🏆 reviewed paper

Analyses on Stopping-off Points and Destination Selection of Nagoya CBD Visitors by applying Spatial-Temporal Position Data

Yoshimasa Matsumoto, Toshiyuki Kaneda

(Yoshimasa Matsumoto, Nagoya Institute of Technology, Gokiso, Showa, Nagoya, Aichi, Japan, cls15084@ict.nitech.ac.jp)
(Prof. Toshiyuki Kaneda, Nagoya Institute of Technology, Gokiso, Showa, Nagoya, Aichi, Japan, kaneda@nitech.ac.jp)

1 ABSTRACT

In recent years, GPS(Global Positioning System) data and various other data have been provided as information on people's shopping-around behaviors within CBD(Central Business District), allowing people's shopping-around behaviors to be obtained in more detail. In addition to conventional person-trip surveys, these data have the potential to reveal new aspects of people's shopping-around behaviors. Furthermore, through the restaurant API(Application Programming Interface), by combining spatial-temporal position data with restaurant data, it will be possible to clarify people's district visitation factors and downtown selection behavior.

In this study, We defined a 5-minute stop in the area as a stopping off points in the spatial-temporal position data, and aimed to clarify the factors of stopping off points in the Nagoya CBD area by latent factor analysis of PLS(Partial Least Squares) regression, and to estimate individual destination selection behavior by multinomial logistic analysis. The results of the latent factor analysis of the PLS regression revealed that the regression coefficients were high for the number and type of restaurants in the 0-12 time period, but high for entertainment facilities in the 12-18 and 18-24 time periods. Furthermore, multinomial logistic analysis revealed a trade-off between proximity of 3 meters and the utility obtained from one restaurant.

Keywords: Shopping-around behaviour, Stopping-off point, Spatial-Temporal Position Data, Urban analysis, Big data

2 INTRODUCTION, PROBLEM AND OBJECTIVES

In recent years, GPS data and various other data have been provided as information on people's behavior within urban centers, allowing people's behavior to be obtained in greater detail. In addition to conventional person-trip surveys, these data have the potential to reveal new aspects of people's behavior. Furthermore, by combining restaurant API data, it will be possible to clarify people's district visitation factors and downtown selection behavior.

Therefore, in this study, the stopping off point is defined as a 5-minute stay in the area in the spatial-temporal location data, and the purpose of this study is to clarify the factors of the stopping off point in the Nagoya CBD by latent factor analysis of PLS regression and to estimate the individual's district selection behavior by multinomial logit analysis.

3 SETTING AREA AND DATA SOURCE

The target area was the central Nagoya area centered on Nagoya Station and Sakae Station, and was divided into 55 districts according to townships (Fig. 1). Table 1 shows the spatial-temporal position data by sample, restaurant API data, commercial concentration statistics data, and Nagoya City Building Survey data used in this study.

The spatial-temporal position data by sample is based on the PT(Person Trip) survey conducted on October 3, 2011 (weekday) as part of the respondents. The Chukyo area data provided includes the minute-by-minute latitude and longitude coordinates of the trips of 208,543 sample members. The restaurant API data contains restaurant information registered on Gurunavi, a restaurant information service.

Symbol	Data name	Source	Number of data	Year
A	spatial-temporal position data	Tokyo University	208,543	2011
B	restaurant API data	Gurunavi	13,144	2019
C	commercial concentration statistics data	Tokyo University	611,839	2014
D	Nagoya City Building Survey data	Nagoya City	32,925	2011

Table 1: Data source

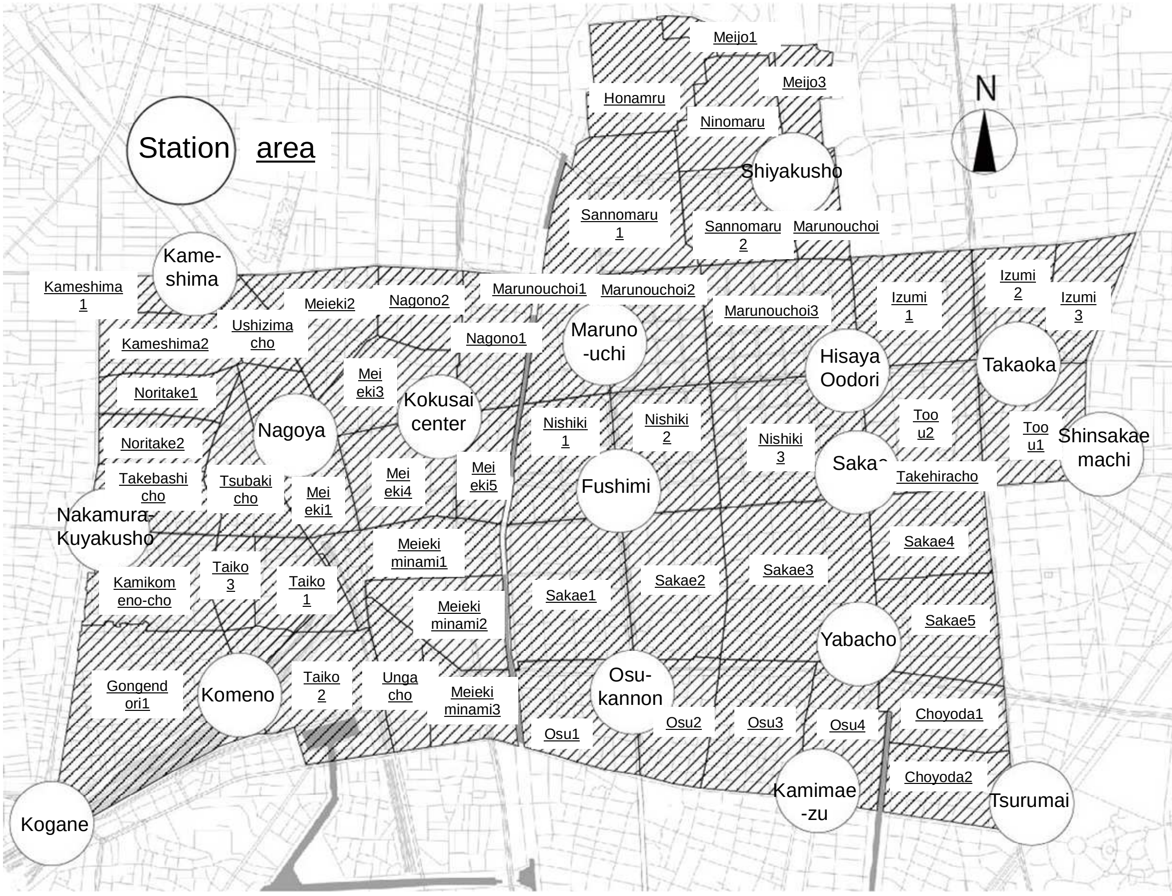

Fig. 1: Targeted Nagoya CBD and area division

4 EXTRACTION OF STOPPING OFF POINTS BY TIME ZONE WITHIN THE CITY CENTER USING SPATIAL-TEMPORAL POSITION DATA

Stopping off points were defined as follows. We extracted the five-minute stay behavior of visitors and defined the five-minute stopping off points as those points (Fig. 2, Table 2). First, visitors who passed through the area were extracted because of the large size of the data (Table 2-ii). If there were five or more points inside the circle, the center point was designated as the five-minute stop point, the beginning point was designated as the start time of the stop, and the number of points was designated as the stop time (Table 2-iii). Next, to determine whether or not the sample was at home, we pseudo-checked whether or not the linear distance between 0:00 and 23:59 was within 25 meter. If it was within 25 meter, the first and last stop points of the sample were deleted as home (Table 2-iii). Finally, based on the time of the start of the stop, the sample was divided into 0-12:00, 12-18:00, and 18-24:00, and divided into stop points at each time of day.

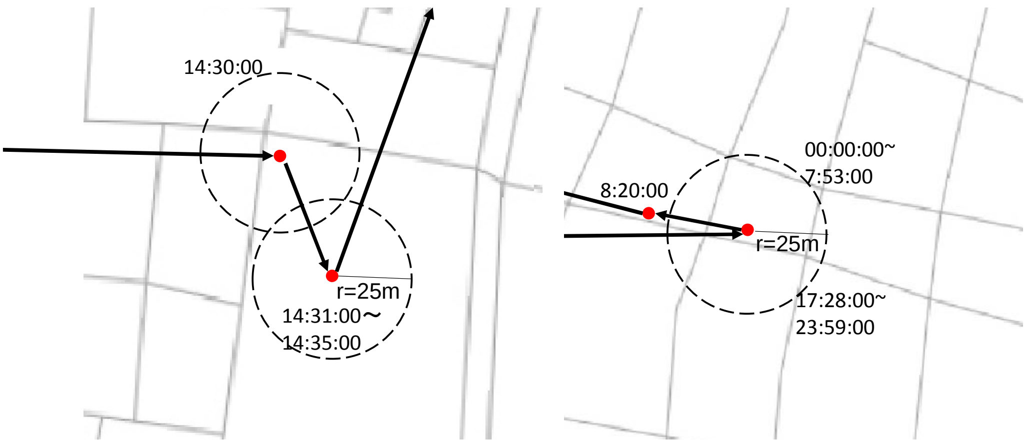

Fig. 2: Determination process:Five-minute stop points definition and decision to go home

Steps		the number of people	Stopping-off points
i	All samples	208,543	
ii	Sample passing through the area	17,445	
Determination process			
iii	Samples of stopping-off in the area	9,910	14,536
iv-1	0:00~12:00 samples		8,775
iv-ii	12:00~18:00 samples		4,220
iv-iii	18:00~24:00 samples		1,514

Table 2: Number of samples in each process

The results of the extraction showed that the spatial-temporal position data had larger values than the results of the PT survey, except for the 8:00 p.m. time slot, suggesting that shorter stop times could be extracted from the data.

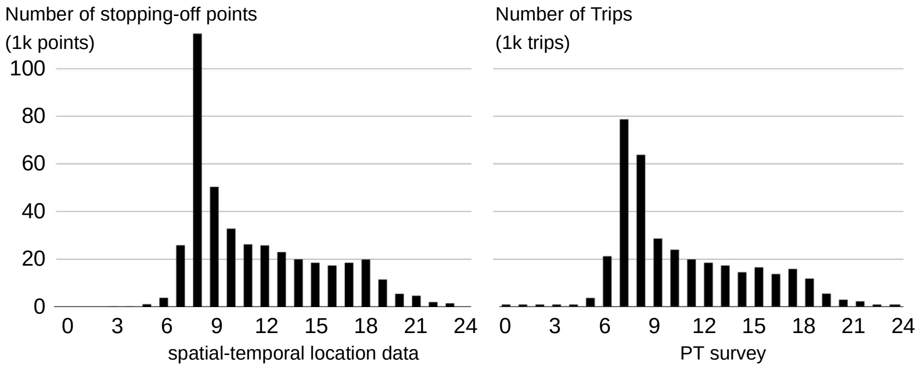

Fig. 3: Comparison of point histograms by time period for PT survey and spatial-temporal location data

Figure 4 shows the spatial distribution of stopping-off points extracted from the patial-temporal position data by sample.

In particular, from 18:00 to 24:00, the distribution is centered on Meieki 1, 3, and 5-chome, Sakae 3-chome, and Nishiki 2 and 3-chome. In addition, it can be seen that many of the stops are for 2 or 3 hours overall.

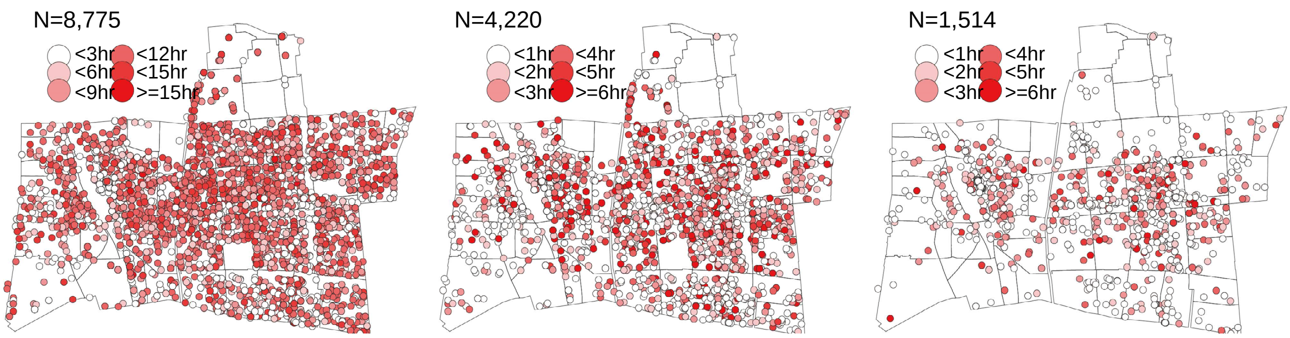

Fig. 4: Spatial distribution of stopping-off points by stop time (left: 0:00-12:00, middle: 12:00-18:00, right: 18:00-24:00)

Next, a histogram of stop times for each time period (0-12, 12-18, and 18-24 hours) is shown in Figure 5. 60-minute stays are prominent in the 0-12 hour time period, but this is thought to be due to a 5-minute transit stop. The peak at 600 minutes can be interpreted as a stay at the office due to employment.

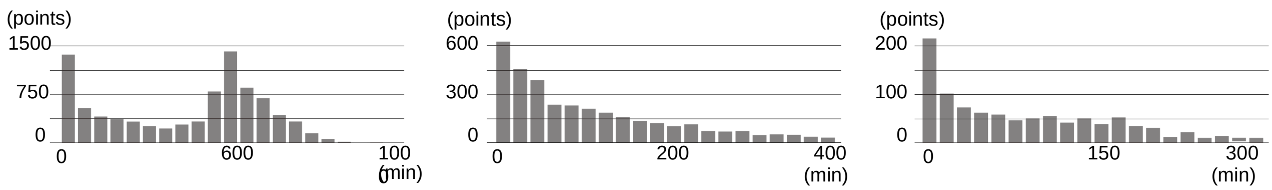

Fig. 5: Histograms of stop times by time zone(left: 0:00-12:00, middle: 12:00-18:00, right: 18:00-24:00)

ISBN 978-3-9504945-1-8. Editors: M. SCHRENK, V. V: POPOVICH, P. ZEILE, P. ELISEI, C.BEYER, J. RYSER

5 CHARACTERISTICS OF INNER-CITY DISTRICTS BASED ON RESTAURANT API

Next, the restaurant API data was used to explore the number of restaurants and restaurant types for each time period.

The restaurant API data was plotted on a map (Figure 6). By time zone, the number of restaurants open between 12 and 6 p.m. increased (2,208), with a particularly marked increase in Osu and Sakae.

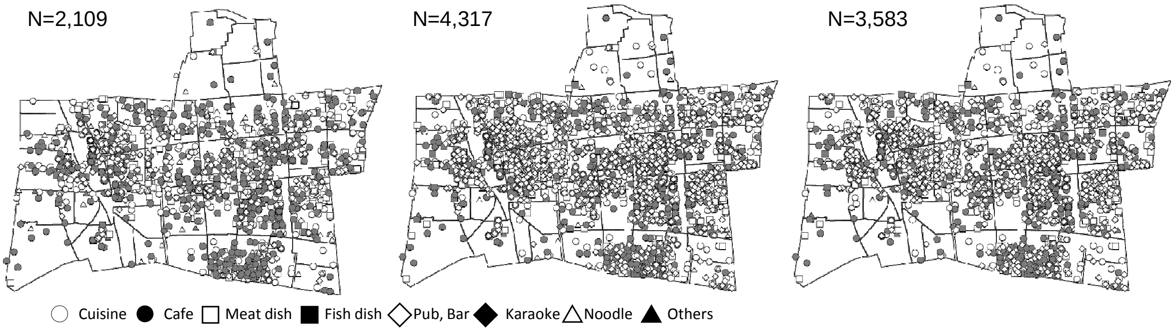

Fig. 6: Space distribution of restaurants by restaurant types (left: 0:00-12:00, middle: 12:00-18:00, right: 18:00-24:00)

Aggregate results by district also show that of the 22 total restaurant types, Meieki 3 and 4 and Sakae 3 and Nishiki 3 have more than 20 restaurants types. (Table 3).

	Right:Restaurants/km2 Left:types							Right:Restaurants/km2 Left:types					
	0~12		12~18		18~24			0~12		12~18		18~24	
Ushijimacho	263	12	338	13	213	10	Ninomaru	10	2	16	3	10	2
Nagono1	253	10	464	16	349	15	Sannomaru1	21	6	37	8	28	7
Nagono2	149	6	191	8	166	6	Sannomaru2	5	1	15	2	15	2
Meieki1	856	19	939	19	714	17	Marunouchi	52	2	73	3	52	2
Meieki2	222	16	475	18	353	16	Marunouchi1	111	10	222	13	185	12
Meieki3	488	17	1515	20	1240	20	Marunouchi2	165	10	266	14	188	12
Meieki4	708	19	1401	20	1057	20	Marunouchi3	126	13	215	17	147	14
Meieki4	247	10	585	17	476	14	Izumi1	231	15	449	17	342	16
Taiko1	57	4	85	5	57	3	Izumi2	175	8	330	14	265	13
Taiko2	0	0	0	0	0	0	Izumi3	96	6	208	12	178	10
Taiko3	72	4	131	7	87	5	Touou1	238	14	363	16	300	16
Takebashicho	55	5	111	7	96	6	Touou2	209	13	489	17	387	15
Tsubakicho	399	16	1231	19	1123	19	Takehiracho	107	4	214	6	150	4
Noritake1	116	5	270	10	250	9	Sakae1	240	15	487	18	389	17
Noritake2	83	4	148	7	138	7	Sakae2	222	15	387	17	260	16
Kamejima1	12	1	113	7	113	7	Sakae3	626	19	1203	21	1077	21
Kamejima2	53	5	99	8	66	7	Sakae4	504	14	1397	20	1282	19
Meiekiminami1	237	14	473	15	340	14	Sakae5	168	13	323	16	261	16
Meiekiminami2	29	4	35	5	17	3	Nishiki1	224	11	385	15	286	14
Meiekiminami3	48	4	56	5	40	4	Nishiki2	292	15	563	19	432	17
Meiekiminami4	9	1	27	3	27	3	Nishiki3	481	18	1737	21	1653	21
Kamikomenocho1	45	7	68	10	36	7	Chiyoda1	12	2	44	5	44	5
Unngacho	118	13	141	14	103	12	Chiyoda2	53	4	124	7	100	6
Gongendori1	18	4	24	6	18	5	Osu1	22	3	50	6	41	6
Makinocho	0	0	0	0	0	0	Osu2	277	9	495	16	448	15
Meijo1	0	0	0	0	0	0	Osu3	411	15	602	19	506	19
Meijo3	20	2	20	2	20	2	Osu4	151	13	244	16	177	13
Honmaru	9	2	19	4	14	3							

Table 3: Number of restaurants and restaurants types by time zone by district

6 ANALYSIS OF STOPS IN NAGOYA CBD USING LATENT FACTORS IN PLS REGRESSION

Next, we get a latent factor analysis of PLS regression, which stands for partial least squares regression, a technique that can be adapted to data with the threat of multicollinearity. PLS regression analysis was used because of the highly correlated combinations found among the variables in this study. With the number of visitors per district area as the objective variable, latent factors are analyzed by time period (0-12, 12-18, and 18-24 hours). The objective and explanatory variables are organized and presented in Table 4. The equation for PLS regression analysis is as follows.

$$X = t_a p_a^T + t_b p_b^T + t_c p_c^T + t_d p_d^T + t_e p_e^T + E$$

$$y = t_a q_a + t_b q_b + t_c q_c + t_d q_d + t_e q_e + f$$

(tn: n-th principal component, pn: n-th loading, E: residual of X, pn: n-th coefficient, f: residual of Y)

	Variables	Source	Note
y	Number of visitors per district area	A	Stopping-off Start Time
a	Number of restaurants per district area	B	Opening time
b	Number of restaurant types per district area	B	Opening time
c	Number of recreational facilities per district area	C+D	
d	Number of cultural facilities per district area	D	
e	Number of retail stores per district area	C	

Table 4: Objective and explanatory variables used in PLS regression analysis by time period

A summary of the analysis results is shown in Table 5, where the number of latent factors with the lowest PressRMSE(Root Mean Squared Error) was selected, resulting in a latent factor number of 2 for all time periods. The regression equation for the relationship between the number of visitors per unit area by district was obtained for these two latent variables.

		0-12	12-18	18-24
Number of components		2	2	2
R2		0.606	0.314	0.109
PressRMSE	first latent variable	0.00052	0.000254	0.000195
	second latent variable	0.00112	0.000433	0.00021
Cumulative contribution rate		72.7	68.4	62.7

Table 5: Summary in PLS regression analysis

Next, we examine the standard partial regression coefficients by time period (Table 6,7,8).

In particular, the absolute values of the number of entertainment facilities (c: 0.448), cultural facilities (d: -0.369), restaurant types (b: 0.213), restaurants (a: 0.157), and retail stores (e: -0.117) were higher between 18 and 24 hours, in that order.

A comparison of the rank order of the standard partial regression coefficients across time periods reveals that the number of restaurants and the number of restaurant types are higher in the 0-12 time period, while the number of entertainment facilities is higher in the 12-18 and 18-24 time periods.

	a	b	c	d	e
Standard Variance Regression Coefficient	0.257	0.429	0.163	-0.083	-0.0028
first latent variable	0.004	89.446	0.012	0.000	0.000
second latent variable	0.222	0.000	0.959	0.005	-0.011

Table 6: Latent variable loadings(0-12)

	a	b	c	d	e
Standard Variance Regression Coefficient	0.152	0.261	0.549	-0.245	-0.028
first latent variable	0.003	40.598	0.007	0.000	0.000
second latent variable	0.184	0.000	0.781	0.005	-0.011

Table 7: Latent variable loadings(12-18)

	a	b	c	d	e
Standard Variance Regression Coefficient	0.157	0.213	0.448	-0.369	-0.117
	0.001	13.478	0.003	0.000	0.000
second latent variable	0.061	0.000	0.305	-0.008	-0.002

Table 8: Latent variable loadings(18-24)

In PLS regression analysis, as in principal component analysis, the latent variables are orthogonal and the axes can be interpreted as an overall indicator of the variables. Therefore, each latent variable is interpreted by looking at the loadings of the latent variables and the distribution of latent factor scores. The spatial distribution of latent factor scores by time period is shown in Figures 7, 8 and 9.

The spatial distribution shows that there is no significant difference between the first and second latent variables by time period. However, looking at the loadings for each latent variable, the first latent variable is

ISBN 978-3-9504945-1-8. Editors: M. SCHRENK, V. V: POPOVICH, P. ZEILE, P. ELISEI, C.BEYER, J. RYSER

0.894 for the number of restaurant types from 0-12:00, 0.406 from 12-18:00, and 0.134 from 18-24:00. Although the loadings decrease as the hours get later, they are higher than the other variables, which can be considered the "axis of restaurant diversity" can be considered.

The second latent variable can be considered as the "urban entertainment axis" since the number of entertainment establishments is 0.959 from 0-12am, 0.781 from 12-18am, and 0.305 from 18-24am, similarly taking higher values than the other variables.

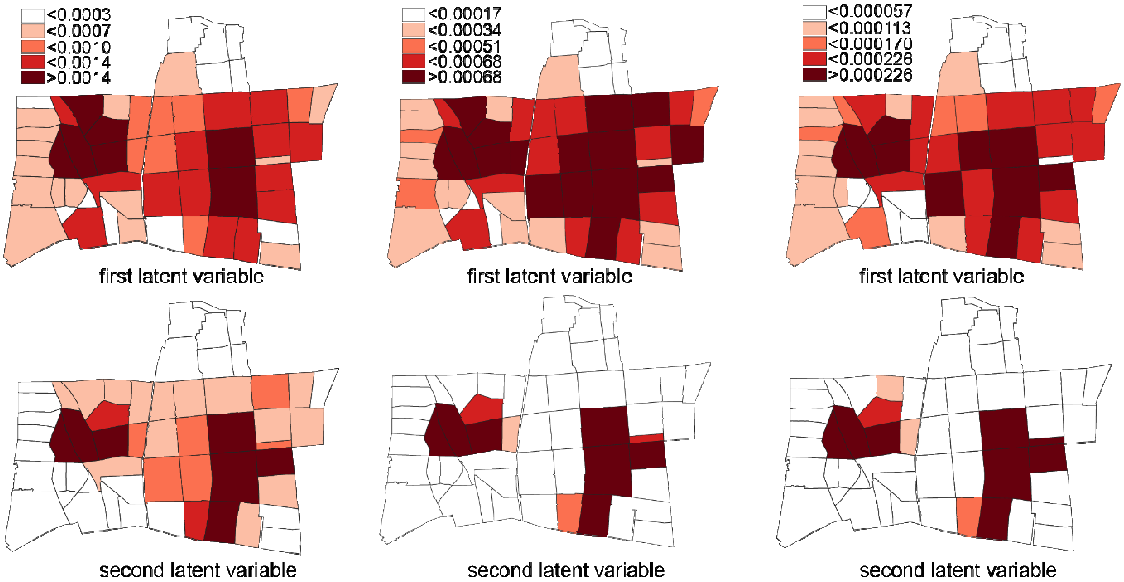

Fig. 7: Spatial distribution of latent factor scores (left: 0:00-12:00, middle: 12:00-18:00, right: 18:00-24:00)

7 ESTIMATING A MODEL OF NIGHTTIME DISTRICT CHOICE BEHAVIOR USING MULTINOMIAL LOGIT ANALYSIS

Next, we consider that there are routine and non-routine free activities in the nighttime after work because of the characteristics of the spatial-temporal location data and the fact that it is a weekday. Therefore, we estimate the district choice probability of visitors to the Nagoya city center area by conducting a multinomial logit analysis, with the presence or absence of individual district choice as the objective variable and the linear distance between the 18-24 pm stop as the explanatory variable in addition to the explanatory variables in the PLS regression analysis. The variables used are shown in Table 9.

	Variables	Source	Note
y	Visitation choice by district	A	Stopping-off Start Time
a	Number of restaurants per district area	B	Opening time
b	Number of restaurant types per district area	B	Opening time
c	Number of recreational facilities per district area	C+D	
d	Number of cultural facilities per district area	D	
e	Number of retail stores per district area	C	
f	Distance between stopping-off points	A	Distance between stopping-off points in 12-18

Table 9: Variables used in multinomial logit analysis

	a	b	c	d	e	f
a	1					
b	-0.139	1				
c	0.047	-0.608	1			
d	-0.047	-0.318	0.296	1		
e	0.467	-0.419	0.362	0.04	1	
f	-0.063	-0.204	0.122	0.083	0.022	1

Table 10: Correlation matrix between variables

Since multinomial logit analysis requires consideration of multicollinearity among explanatory variables, correlation coefficients among each explanatory variable were obtained and variables were selected. As a result, the number of restaurants (a), the number of cultural facilities (d), and the distance between stops (f), which are not subject to multicollinearity, were used in the analysis (Table 10).

The equation for individual utility obtained from the multinomial logit analysis is as follows:

$$\log\ (y) = 0.606 \log (a) - 0.216 \log (d) - 1.995 \log (f) + 6.151$$

Using the cross-validation method based on the parameters obtained, a high value of 0.984 was obtained for the success rate.

A summary of the analysis results is shown in Table 11. The negative values for the distance between stops indicate that the longer the distance, the smaller the utility an individual can obtain from the district. We also find that the individual's utility increases as the number of restaurants per unit area increases. This suggests that there is a trade-off between the utility gained from one restaurant and the proximity of 3 meters.

	parameter	odds	VIF	p-value	AIC	Success rate
intercept	6.151				6937.4	0.984
a	0.606	1.833	1.024	<0.05		
d	-0.216	0.805	1.011	<0.05		
f	-1.995	0.136	1.019	<0.05		

Table 11: Variables used in multinomial logit analysis

From the obtained model equation, the individual's probability of choice by district is estimated and illustrated in the figure. The estimated probabilities of choice by district are Meieki 1, Nishiki 3, Meieki 4, and Sakae 3, in descending order of probability. In addition, there are two distributions, one centered on Meieki and the other centered on Sakae (Figure 8 left).

Next, using this estimated model equation, we estimate where and how much the number of restaurants and facilities by district should increase to increase the probability of individuals choosing any given point over competing districts. As an example, we selected the Osu 1, 2, 3, and 4 districts and estimated the district-specific selection probabilities for the case in which the number of restaurants was increased tenfold from 1,172.8 to 11,728.3. The results showed an effect of reducing the selection probability of Sakae 3 by one rank and increasing the selection probability of Osu 3 by one rank (Figure 8, right).

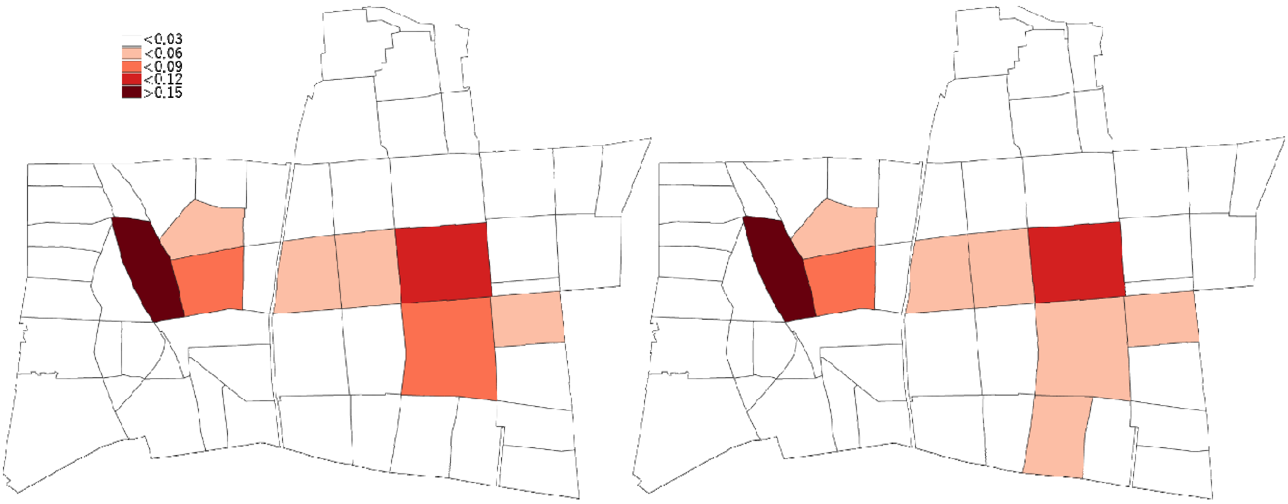

Fig. 8: Estimated probability of selection by district (left: existing state, right: Increase the number of restaurants in Osu area by 10 times)

8 CONCLUSION

This study clarified the factors of stops in the central Nagoya CBD and estimated district-specific selection probabilities based mainly on the spatial-temporal location data and the restaurant API data.

ISBN 978-3-9504945-1-8. Editors: M. SCHRENK, V. V: POPOVICH, P. ZEILE, P. ELISEI, C.BEYER, J. RYSER

1) The standard partial regression coefficients of the PLS regression showed that the number and type of restaurants were the top factors during 0-12 hours, while entertainment facilities were the top factors during 12-18 hours and 18-24 hours. Regardless of the time of day, the first latent variable could be interpreted as the restaurant diversity axis and the second latent variable as the urban entertainment axis.

2) A multinomial logit analysis revealed a trade-off between the 3-meter proximity and the utility gained from one restaurant. In Nagoya CBD, the probability of choice is high in Meieki and Sakae, but increasing the number of restaurants in the Osu area increases the probability of choice in the Osu area and decreases the probability of choice in Sakae 3.

9 REFERENCES

Y. Sakurai, S. Miyazaki and A. Fujii, Construction of a Choice Behavior Model for Commercial Areas Using a Multinomial Logit Model and Analysis of Trade Areas, Journal of Urban Planning, 2011, Vol.46, No.3, p.427-432

Suehashi, K., Saito, S., and Kumada, T., Development of an aggregate circulation Lukoff model for predicting sales based on frequency of visits to a commercial district in Komakura, in Proceedings of the City Planning Institute of Japan, 1998, Vol.33, No.3.

Edward L. Glaeser, Hyunjin Kim, Michael Luca, Nowcasting the Local Economy:Using Yelp Data to Measure Economic Activity,Harvard Business School NOM Unit Working Paper No. 18-022

Akihiko Namie, An Attempt of Regional Analysis Using PLS Regression: A Case Study of Garbage Discharge in Fukui Prefecture, Geographical Review, 2007, Vol. 80, No. 4, p.178-191