due to the spread of information sensors with advanced functions, detailed observation on a three-dimensional urban network is now becoming possible, not on a two-dimensional network that has been analyzed so far.

The biggest challenge of the route choice model lies in the generation of route choice sets. It is difficult to explicitly enumerate all possible paths on a huge network of real space, especially in a three dimensional network. Because of these problems, route choice models that implicitly estimate parameters without enumerating choices have been proposed by assuming that future state selections are performed sequentially in each state(Dial, 1971; Akamatsu, 1996; Oyama et al., 2016). Recursive Logit model (RL) proposed by Fosgerau et al. (2013) describes behavior consistent with the path-based route choice model by link-to-link transition probability applying Markov chain allocation(Akamatsu, 1996).

True route alternatives are required for parameter estimation of the route choice model. Routes are generally described as a sequence of links on a network by a map matching algorithm. Although various map matching methods(Bierlaire et al., 2013; Chen et al., 2015) have been proposed, explicitly associating behavior data with different observation errors with links on the network will bias estimated parameters. In particular, when the observation data is sparse, the method specifying the route based on observation data(Akgun et al., 2000; Prato et al., 2006) has not yet been established. Since the observation of the Wi-Fi data is probabilistic, it is necessary to consider such cases explicitly by handling individual scheduling(Spiess and Florian, 1989 ; Bell, 2009; Ma and Fukuda, 2015) , psychological factor (Kazagli et al., 2016) and observation conditions stochastically.

Although there is a method to model behavior by taking behavior as a movement of fluid particles on a two-dimensional plane (Edie, 1963; Nikolic and Bierlaire, 2016) instead of movement on the network, it is not suitable for policy evaluation.

Bierlaire et al.(2008) defined the spatial domain where behavior data could exist as DDR (Domain of Data Relevance) and proposed a route choice model which treats actual routes implicitly and estimates the parameters taking clearly the choice probability and the acquisition probability of the position information.

There is no frame of the route choice model based on three-dimensional behavioral data using Wi-Fi methods. We aim to propose a framework of 3-D route choice model based on data collected through different observation methods.

2 Formulation based on Lagrangian and Eulerian approaches

There are two types of pedestrian behavior analysis which are the Lagrangian approach and the Eulerian approach. The Lagrangian approach is a method to describe behavior in a form that tracks temporal changes of different position coordinates for each pedestrian, and most of the analysis of probe data(Nikolic and Bierlaire, 2016; Danalet et al., 2014) belongs to this. On the other hand, the Euler type approach is a method to describe the movement of a pedestrian observed at each fixed point at each time by fixing objects of interest to certain position coordinates.

In order to analyze the path of each pedestrian with the former approach, it is necessary to observe

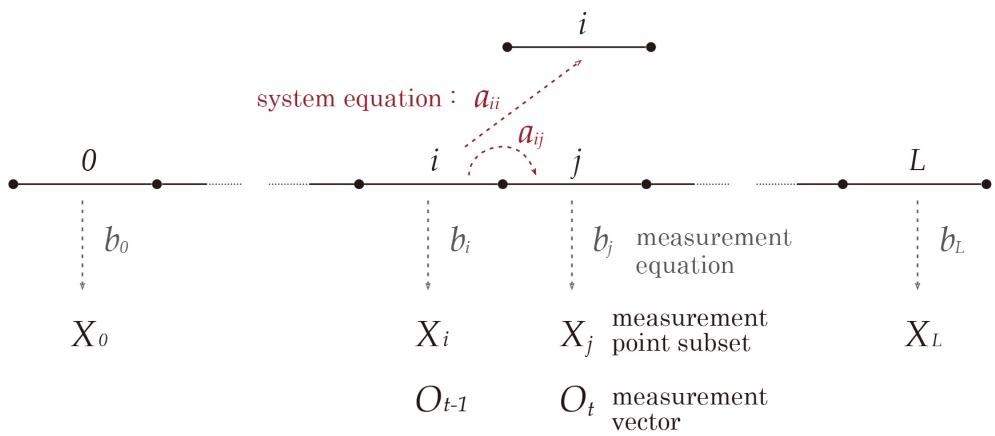

Figure 1: The concept of Lagrangian route choice model

continuous position data. With the latter approach, it is necessary to observe the position data of each pedestrian at all fixed points.

It is possible to represent the inter-link transition probability $\Gamma_t(i,j)$ for each link at time t by giving the observation probability $\boldsymbol{b}$ at each Wireless LAN access point (AP) using Equation(1).

$$\Gamma_t(i,j) = \frac{\alpha_t(i)a_{ij}b_i(X_t)\beta_{t+1}(j)}{\alpha_T(L)} \tag{1}$$

It is important that all observation points are always obtained from known observation point sets $\boldsymbol{X}$. The parameters can be estimated based on algorithm in Figure 2. We aim to incorporate the concept of DDR (Bierlaire and Frejinger, 2008) into estimation of parameters in a spatial choice model without uniquely identifying the existence space of the data. We assume that the space on the network is decomposed as shown as Figure 3 and membership constraints are given to each element. We can define incidence matrices $\boldsymbol{J}$, $\boldsymbol{K}$, $\boldsymbol{L}$, $\boldsymbol{M}$, $\boldsymbol{N}$. The likelihood function can be formulated as follows.

$$P_{nt}(i|F_i) = \prod_{\tau}^{t} P_{n\tau}(i_\tau|F_i) \tag{2}$$

$$P_{n\tau}(i_\tau|F_i) = \sum_{f\in F_i}\sum_{b\in B_i}\sum_{h\in H_i}\sum_{s\in S_i} P_n(s|S_i)P_n(h|H_i)P_n(b|B_i)P_n(f|F_i)\sum_{p\in C_n(s)} P_n(p|C_n(f);\theta) \tag{3}$$

$P_n(f|F_i)$: The probability that the actual OD lots pair is f given that relevant OD lots pair sets are F_i

$P_n(b|B_i)$: The probability that the actual OD buildings pair is b given that relevant OD lots pair(building

Step0 : choose initial parameters

Step1 : preparation for estimation

Step1-1 : Forward Algorithm

calculate $\boldsymbol{\alpha_t(j) = \alpha_{t-1}(j)a_{j,j}b_j(O_{t-1}) + \alpha_{t-1}(j-1)a_{j-1,j}b_{j-1}(O_{t-1})}$

Step1-2 : Backward Algorithm

calculate $\boldsymbol{\beta_t(j) = \beta_{t+1}(j+1)a_{j,j+1}b_j(O_t) + \beta_{t+1}(j)a_{j,j}b_j(O_t)}$

Step2 : calculate

Repeat

Step2-1 : E-step

calculate $\boldsymbol{\Gamma_t(i,j) = \frac{\alpha_t(i)a_{i,j}b_i(O_t)\beta_{t+1}(j)}{\alpha_T(L)}}$

Step2-2 : M-step

recalculate $\boldsymbol{a_{i,j} = \frac{\sum_t \Gamma_t(i,j)}{\sum_t \sum_j \Gamma_t(i,j)}}$, $\boldsymbol{b_j = \frac{\sum_{t \in O} \sum_k \Gamma_t(j,k)}{\sum_t \sum_k \Gamma_t(j,k)}}$

Until : convergence reached according to the criterion

Return : optimal parameters

Figure 2: Lagrangian estimation

pair sets) is $f(= B_i)$.

$P_n(h|H_i)$: The probability that the actual OD floors pair is h given that relevant OD buildings pair(floor pair sets) is $b(= H_i)$.

$P_n(s|S_i)$: The probability that the actual OD pair is s given that relevant OD floor pair(OD pair sets) is $h(= S_i)$.

$P_n(i|p)$: Measurement model. The probability to measure i given that that actual route is p.

$P_n(p|C_n(s);\theta)$: Route choice model. The probability that individual n selects path p from path choice sets $C_n(s)$.

3 Case Study

3.1 Data characteristics

Although Wi-Fi data has three-dimensional position information, it is necessary to supplement uncertain factor with the behavior model since the time interval and the distance interval of each observation data are longer than GPS data. The position coordinates are always obtained on lots and can be obtained at a plural points at the same time. Furthermore, since the number of observation

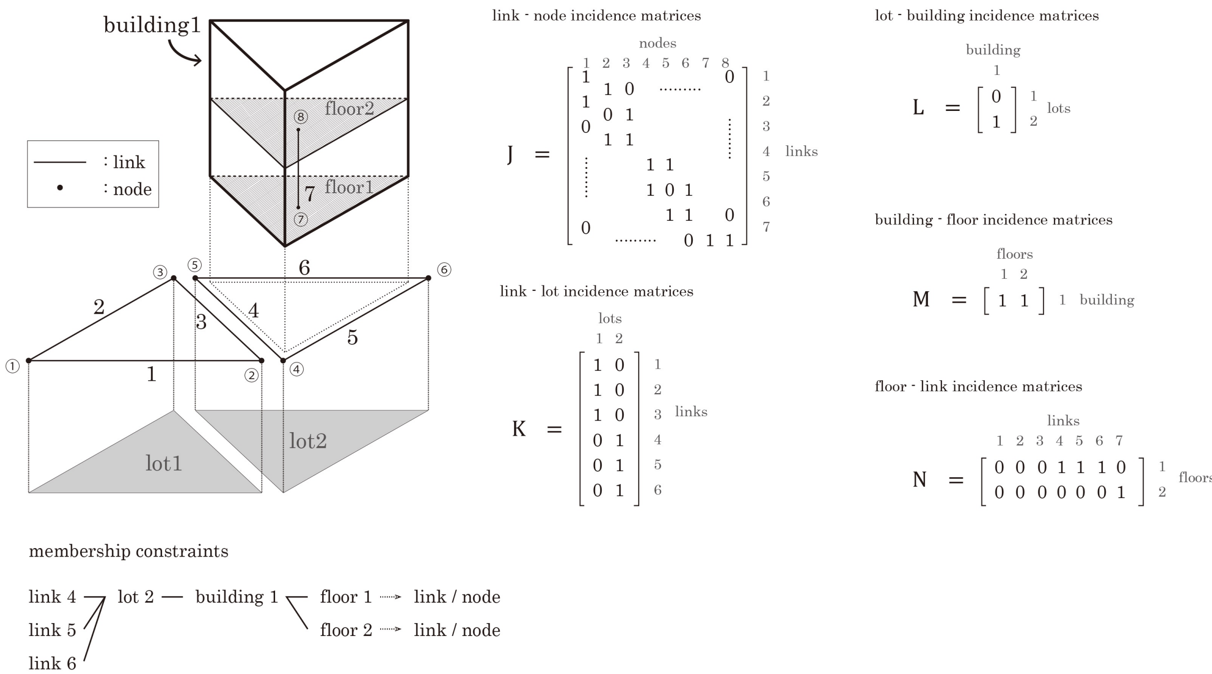

Figure 3: The concept of Eulerian model

Table 1: Comparison between characteristics of survey data

	Person Trip Survey	Probe Person Survey	Wi-Fi Survey
measurement	questionnaire	GPS data	Wireless LAN
unit	zone(500 – 10 km)	dot(coordinate)	dot(coordinate)
error	depend on memory, unanswered	hundreds m – dozens m	dozens cm – dozens m
personal attribute	acquirable	acquirable	unacquirable
sample size	big(a small-bias sampling)	small(sampling bias)	enormous(a small-bias sampling)
information	position, destination, transportation	positional coordinate, destination, transportation, route, acceleration, sojourn time	positional coordinate, route, sojourn time, acceleration
characteristics	wide area	2-D space, short sampling interval	3-D space, long sampling interval, simultaneously measured in plural parts

points is limited for one trip, the choice set tends to be enormous. Table 1 show a comparison between characteristics of survey data.

3.2 Estimation result

We estimate route choice model in the simple network (Figure 4) changing the probability of OD pair. We assume that there are many people who select the link between the two lots.

Here, the travel time of all the links are equal and only the travel time of link between floor 1 and floor 2 is set to 0. Now we assume that an observation of AP1 is obtained at time t_1, observations of AP2 and AP3 are obtained at time t_4 and observations are not obtained at time t_2 and t_3.

Table2 shows the result of numerical estimation. We can find that different values are estimated

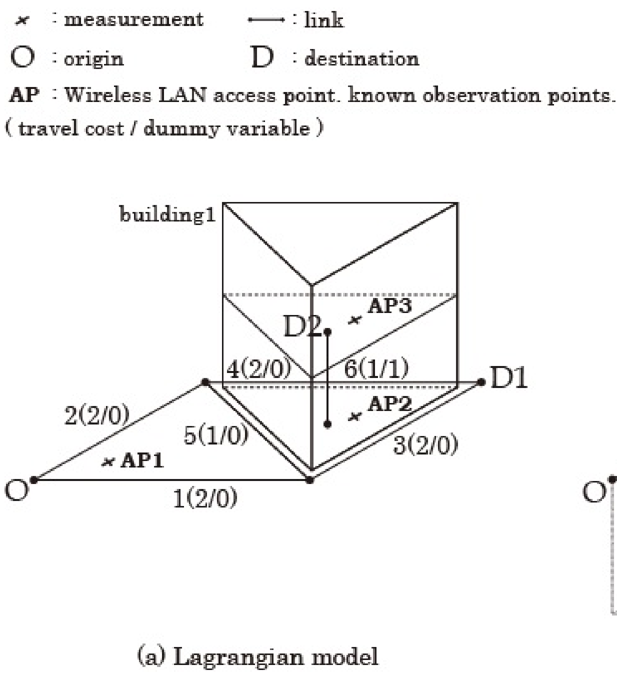

Figure 4: Illustrative example in 3-D network

between when the route is uniquely specified and when it is not. In addition, it is assumed that more accurate values can be obtained when supplementary information such as radio field intensity is obtained and the probability of OD pair is set.

4 Conclusion

In this paper we have proposed a framework of 3-D route choice modeling based on data with different observation methods and demonstrated that the parameters of the spatial choice model are estimated differently depending on whether or not to specify the existence space. This means that

Table 2: Estimation result of 3-D route choice model

	$(\kappa_1 = 1.0, \kappa_2 = 0)$	$(\kappa_1 = 0.75, \kappa_2 = 0.25)$	$(\kappa_1 = 0.5, \kappa_2 = 0.5)$	$(\kappa_1 = 0.25, \kappa_2 = 0.75)$	$(\kappa_1 = 0, \kappa_2 = 1.0)$
time	-0.03736	0.769	0.796	0.825	0.365
	(-5.540×10^{-2})	(1.123×10^{-4})	(6.397×10^{-2})	(6.630×10^{-2})	(2.070×10^{-2})
cost	0.369	0.140	0.139	0.139	0.716
	(2.007)	(2.149)	(7.899×10^{-3})	(1.291×10^{-2})	(4.069×10^{-2})
dummy	-1.099	-1.140	-1.205	-1.284	-1.387
	(-5.993)	(-5.849)	(-5.716)	(-5.612)	(-5.546)
sample	100	100	100	100	100
L0	-438.203	-438.203	-438.203	-438.203	-207.944
LL	-391.202	-228.006	-287.058	-228.006	-160.938
ρ^2	0.107	0.480	0.345	0.480	0.226
$\overline{\rho}^2$	0.100	0.473	0.338	0.473	0.212

(κ_1, κ_2) = (Probability that the OD pair is OD1, Probability that the OD pair is OD2)

(t-value)

it is necessary to take sufficient consideration as to what kind of modeling to apply depending on the purpose of urban planning. Establishing an analytical framework using sparce data is extremely effective in sensing design and network design in urban planning. It is also important that the latter model proposed in this research can be applied to both GPS data and Wi-Fi data. We will estimate using real data and consider what kind of description of behavior is appropriate for pedestrian behavior that we want to analyze.

References

1. Edie, L. C. (1963), ”Discussion of traffic stream measurements and definitions”. Port of New York Authority.

2. Dial, R. B. (1971), ”A probabilistic multipath traffic assignment model which obviates path enumeration”,*Transportation research*, Vol.5, No.2, pp.83-111.

3. Spiess, H., and Florian, M. (1989), ”Optimal strategies: a new assignment model for transit networks”, Transportation Research Part B: Methodological, Vol.23, No.2, pp.83-102.

4. Bell, M. G. (1995), ”Alternatives to Dial's logit assignment algorithm”, *Transportation Research Part B: Methodological*, Vol.29, No.4, pp.287-295.

5. Akamatsu, T. (1996), ”Cyclic flows, Markov process and stochastic traffic assignment”, *Transportation Research Part B: Methodological*, Vol.30, No.5, pp.369-386.

6. Akgun, V., Erkut, E., and Batta, R. (2000), ”On finding dissimilar paths. European Journal of Operational Research”, Vol.121, No.2, pp.232-246.

7. Prato, C., and Bekhor, S. (2006), ”Applying branch-and-bound technique to route choice set generation”, Transportation Research Record: Journal of the Transportation Research Board, (1985), pp.19-28.

8. Bierlaire, M., and Frejinger, E. (2008), ”Route choice modeling with network-free data”, *Transportation Research Part C: Emerging Technologies*, Vol.16, No.2, pp.187-198.

9. Frejinger, E., Bierlaire, M., and Ben-Akiva, M. (2009), ”Sampling of alternatives for route choice modeling”, *Transportation Research Part B: Methodological*, Vol.43, No.10, pp.984-994.

10. Bell, M. G. (2009), ”Hyperstar: A multi-path Astar algorithm for risk averse vehicle navigation”. Transportation Research Part B: Methodological, Vol.43, No.1, pp.97-107.

11. Hato, E. (2010), ”Development of behavioral context addressable loggers in the shell for travel-activity analysis”, Transportation Research Part C: Emerging Technologies, Vol.18, No.1, pp.55-67.

12. Bierlaire, M., Chen, J., and Newman, J. (2013), ”A probabilistic map matching method for smartphone GPS data”, *Transportation Research Part C: Emerging Technologies*, Vol.26, pp.78-98.

13. Fosgerau, M., Frejinger, E., and Karlstrom, A. (2013), ”A link based network route choice model with unrestricted choice set”, *Transportation Research Part B: Methodological*, Vol.56, pp.70-80.

14. Danalet, A., Farooq, B., and Bierlaire, M. (2014), ”A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures”, *Transportation Research Part C: Emerging Technologies*, Vol.44, pp.146-170.

15. Chen, J., and Bierlaire, M. (2015), ”Probabilistic multimodal map matching with rich smartphone data”, *Journal of Intelligent Transportation Systems*, Vol.19, No.2, pp.134-148.

16. Ma, J., Fukuda, D., (2015), ”A hyperpath-based network generalized extreme-value model for route choice under uncertainities”, Transportation Research Procedia, the 21st International Symposium on Transportation and Traffic Theory, Vol.7, pp.44-58.

17. Mai, T., Fosgerau, M., and Frejinger, E. (2015), ”A nested recursive logit model for route choice analysis”, *Transportation Research Part B: Methodological*, Vol.75, pp.100-112.

18. Nikolic, M., and Bierlaire, M. (2016), ”Data-driven characterization of pedestrian flows”, *In hEART 2016, 5th Symposium of the European Association for Research in Transportation*, No. EPFL-TALK-223712.

19. Oyama, Y., Chikamatsu, K., Shoji, Y., Hato, E. and Koga, M.(2016), ”Trajectory-oriented traffic management using sequential discount rate: a case study of the Great East Japan Earthquake”, paper presented at the 11th ITS European Congress.

20. Kazagli, E., Bierlaire, M. and Flotterod, G. (2016),”Revisiting the route choice problem: A modeling framework based on mental representations”, *Journal of Choice Modelling*, Vol.19, pp.1-23.