Study of the relationship between rental house prices and people congestion of urban areas based on mobile phone GPS data

Yinglan Qin
The University of Tokyo
Room 434, General Research Building, Kashiwanoha 5-1-5 Kashiwa city, Chiba, Japan
1187170943@edu.k.u-tokyo.ac.jp

Yuki Akiyama
The University of Tokyo, Center for Spatial Information Science
Room 404, General Research Building, Kashiwanoha 5-1-5 Kashiwa city, Chiba, Japan
aki@csis.u-tokyo.ac.jp

Yoshiki Ogawa
The University of Tokyo, Institude of Industrial Science
Ce-509 Production Engineering Research Center Building, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan
Cogawa@csis.u-tokyo.ac.jp

Ryosuke Shibasaki
The University of Tokyo, Center for Spatial Information Science
General Research Building, Kashiwanoha 5-1-5 Kashiwa city, Japan
shiba@csis.u-tokyo.ac.jp

SAKURAMACHI Ritsu
Zenrin Datacom
c-6floor, Minato, Konan, 2 Chome-15-3 Tokyo, Japan
r_sakuramachi@zenrin-datacom.net

Abstract

Rental house price is affected by house function, surrounding environment and so on. However, there are less research of finding the relationship between rent prices with urban people congestion. There is also less research of urban people congestion by quantitative assessment. Since that, we will use location-based big data collected by mobile phone GPS data to define people congestion. Besides that, we are also trying to verify if there are some relationship between rental house prices with area people congestion in city. As a result, it is hard to evaluate rental house prices quantitatively only from people congestion but it was found to have a certain influence on rental house prices when calculated with multiple house related variables.

Keywords: rental house price； micro geodata； people flow; scikit-learn

1 Introduction

In the real estate industry in Japan, rent quotations are often determined based on the experience of house estate agents and there are still few studies that statistically analyze rental house pricing. The rent rate is affected by various factors, such as the value of the building itself (i.e., the facilities owned by the building, Morimoto (1995)), value of the area in which the property is located, surrounding environment (Luttik (2000)), timing of distribution, and state of the economy factors (Tsatsaronis and Zhu (2004)). In the world range, Kershaw and Rossini (1999) estimate constant quality house price indices by considering some Property Characteristics. But the number of samples are small and it didn't show the results by each single building but average.

In Japan, Akiyama and Ogawa (2018) is hitherto one of the few studies that have statistically analyzed the determinants of rent formation of each postal code unit all over Japan. In their study, a large number of explanatory variables (82 types) and the number of samples are used to analyze the factors affecting the prices of rental properties.

However, within this research method, it is difficult to collect and organize data as explanatory variables and, because the updating frequency of these explanatory variable data is low, making is also difficult to identify the most recent quotations. However, it is necessary to understand rent quotes, whose value changes on a daily basis.

Therefore, we use the resident population data for each area based on the movement history of mobile phones and quantitatively estimate the prosperity of the city, analysing whether the rent rate can be quantitatively evaluated based on the level of congestion in the city or not. The possibility of replacing a large number of explanatory variables which is necessary for explaining rent formation in Akiyama and Ogawa (2018) by using the prosperity of the city also be investigated.

2 Data

2.1 People flow

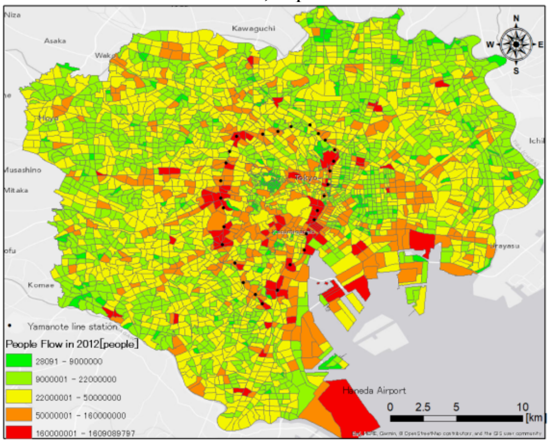
To quantitatively estimate the level of congestion in the city, Yamamoto et al. (2017) estimated the resident population for each street using the "Konzatsu-Tokei (R) " Data refers to people flows data collected by individual location data sent from mobile phone under users' consent, through Applications provided by NTT DOCOMO, INC. Those data is processed collectively and statistically in order to conceal the private information. Original location data is GPS data (latitude, longitude) sent in about every a minimum period of 5 minutes and does not include the information to specify individual. ※ Some applications such as "docomo map navi" service (map navi・local guide).

In this study, data for an entire year in 2012 on the 23 wards of Tokyo are used. The 23 wards as well as 23 special districts in eastern Tokyo include 70% of the population in Tokyo. We estimate the number of individuals residing in the city on a daily basis and sum them up them for one year, which represents the people flow.

Figure 1 shows the total people flow in 2012. A large number of people reside in the center of Tokyo especial around the Yamanote line (black point in Figure 1), which is one of the busiest train lines in Tokyo. Moreover, Haneda

Airport and train stations become a high people flow area. Further, we also distinguish the data as day-time, from 9 a.m. to 6 p.m. (JST), which is the working schedule of most individuals, and the remainder as night-time. Weekdays and holidays are also taken into account. This study considers both weekends and public holidays as holiday. The number of holidays and weekdays is 115 and 250 days, respectively. For comparing the people flow, all data are averaged daily.

Figure 1: Yearly People Flow in 2012 in Tokyo’s 23 wards, Japan



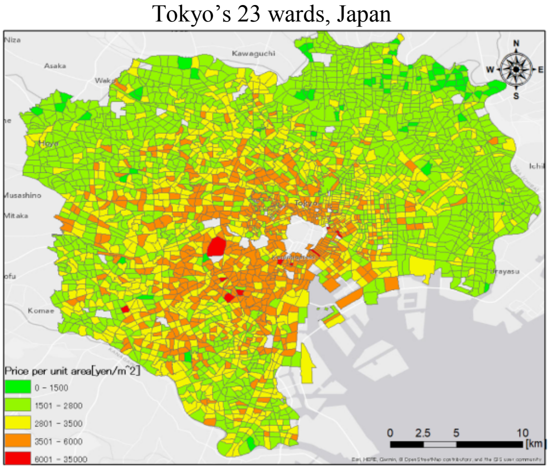
Source: “Konzatsu-Tokei®” ©ZENRIN DataCom CO.,LTD.

2.2 House prices and other parameters

To better grasp information on rental house prices and property attributes (building floor, building area, nearest station, walking time from nearest station, etc.), the "Real estate data library nationwide 1999–2016 dataset" of At Home Co. Ltd. is used. The data not include commercial facilities.

To match the data with the timing of the population data, we have studied rental property data on all 23 wards of Tokyo in 2012. We also acquired the longitude and latitude of buildings using the "CSV address matching service" of the University of Tokyo Spatial Information Science Research Center.

Figure 2: Rent house price per unit area in 2012, Tokyo’s 23 wards, Japan



Source: At Home Co. Ltd.

Figure 2 shows the rental house price per unit area The 23 wards in the west area in Tokyo have higher unit house prices compare to the other areas. In addition to the 82 types of variables analyzed by Akiyama and Ogawa (2018), history in Tokyo also thought effect.

2.3 Data processing

Using Python and the ArcGIS 10.6 Spatial Join function, we constructed a database that compares rent data with the congestion in the city by combining the data described in 2.1 and 2.2 with the urban area analyzed.

After excluding extreme values, the upper and lower 2.5% of the observations (total 5%) were removed, with all the parameters used in the analysis. The dataset included a total of 1,695,586 buildings. As these data are then averaged by ward, the number of wards becomes 2,538.

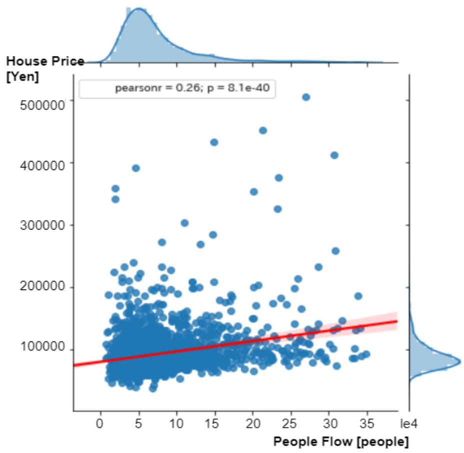
3 Results

3.1 Linear regression

Figure 3 shows the relationship between the daily people flows and rent house prices in the 23 wards of Tokyo. The coefficient of determination is 0.067 and the correlation coefficient 0.258. It is difficult to determine whether there is any correlation between the two variables. Moreover, other aggregations of the people flow do not show high correlations as well, although weekday and day-time show higher correlations compared to night-time and holidays (Table 1).

One reason for this might be that, since people that live in the city have variable lifestyles, even data show the same level of people flow, it could be caused by different reason. Therefore, it is not easy to identify which behaviors affect house pricing more. Therefore, it is difficult to evaluate rent house prices only from the people flow in the city.

Figure 3: Linear regression of rental house price and people flow



Source: “Konzatsu-Tokei®” ©ZENRIN DataCom CO.,LTD.

Table 1: Correlation coefficient of rental house price and people flow in day, night, holiday, and weekday

	day	night	holiday	weekday
R	0.271	0.248	0.146	0.291

Source: "Konzatsu-Tokei®" ©ZENRIN DataCom CO.,LTD.

3.2 SDG regression

Since it is difficult to evaluate rent house prices only from the people flow in the city, the following sections examine the influence of the people flow by comparing its effect to building areas, which have been proved to affect house prices (Akiyama and Ogawa (2018)) and also show a high correlation (R = 0.667, Figure 4, upper panel).

For calculations, the SGDRegressor module in scikit-learn is used. The Stochastic Gradient Descent (SGD) is considered a practical module when applied to large-scale and sparse machine learning problems. 80% data are used as training data, the rest of them are used as test data. Maximum number of iteration is set as 1000.

For comparing people flows in same level, all the people flow data are show by daily. A total of six parameters describe the daily people flow during day-time [people/day], night-time [people/day], total daily people flow [people/day], daily people flow during weekdays [people/day], in holiday [people/day], and building area [m^2]. All these six parameters are standardized using the StandardScaler module in Python.

For their better understanding, we compare the predicted SDG results to real house price by house price level (lower panel of Figure 4). The orange points show the difference calculated by all parameters. The blue points show the results only using building area. From the lower panel of Figure 4, increasing the variable of people flow leads to more realistic values than only using building area for all house price levels. The mean squared error (MSE) for only use building area is 5.695042e+08, while the MSE when using all parameters is 4.208356e+08.

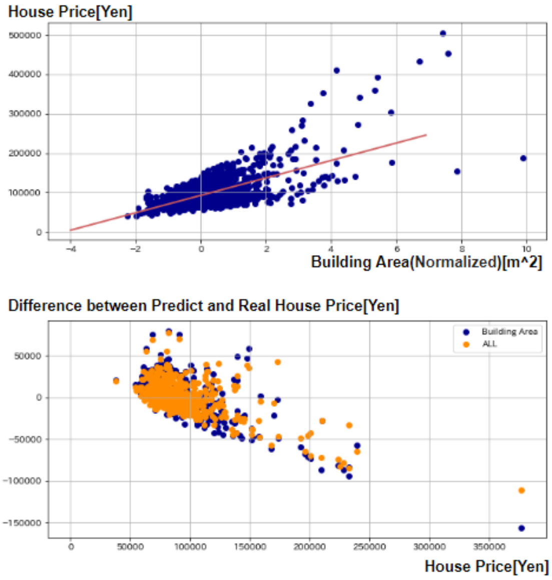
Table 2: Coefficients of six parameters' effects on rental house prices in the SGD regression

People flow					Building area
day	night	weekday	holiday	sum	
63,402	-45,350	3,592	-13,892	-1,195	21,397

Source: "Konzatsu-Tokei®" ©ZENRIN DataCom CO.,LTD.

Table 2 shows the effect of each variable on house prices. The coefficients in Table 2 are the results of 100 averages. We examine which variable influences the explanatory variable most. Day- and night-time daily people flows show significant effects on house prices, having a higher influence than building area. A possible reason for the results might be the higher people flows in the morning, which means the location is more convenient as a housing environment. Conversely, because of noise and other problems, housing environments with high people flows in the evening are evaluated lower. High people flows during holidays show significant negative effects in house prices possibly due to the same reason. Appropriate classification should be considered in future work.

Figure 4: Linear regression of rental house price and building area and difference between predict and real house price by house price level



Source: "Konzatsu-Tokei®" ©ZENRIN DataCom CO.,LTD.

3.3 Lasso regression

Here, another four parameters that have enough valid values, including in rent house price data, are used in the analysis. This section compares the Lasso and SGD regressions results. The least absolute shrinkage and selection operator (Lasso) module in scikit-learn is considering as a good way to compress the dimension. The coefficients of those variables do not influence explanatory variable become 0. As such, it is used to confirm whether all parameters are affected by house prices.

The 10 parameters include the five daily people flows that have already been introduced. Moreover, building floor, total building floor, building area, walking time from station, and age of building are used. Table 3 shows the coefficients on each parameter. The results of the SGD regression are the averages of 100 iterations. There are no coefficients become 0 after sparse estimation, which means all variables have significant effects on house prices. Each coefficients do not show significant differences when comparing the two results, expect the sum of daily people flows shows a significant relative difference: it is negative in the SGD regression and positive in the lasso regression.

Day- and night-time daily people flows show significant effects even in the results have more parameters compare to 3.2. It could become valuable variables when identify quotations of rental house price. The MSE of the SGD regression is 3.692990e+08, while that of the lasso regression is 3.688060e+08. The MSE of SGD is smaller compared to the result in 3.2. In other words, more meaningful parameters would help obtain realistic values of house prices. Since 5 other parameters is stable, it will not conflict with purpose which is identify the most recent quotations of rental house price.

Table 3: Coefficients of the 10 parameters on rental house prices in the SGD and lasso regressions

Model	People Flow				
	day	night	weekday	holiday	sum
SGD	51,153	-37,549	1,236	-8,807	-1,535
Lasso	43,450	-54,491	2,842	-8,077	20,890
	Floor	Total floor	Building area	Walking time from station	Age of building
SGD	-7,799	12,000	21,097	-3,539	-2,470
Lasso	-7,823	11,994	21,099	-3,528	-2,467

Source: “Konzatsu-Tokei®” ©ZENRIN DataCom CO.,LTD.

4 Conclusions and future work

In this study, we quantitatively estimate urban congestion based on resident population data of each area based on the GPS trajectory of mobile phones in all 23 wards of Tokyo, and quantitatively analyze if it is possible to evaluate the rent quotient by the level of urban congestion. The main results of this study are as follows. First, we could not evaluate rental house prices quantitatively only from urban people congestion expressed by the resident population. Second, from the results of SGD and lasso regression analyses using multiple variables, urban congestion was found to have a certain influence on rental house prices. Furthermore, day- and night-time congestion have positive and negative influence on rental house prices, respectively. Understanding the reason why these results were obtained requires further research. More appropriate way to separate time required as well.

There are two goal of this study. One is to reduce the time for data collection and organization when using a large number of variables to obtain rental house price quotations. The other one is to identify the most recent quotations of rental house price. From the results, it is hard to replace a large number of variables by people flow data. However, there are possibility to identify the most recent quotations by adding people flow as one variable.

In future studies, we will consider the calculation method for urban congestion using GPS data on people flows. Moreover, we will use the data with high immediacy along with other data for a quick update cycle to obtain more accurate results that are closer to recent reality.

5 Acknowledgements

This study was supported by the research and practice grants of Jusoken. In addition, the authors were given the “Konzatsu Toukei ®” from ZENRIN DataCom Co.LTD. The data is the person flow data developed by comprehensive and analytical processing of the GPS data with permission from phone users. Moreover, the authors were given a various spatial data from the joint research with Center for Spatial Information Science (CSIS), the University of Tokyo (research ID: 122). We would like to thank them for their great contributions.

6 References

Morimoto, N. (1995) Unit Rent,Floor Area and Location of Apartment House, URBAN HOUSING SCIENCES No 11, 1995 Autumn.

Luttik, J. (2000) The value of trees, water and open space as reflected by house prices in the Netherlands, Landscape and Urban Planning, Volume 48, Issues 3–4, Pages 161-167.

Tsatsaronis, K., Zhu, H. (2004) What Drives Housing Price Dynamics: Cross-Country Evidence, BIS Quarterly Review, March 2004.

Kershaw, P., Rossini, P. (1999) Using Neural Networks to Estimate Constant Quality House Price Indices, Fifth Annual Pacific-Rim Real Estate Society Conference.

Akiyama, Y., Ogawa, Y. (2018) Study of rent formation mechanism using micro geodata - Focusing on the difference between the value of housing and market value as seen from the household, Sumikin Research and Papers Collection, Practical Research Report Collection, 44, 1-12.

Yamamoto, Y., Akiyama, Y., Shibasaki, R., Shibasaki, R. (2017) Estimation of interregional capital flows using large-scale intercompany transaction data including establishment information and large scale human flows data, The 26th Geoinformatics Society Conference Papers Collection, C-6-1.