Application of eigenvector spatial filtering to travel destination choice model: A case study of municipality-size choice in Hokkaido Island, Japan

Wataru Nakanishi [a] and Hiromichi Yamaguchi [b]

a) nakanishi@plan.cv.titech.ac.jp, Department of Civil and Environmental Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 1528552 Japan, +81-3-5734-2575

b) hyamaguchi@se.kanazawa-u.ac.jp, Institute of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa 9201192 Japan, +81-76-264-6394

Keywords: destination choice modelling, spatial discrete choice model, eigenvector spatial filtering, high-resolution data

1. Introduction

Many researchers have focused on the destination choice in long-distance travel. Discrete choice modelling is a representative way to describe this problem (e.g., Kato et al., 2011; van Nostrand et al., 2013). Here, we assume that the "utility function" of visiting each destination can be constructed by attribute variables like attractiveness and travel cost. Nowadays the data of the number of traveller can be obtained regularly and frequently with large sample size and spatio-temporally high resolution. For example, "Mobile spatial statistics" (MSS) by NTT DOCOMO (Terada et al., 2013) provides the population of residential zone and staying place zone pairs at every one hour. Nevertheless, utilisation of such data for destination choice modelling is not fully developed yet.

As spatial resolution of data becomes higher, naïve application of multinomial logit model to such data may lead to serious estimation bias caused by spatial correlation. Destinations that are close to a destination with higher utility may also have higher utility compared to other alternative destinations. In addition, it is difficult to find enough attribute variables that can distinguish the difference of utility among contiguous destinations, as they tend to have similar values. Spatial discrete choice model (e.g., Smirnov, 2010) is a common way to deal with this problem, although almost all of them need heavy computational load. Then, Yoshida and Tsutsumi (2013) and Wang et al. (2013) proposed the method to mitigate this problem. They introduced eigenvector spatial filtering (ESF) method to multinomial logit model (ESF-based MNL) and applied it to land use/price prediction. However, to authors' knowledge, this method is only appeared in Shinha (2017) whose objective is also land use modelling. The advantages of ESF-based MNL are that a) parameter estimation is as easy as conventional MNL as the utility function is given by a form of linear combination, and b) the estimated result can be interpreted as missing explanatory variables and/or spatial map patterns.

In this paper, we apply ESF-based MNL to long-distance travel destination choice model and estimate parameters and spatial map patterns using MSS data. We show the improvement in model accuracy in terms of likelihood ratio and Akaike information criterion (AIC) by comparison with the result of conventional non-spatial MNL. In addition, the obtained spatial map patterns are interpreted.

2. Methodology

2.1. Data

The population of residential and staying place zone pairs in Japan at 1 P.M. every day are collected for two weeks in June 2017 by MSS data. We only use the data whose residential zone is “Tokyo Metropolis” and staying place zones are 184 municipalities in Hokkaido Island (Fig. 1), which is located around 1000 km north from Tokyo. We regard that the staying place zones are the destinations of people in Tokyo. We obtain the number of people $n_{d,i}$, who stay in each municipality i for each day d, and calculate the summation for two weeks (14 days) as $N_i = \sum_d n_{d,i}$, i=1,2,…,184. The destinations are regarded to be selected by random utility maximisation; the observed probability for selecting municipality i is given by $N_i / \sum_j N_j$, j=1,2,…,184.

As for explanatory variables, we use following data; population (source: 2015 national census), added value per capita in both total and accommodation industry (source: 2016 national economic census), travel time from Tokyo to each municipality’s office, and flight frequency on each route (calculated by using road network data and flight timetable). These are selected by referring to the previous literatures.

2.2. Spatial discrete choice model: ESF-based MNL

MNL is represented by utility function, $U_i = \boldsymbol{\beta}\mathbf{X}_i + \varepsilon'_i$. Here, U_i is the utility to choose the municipality i as the destination, $\mathbf{X}_i$ is a vector of attribute explanatory variables of i, and $\boldsymbol{\beta}$ is a parameter vector. ε'_i represents the error term that follows i.i.d. Gumbel distribution, although this assumption may not hold true because of the spatial correlation as mentioned before.

We solve this problem using ESF by adding new variables $\mathbf{E}_i$ to the utility function, which represents the spatial map pattern. $\mathbf{E}$ is calculated based on the theory of Moran’s I statistics, which is popular as representative of spatial dependence. ESF is the methodology to represent the spatial correlation by a set of orthogonal and uncorrelated vectors $\mathbf{E}$ calculated from Moran’s I and proximity matrix (e.g., Tiefelsdorf and Griffith, 2007). We expect that the term $\mathbf{E}$ excludes the spatial correlation within the error term and makes i.i.d. assumption hold true. We call this method as ESF-based MNL. Its utility function is given by $U_i = \boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\gamma}\mathbf{E}_i + \varepsilon_i$. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are parameter vectors and ε_i represents the error term that follows i.i.d. Gumbel distribution. For the municipality i=1,2,…,184, the selection probability of each municipality $P(i)$ is given by

$$P(i) = \frac{\exp(\boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\gamma}\mathbf{E}_i)}{\sum_j \exp(\boldsymbol{\beta}\mathbf{X}_j + \boldsymbol{\gamma}\mathbf{E}_j)}, \qquad j=1,2,\ldots,184$$

The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are estimated by the maximum likelihood method as the conventional MNL (Train, 2003). The log-likelihood function ln$L$ is given by

$$\begin{aligned}\ln L &= \sum_i \ln\left(P(i)^{N_i}\right) = \sum_i N_i \ln P(i) \\ &= \sum_i N_i \left((\boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\gamma}\mathbf{E}_i) - \ln \sum_j \exp(\boldsymbol{\beta}\mathbf{X}_j + \boldsymbol{\gamma}\mathbf{E}_j) \right)\end{aligned}, \quad i,j=1,2,\ldots,184$$

3. Results

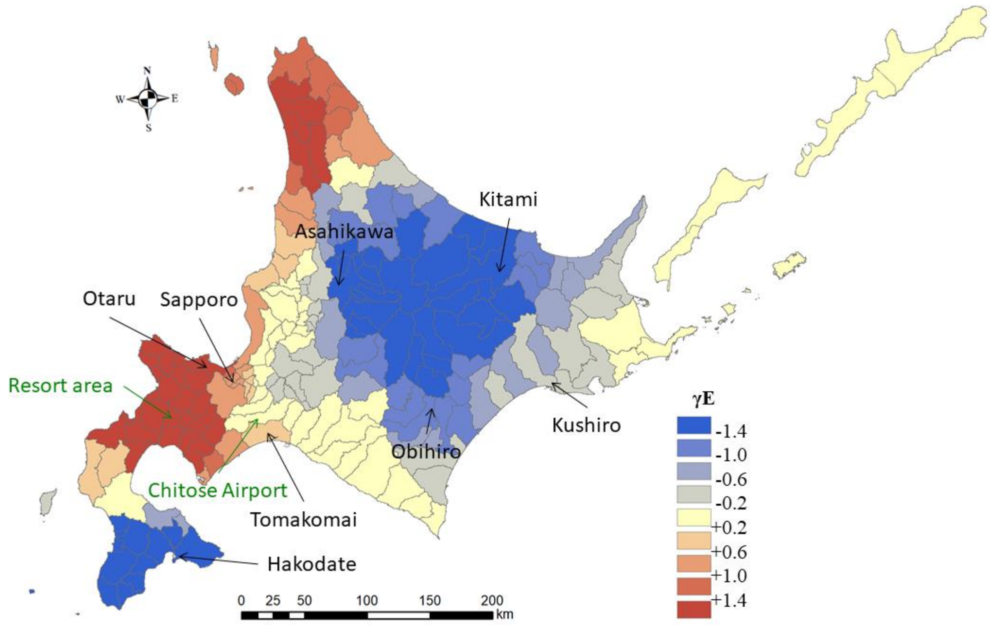
In addition to the explanatory variables explained in *2.1.*, we extract seven vectors as **E** for ESF-based MNL (detail in this process will be explained in the conference). The estimated results of both ESF-based MNL and conventional non-spatial MNL are shown in Table 1. First of all, all parameters are well obtained in terms of *t*-statistic and sign condition. Moreover, we confirm the improvement in both log-likelihood ratio and AIC in ESF-based MNL. This implies that we should consider the spatial correlation in destination choice in high resolution and ESF-based MNL is the useful method to deal with this problem. Actually, the coefficients of travel time and flight frequency are greatly different between non-spatial and spatial models. This result might show the estimation bias in the non-spatial model, which should be further examined.

Fig. 1 shows the estimated spatial map pattern, namely the value of $\boldsymbol{\gamma}\mathbf{E}$. The meaning of these values are the difference in utility to choose each destination if the values $\boldsymbol{\beta}\mathbf{X}$ are the same. Thus, we may interpret this map pattern as a) the resort area near Sapporo city is more likely to be visited compared to other areas, and b) eastern and southern Hokkaido area are less likely to be visited despite their attractiveness and accessibility.

Table 1. Estimated parameters

Explanatory variables		Spatial: ESF-based MNL		Non-spatial: Conventional MNL	
		Estimate	*t*-statistic	Estimate	*t*-statistic
Population [mil.]	β_1	8.95	113.27 **	8.84	185.40 **
Added value in total per capita [mil. JPY]	β_2	0.22	68.98 **	0.31	103.84 **
Added value in accommodation industry per capita [mil. JPY]	β_3	1.65	43.42 **	0.66	22.65 **
Travel time [hour]	β_4	-3.30	-461.52 **	-1.78	-385.20 **
Flight frequency (1 / # flights per day)	β_5	-2.94	-30.37 **	-12.33	-128.29 **
Additional vectors of spatial correlation	γ_1	-10.63	-89.00 **		
	γ_2	-8.82	-59.30 **		
	γ_3	8.80	74.71 **		
	γ_5	8.62	83.92 **		
	γ_6	-4.90	-25.77 **		
	γ_7	9.00	46.11 **		
	γ_8	5.18	58.49 **		
Log likelihood ratio			0.392		0.366
AIC			272,453.0		283,971.2
# samples					42,972

** represents 1% significant.



The values of **γE** and representative cities are drawn in the map. This can be interpreted as spatial map pattern: difference in utility if the values **βX** are the same.

Figure 1. Estimated spatial map patterns

4. Conclusion

We proposed the method of long-distance travel destination choice model for spatially high-resolution data. We introduced ESF-based MNL and showed the improvements in the estimated result. The main contributions of this paper are that a) the increase in estimation reliability in terms of spatial correlation increases the usefulness of analyses with spatially high-resolution data, and b) we can further utilise the proposed model for quantitative analyses of the spatial effect, e.g., how the increase in attractiveness of a certain zone affects in the neighbourhood areas. To that end, we will discuss on a) better selection of explanatory variables to improve the model accuracy, b) results on application to other dataset, e.g., other origins and long period data, and c) consideration of day of departure (weekday/end). In addition, as MSS data do not provide trip purpose, we should integrate other data to discuss the difference among them as a future work.

Acknowledgements

This work was supported by JSPS Kakenhi 17H01297 and 17K14736, and MEXT LEADER project.

References

Kato, H., Endo, K., Kurita, Z., Kaneko, Y., Kato K., Shimizu, T. and Tanabe, K.: Inter-Regional Travel Demand Analysis Using Integrated Model For Practical Travel Demand Forecast, *Journal of the Eastern Asia Society for Transportation Studies*, Vol.9, pp.239-254, 2011.

van Nostrand, C., Sivaraman, V. and Pinjari, A. R.: Analysis of long-distance vacation travel demand in the United States: a multiple discrete–continuous choice framework, *Transportation*, Vol.40, No.1, pp.151-171, 2013.

Shinha, P.: Modeling Land Use Change Using an Eigenvector Spatial Filtering Model Specification for Discrete Responses, In: Griffith D., Chun Y., Dean D. (eds.) *Advances in Geocomputation*, Springer, 2017.

Smirnov, O. A.: Modeling spatial discrete choice, *Regional Science and Urban Economics*, Vol. 40, No. 5, pp. 292–298, 2010.

Terada, M., Nagata, T., and Kobayashi, M.: Population Estimation technology for Mobile Spatial Statistics. *NTT DOCOMO Technical Journal*, Vol. 14, No.3, 2013, pp.10-15.

Tiefelsdorf, M. and Griffith, D. A.: Semiparametric filtering of spatial autocorrelation: the eigenvector approach, *Environment and Planning A*, Vol.39, No.5, pp.1193-1221, 2007.

Train, K.: *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, 2003.

Wang, Y., Kockelman, K. M. and Wang, X.: Understanding spatial filtering for analysis of land use-transport data, *Journal of Transport Geography*, Vol.31, pp.123-131, 2013.

Yoshida, T. and Tsutsumi, M.: Application of eigenvector-based spatial filtering approach to a multinomial logit model for land use data, *Proceedings of the Seventh World Conference of the Spatial Econometrics Association*, 2013.