モビリティデータに基づく東京における消費者行動の探求：新しいショッピングセンターの開発

劉璐瑶*・馬玨**・龐岩博**・澁谷遊野**・関本義秀**

Exploring Consumer Behavior in Tokyo based on Mobility Data: Development of a New Shopping Center

Luyao LIU*, Jue MA**, Yanbo PANG**, Yuya SHIBUYA**, Yoshihide SEKIMOTO**

Abstract Social changes have intensified the complexity of consumer behavior, while technological developments have facilitated the exploration of its attributes. Based on GPS data, we introduced a method to analyze both categorical and continuous variables, assigning weighted values to their impact on consumer behavior. Additionally, we improved the Pseudo People Flow Model to generate people activity data based on user input scenarios. This enhanced model provides decision support by simulating scenarios before implementing business strategies and urban planning, offering a more precise and user-friendly tool for predicting and optimizing consumer behavior. We identified significant variables influencing on consumer behavior in Tokyo and tested their impact on people activity in Shibuya ward through a user-friendly interface when developing a new Shopping center (SC).

Keywords: 消費者行動 (Consumer Behavior), ショッピングセンター (Shopping Center), 人流シミュレーション (People Flow Simulation), エージェントベースの分析 (Agent-based model), 東京 23 区 (Tokyo 23 Wards)

1. Introduction

1.1 Background

As social development and economic growth evolve, consumer behavior and urban planning are becoming increasingly complex (Wlodarczyk, 2021). These changes necessitate the integration of big data analysis and intelligent technologies for a deeper understanding of consumer behavior. However, current research on consumer behavior mainly focusses on marketing and psychology (Ma et al., 2023). Additionally, analyzing and predicting consumer behavior is challenging primarily due to its dynamic nature (Birkin, 2019) and the complexity of its determinants (Rozenkowska, 2023). Furthermore, models available to the public are often in a prototype development stage, which lowers their efficiency and applicability to the public (Shibuya et al., 2022).

1.2 Research objective and workflow

Our study aims to develop an easy-to-use model for simulating consumer behavior prior to implementing new business strategies and urban planning. Our research includes data construction, consumer behavior analysis and simulation analysis. First, we build a SC attribute database and a consumer point dataset with GPS data. Second, we integrate categorical and continuous variables to identify significant variables influencing consumer behavior, and then standardize these variables. Finally, the resulting weighted values are used to improve the destination component of the Pseudo People Flow Model, simulating people activity when a new SC is introduced.

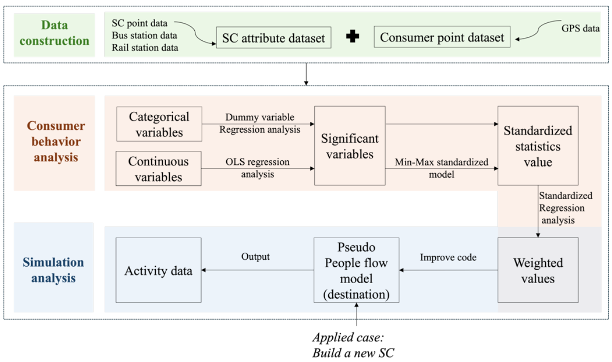

Figure 1 Research framework

* 学生会員 東京大学大学院新領域創成科学研究科 (The University of Tokyo) **liuluyao@g.ecc.u-tokyo.ac.jp**
** 正会員 東京大学空間情報科学研究センター (The University of Tokyo)

2. Data and methodology

2.1 Study area and data sources

Tokyo 23 wards, one of the most developed cities globally, attract consumers from both within and outside the city due to their numerous and diverse businesses. Furthermore, the robust brick-and-mortar retail sector in these wards provides an excellent context for studying consumer behavior across various groups.

In this paper, a SC is defined as having a retail area of at least 1,500 square meters and containing a minimum of 10 tenants. We identified 229 SC points in Tokyo 23 wards with important SC information, provided by the Japan Council of Shopping Centers (Figure 2), and used bus station data (2022) and rail station data (2022) with coordinates, provided by Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan.

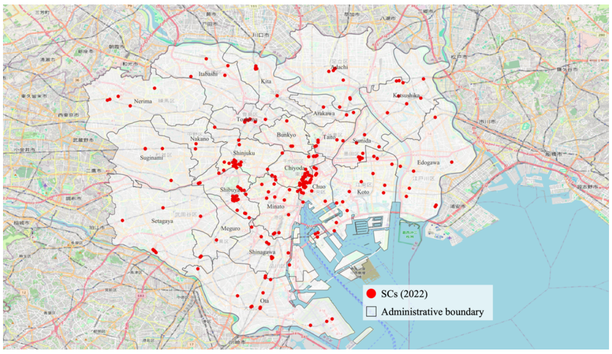

Figure 2 Tokyo 23 wards with SC points (2022)

For mobility data, we utilized GPS data (2022) provided by AGOOP Corp., which includes anonymized location information from mobile devices, and Pseudo People Flow Data version 2.0 with synthesized human movement patterns, provided by Center for Spatial Information Science (CSIS), The University of Tokyo.

2.2 Data processing

First, considering data available, research needs and related literatures, we utilized SC point data (2022), bus station data (2022) and rail station data (2022) to construct our SC attribute dataset using QGIS software. This dataset includes SC building purpose, location, parking lot capacity, site area, total floor building area, area ratio of all tenements, area ratio of retail (R), food & beverage (FB) and service (S) tenements, quality ratio of R, FB and S tenements, quality of R tenements, quality of FB tenements, quality of S tenements, quality of all tenements, the distance to the nearest rail station, the distance to the nearest bus station, and the distance to the nearest SC.

Second, we preprocessed GPS data (2022) utilizing the "scikit-mobility" framework to apply noise filtering, stop detection and trajectory compression (Pappalardo et al., 2022). We set thresholds as 10 minutes and 500 meters to identify data points that demonstrate consumer behavior from May 2, 2022, to May 14, 2022.

2.3 Identification of significant variables of consumer behavior

In this study, we assessed the impact of various SC attributes on their attraction, as measured by the number of visitors.

Firstly, dummy variable regression was used to examine the effects of different categorical variables on dependent variables (Zhou et al., 2021). In this analysis, the first type of each categorical variables was excluded as the reference type to avoid collinearity issues. Significant variables were identified by analyzing the p-values, with those below 0.05 considered to have a meaningful impact. In our research, categorical variables include "building purpose", classified into commercial building, office building, station building, underground street, elevated structure, residential building, complex building, airport building and hotel building, and "location" classified into urban area and surrounding area.

Secondly, an Ordinary Least Squares regression model was used to identify and exclude non-significant variables from dependent variables through p-value analysis, with those below 0.05 considered to have a significant impact. We standardized these variables using the min-max method, which scales the data to a range of 0 to 1, ensuring comparability across different scales (Le et al., 2019).

Finally, these variables, along with the date, were combined as a panel data and analyzed using a standardized regression model to conduct a comprehensive assessment of their respective influences over 14 days. This approach can capture the time series

Table 1 Regression results of categorical variables and continuous variables (P-value)

Variables/Visit data of May. 2022	2	3	4	5	6	7	8	9	10	11	12	13	14	15
office building	0.02	0.03	0.04	0.04	0.02	0.05	0.04	0.01	0.01	0.01	0.01	0.01	0.05	0.05
station building	0.02	0.07	0.06	0.05	0.02	0.03	0.04	0.01	0.01	0.02	0.01	0.01	0.03	0.03
underground street	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
elevated structure	0.80	0.73	0.71	0.71	0.83	0.79	0.75	0.86	0.86	0.88	0.92	0.87	0.79	0.73
residential building	0.14	0.14	0.16	0.15	0.14	0.16	0.16	0.15	0.14	0.14	0.15	0.14	0.16	0.16
complex building	0.53	0.61	0.64	0.63	0.50	0.56	0.65	0.41	0.42	0.39	0.35	0.33	0.57	0.62
airport building	0.52	0.61	0.58	0.60	0.50	0.52	0.57	0.48	0.46	0.45	0.46	0.52	0.50	0.56
hotel building	0.93	0.96	0.92	0.94	0.94	0.94	0.91	0.79	0.86	0.83	0.78	0.76	0.99	0.95
surrounding area	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
parking lot capacity	0.49	0.38	0.42	0.41	0.51	0.46	0.39	0.60	0.57	0.57	0.61	0.59	0.49	0.44
site area	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02
total floor building area	0.24	0.25	0.27	0.28	0.26	0.27	0.26	0.26	0.22	0.21	0.23	0.25	0.02	0.28
quality of R, FB and S tenements	0.74	0.72	0.63	0.68	0.62	0.65	0.68	0.68	0.69	0.66	0.76	0.79	0.26	0.68
area ratio of R, FB and S	0.42	0.37	0.35	0.39	0.41	0.41	0.41	0.47	0.43	0.44	0.47	0.49	0.41	0.42
area ration of all tenements	0.77	0.59	0.54	0.64	0.70	0.61	0.63	0.88	0.78	0.81	0.86	0.91	0.63	0.63
quality ratio of R tenement	0.85	0.82	0.83	0.84	0.80	0.80	0.82	0.79	0.78	0.74	0.73	0.75	0.81	0.83
quality ratio of FB tenement	0.90	0.87	0.89	0.90	0.86	0.88	0.89	0.84	0.83	0.79	0.77	0.78	0.89	0.91
quality ratio of S tenement	0.98	0.99	0.99	0.97	0.97	0.96	0.99	0.95	0.94	0.90	0.88	0.91	0.97	1.00
quality of all tenements	0.14	0.10	0.14	0.12	0.19	0.18	0.15	0.20	0.19	0.20	0.17	0.16	0.20	0.18
distance to the nearest SC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
distance to the nearest bus station	0.23	0.27	0.31	0.28	0.23	0.26	0.24	0.26	0.22	0.21	0.22	0.23	0.25	0.25
distance to the nearest rail station	0.35	0.45	0.43	0.42	0.32	0.39	0.40	0.29	0.31	0.30	0.28	0.26	0.37	0.38

and cross-sectional dimensions of the data and ensures that differences in scale do not distort the regression results (Singh et al., 2022).

2.4 Improvement of Pseudo People Flow Model

The activity generator in the Pseudo People Flow Model can predict the activity purpose and destination of everyone, based on a time-inhomogeneous Markov chain model (Kashiyama et al., 2024). Regarding the destination, it is determined in a two-step process. First, the model selects a gird with broader boundary level, such as cities, grids, or school districts. Second, the location at the building level is chosen using building data. When the activity purpose is “Leisure and Business”, the allocation probabilities of each grid are determined using the number of employees for the business type, which is referred to as grid facility capacity.

In this paper, we focus on consumer behavior and improve the accuracy of “SHOPPING” component simulation results. To achieve this, we assign different weighted values to various significant variables, reflecting the varying impacts of different types of SCs. Our enhanced simulation package follows a two-step process. First, when the activity purpose is “SHOPPING”, the choice probability for each gird is calculated based on new SC parameters provided by the user, which act on economic data. These parameters include coordinates, SC building purpose, location, site area, and distance to the nearest SC (automatically calculated based on coordinates). Then, the choice probability for each SC building within the chosen grid is determined using the building data (Figure 3).

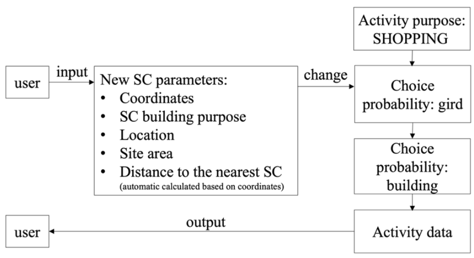

Figure 3 The SC agent-based model workflow

3. Results

3.1 Consumer behavior attributes in Tokyo 23 wards

In general, regression results revealed that “commercial building”, “office building”, “station building”, “underground street”, “urban area” and “surrounding area”, “site area” and “distance to the nearest SC” variables had p-values smaller than 0.05 (Table 1).

Standardized regression analysis identified normalized weighted values for each significant variable, and showed this model passed the collinearity test (VIF<5) (Figure 4).

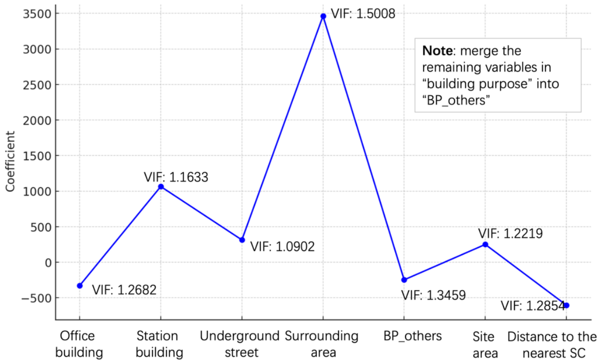

Figure 4 Standardized regression results (Coef and VIF)

3.2 Simulation results

Shibuya ward, with its diverse commercial facilities and a wide range of consumer groups, serves as a model for commercial development in Tokyo. In this study, a new station building with a floor area of 8,000 square meters was constructed near Shibuya Station in the urban area (139.693, 35.664), and its impact on pedestrian flow in the Shibuya area was simulated.

Figure 5 and Figure 6 illustrate the impact of a new SC on people activity in Shibuya ward, highlighting an increase concentration of activity around the station and the new SC, while activity near Yoyogi park decreased. The new SC mainly influences consumer behavior during noon and at night. The simulation reveals a significant increase in people activity during peak times, suggesting that the new SC effectively attracts consumers.

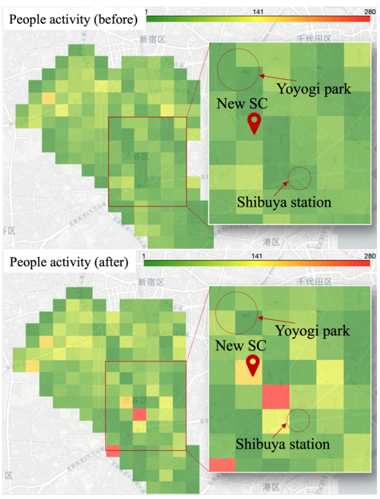

Figure 5 People flow in 11 a.m. (before and after)

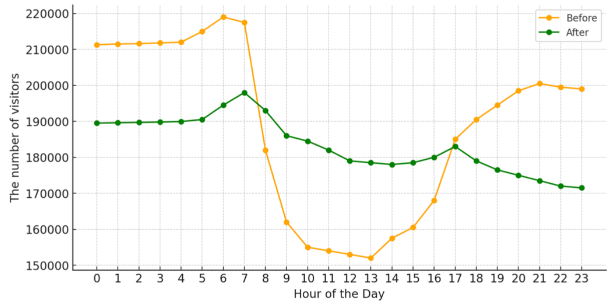

Figure 6 People flow in 24 hours (before and after)

4. Conclusion and future work

This study primarily identified significant variables affecting on consumer behavior in Tokyo and developed an enhanced agent-based model to simulate the impact of new SCs, utilizing two types of variables and enhancing the Pseudo People Flow Model. These efforts establish a robust framework for urban planners and business strategists to simulate and predict the potential impacts of their decisions. However, the current model has limitations in integrating different types of people and retail types in-depth. Additionally, a single case study is insufficient to verify the reliability of simulation results. Future enhancements will include the integration of additional consumer and retail types, and validation against real-world data to improve predictive accuracy.

References

Wlodarczyk, K. (2021). Trends of Evolution in Consumer Behavior in the Contemporary World. *PROBLEMY ZARZADZANIA-MANAGEMENT ISSUES*, *19*(1), 39–50.

Ma, Z., Guo, G., & Jing, L. (2023). Consumer behavior research in the 21st century: Clusters, themes, and future research agenda. *International Journal of Consumer Studies*, *48*(1), e12980.

Birkin, M. (2019). Spatial data analytics of mobility with consumer data. *Journal of Transport Geography*, *76*, 245–253.

Rozenkowska, K. (2023). Theory of planned behavior in consumer behavior research: A systematic literature review. *International Journal of Consumer Studies*, *47*(6), 2670–2700.

Shibuya, Y., Pang, Y., & Sekimoto, Y. (2022). A scenario-based generation of pseudo people flows in urban spaces. *Proceedings of GISA 2022*, Tokyo, Japan, 1-10.

Pappalardo, L., Simini, F., Barlacchi, G., & Pellungrini, R. (2022). scikit-mobility: A Python Library for the Analysis, Generation, and Risk Assessment of Mobility Data. *Journal of Statistical Software*, *103*, 1–38.

Zhou, W., Cheng, Y., Ding, S., Chen, L., & Li, R. (2021). A grey seasonal least square support vector regression model for time series forecasting. *ISA Transactions*, *114*, 82–98.

Le, T.-H., Chang, Y., Taghizadeh-Hesary, F., & Yoshino, N. (2019). Energy insecurity in Asia: A multi-dimensional analysis. *Economic Modelling*, *83*, 84–95.

Singh, A., Jenamani, M., Thakkar, J. J., & Rana, N. P. (2022). Quantifying the effect of eWOM embedded consumer perceptions on sales: An integrated aspect-level sentiment analysis and panel data modeling approach. *Journal of Business Research*, *138*, 52–64.

Kashiyama, T., Pang, Y., Shibuya, Y., Yabe, T., & Sekimoto, Y. (2024). Nationwide synthetic human mobility dataset construction from limited travel surveys and open data. *Computer-Aided Civil and Infrastructure Engineering*, 39(2), 146-165.