

LEARNING GEOGRAPHICAL DISTRIBUTION OF VACANT HOUSES USING CLOSED MUNICIPAL DATA: A CASE STUDY OF WAKAYAMA CITY, JAPAN

H. Baba^{1,*}, Y. Akiyama^{1,2}, T. Tokudomi³, Y. Takahashi⁴

¹ Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba, Japan - (hbaba, aki)@csis.u-tokyo.ac.jp

² Dept. of Urban and Civil Engineering, Tokyo City University, 1-28-1, Tamazutsumi, Setagaya-ku, Tokyo, Japan - akiyamay@tcu.ac.jp

³ Wakayama Data Utilization Promotion Center, 3-17, Higashikuramae-cho, Wakayama, Wakayama, Japan - tokudomi_t0001@pref.wakayama.lg.jp

⁴ Statistics Bureau of Japan, 3-17, Higashikuramae-cho, Wakayama, Wakayama, Japan - y7.takahashi@soumu.go.jp

Commission IV

KEY WORDS: Vacant house, Municipal data, Closed data, Resident registration, Building registration, Machine learning, XGBoost

ABSTRACT:

Vacant housing detection is an urgent problem that needs to be addressed. It is also a suitable example to promote utilisation of smart data that are stored in municipalities. This study proposes a vacant housing detection model that uses closed municipal data and considers accelerating the use of public data to promote smart cities. Employing a machine learning technique, this study ensures high predictive power for vacant housing detection. The model enables us to handle complex municipal data that include non-linear feature characteristics and substantial missing data. In particular, handling missing data is important in the practical use of closed municipal data because not all of the data are necessarily absorbed to a building unit. Consequently, the model in this analysis showed that the accuracy and false positive rate are 95.4 percent and 3.7 percent, respectively, which are high enough to detect vacant houses. However, the true positive rate is 77.0 percent. Although the rate is not low to some extent, selection of features and further collection of extra samples may improve the rate. Geographic distribution of vacant houses further enabled us to check the difference between the actual and estimated number of vacant houses, and more than 80 percent of 500-meter grid data are with below 10 errors, which we think, provides city planners with informative data to roughly grasp geographical tendencies.

1. INTRODUCTION

With the advent of population decline in Japan, an increasing number of vacant dwellings has been emerging. According to the Housing and Land Survey conducted by the Ministry of Internal Affairs and Communications, approximately 13.6 percent of total housing units are vacant across Japan. The geographic tendency of vacant dwellings varies between urban and suburban areas, which makes the decay of depopulated suburban areas distinctive (Baba, Asami, 2017). The existence of vacant houses impedes the efficient use of available land; and once houses are dilapidated, they sometimes give a negative impact on adjacent neighbours (Han, 2014).

Cities worldwide, therefore, have concerned about the existence of vacant dwellings, and in particular depopulated cities carry out policy measures. Detroit, USA, is exemplified as active user of vacant properties that employ a land bank (Alexander, 2005). Eastern Germany implemented physical reconstruction of the city form, which contributed to the reduction in vacant housing stock (Bernt, 2019; Radzinski, 2016). Despite active policy measures to prevent housing vacancy, the number of vacant houses is likely to increase as long as cities are shrinking. We, therefore, need an efficient way to conduct a survey that detects vacant houses.

A frontier approach in dealing with the problem is to use open public data (Bourne, 2019; Cheshire et al., 2018). Spontaneous detection of vacant houses enables municipalities to implement

efficient plans and guidelines in dealing with the problem. For example, Bourne (2019) estimated the extent and value of low-use domestic properties using publicly available data. Although researchers using open data afford to collect and update them, the data sometimes fail to estimate the model accurately due to too many aggregate units that reflect various characteristics such as building, household and so forth.

To improve the accuracy of vacant housing detection, researchers draw attention to take advantage of closed municipal information such as resident and building registrations. Akiyama et al. (2020) took advantage of closed municipal data and estimated the vacancy rate in Japanese cities. Nevertheless, there is a room for improving the accuracy of vacant housing detection, because the model simply sets the thresholds of featured variables and takes an average vacancy rate per grid area.

This study proposes a vacant housing detection model using closed municipal data and considers accelerating the use of public data for promotion of smart cities. Employing a machine learning technique, this study ensures high predictive power for vacant housing detection, and the model enables us to handle complex municipal data, which include substantial missing data. Since the analysis in this study separates the test samples from the obtained ones, the model may contribute to the spatial extrapolation of vacant housing distribution.

* Corresponding author

2. PREVIOUS STUDIES AND TARGET AREA

2.1 Emergence of vacancy

Increase in vacant properties is, in many cases, triggered by urban shrinkage (Couch, Cocks, 2013; Hollander et al., 2018; Radzimski, 2016). Championed by Oswalt (2006), urban shrinkage has appeared to be the main concern in the field of urban planning and policy (Audirac, 2018; Großmann et al., 2013; Hoekstra et al., 2018). Urban shrinkage occurs mainly on cities in developed countries such as Germany (Bontje, 2004), the United States (Schilling, Logan, 2008) and Japan (Hattori et al., 2017). Most studies define population decline as a primary indicator for urban shrinkage (Blanco et al., 2009). Together with the increase in vacant properties, population change has a negative impact on downtown decline (Hollander et al., 2018).

The emergence of vacant properties is fundamentally determined by the excess in supply versus shrink in demand. Since even in populated cities competitive supply yields structural vacancy (Wheaton, 1990), cities with population decline accelerate the situation (Couch, Cocks, 2013). Some cities in Japan represent the shrinking ones that suffer from an increase in vacant houses due to long-term population decline (Mallach et al., 2017). Although the phenomenon is gradually proceeding, neglecting this issue may result in serious hollowing out of population in cities.

Deteriorated vacant properties might exert negative externalities on adjacent neighbourhoods (Baba, Hino, 2019; Whitaker, Fitzpatrick, 2016), and city planners have attempted to control the number of vacant properties (Dol et al., 2017; Radzimski, 2016). In the case of the United States, Detroit Land Bank Authority, an organization that promotes the active use of vacant properties, plays a prominent role in regulating vacancies (Alexander, 2005). In contrast, eastern Germany physically regulates the number of vacancies through demolition of deteriorated buildings (Nelle et al., 2017). Nevertheless, proper intervention on vacant housing regulation is difficult, because the emergence of vacant housing is not fully understood, and geographic distribution may be associated with negative externalities (Han, 2014). We, therefore, need to identify the geographic distribution of vacancies as a first step.

2.2 Detection of vacant houses

In response to the increasing demand for vacant housing detection, researchers try to detect vacancies using various data such as satellite images, GPS tracking data and public survey results. Du et al. (2018) examined census-level vacancy rate using night time satellite images. Bourne (2019) took advantage of public data by examining the value of low-use domestic properties. The above studies provide us the possibility of estimating vacancy rate per census tract level. However, housing vacancy is associated with various factors such as household condition, building characteristics, geographic constraints and so forth, which are difficult to obtain from open data sources.

Recent pieces of research have taken advantage of closed data such as mortgage deeds (Fisher et al., 2015) and building registration (Baba, Hino, 2019), which articulate the characteristics of vacant housing. Some significant data sources

include closed municipal information stored in all municipalities, which enable us to extract important attributes such as household age, year when a building is constructed and amount of water consumption. A frontier research employing closed municipal information is the study by Akiyama et al. (2020), who took advantage of resident registration, building registration and monthly amount of water consumption. While the method is a combination of cross tabulations, the authors indicated the possibility of actively using closed municipal data.

Building upon previous research, recent analyses have tried to include machine learning methods to improve predictive powers and infer causal effects (Athey, 2017). The methods range from health inspector allocation problem (Glaeser et al., 2016) to economic well-being prediction using mobile data (Blumenstock et al., 2015). The results obtained from such machine learning techniques help city planners implement policies smoothly via evidence-based policy making (Howlett, 2009). We try to use a machine learning method to detect the location of vacant houses utilizing closed municipal information. Although the aim is similar to that of several pieces of previous research, our method extends the traditional way by exploring a modern method and considers how results can be interpreted to advance a smart city framework.

2.3 A case study area: Wakayama city, Japan

The city of Wakayama, which is the case area for this study, is located in the outskirts of the Osaka metropolitan area and is embraced by mountains and an ocean bay (Figure 1¹). The population as of 2020 amounts to 355,514. Because of the suburban setting of the city, inhabitants in Wakayama move to the central districts of Osaka, resulting in population decline of approximately 15,000 inhabitants in this decade. The number of vacant houses as of 2013 accounted for 15.8 percent, which is 2.3 percent higher than the national average.

In response to the increasing number of vacancies, the city conducted a vacant housing field survey to understand the overall quantity and geographical distribution of the vacant houses. While the field survey brings city planners many informative cues to deal with the problem, conducting field surveys is basically an expensive and laborious task. It is true that city planners require precise locations of vacant houses in a specific term, but they also figure out the estimated locations in a timely manner. Therefore, not overly accurate but affordable ways of detecting vacant houses have been desired.

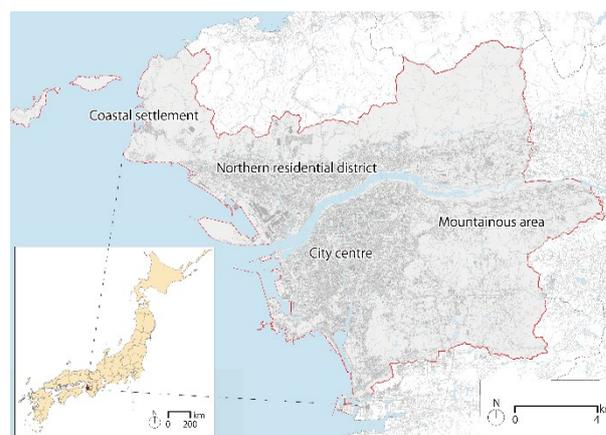


Figure 1. Geographic location of Wakayama city

¹The background map is created using digital residential map by Zenrin Co. LTD and Digital Map (Basic Geospatial Information) by Geospatial Information Authority of Japan.

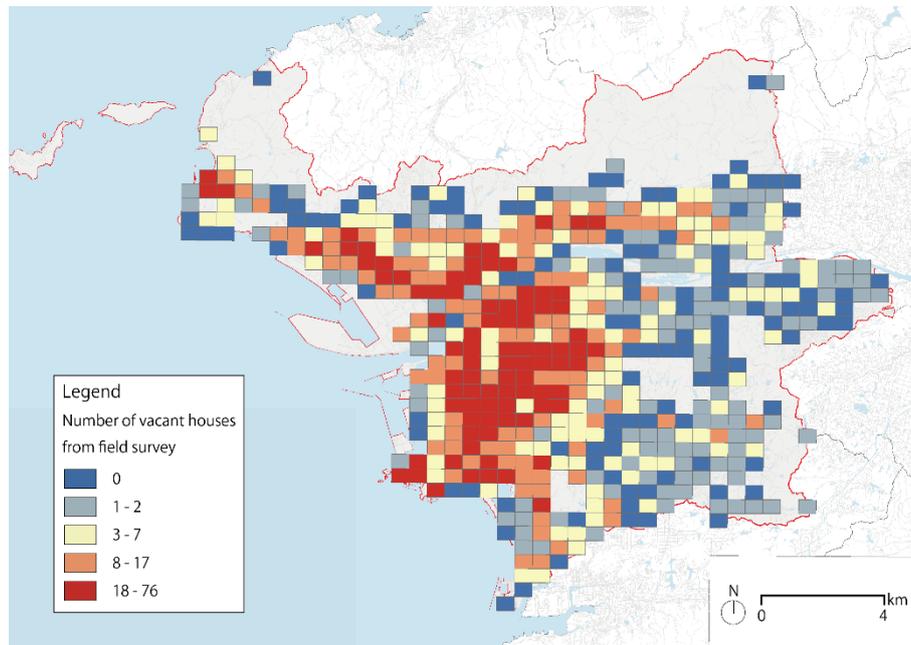


Figure 2. Geographic distribution of vacant houses per 500-meter grid cell

The number of vacant housing for a geographic location and the result of field survey from 2016 to 2017 are illustrated in Figure 2². The centre of the city is located along the bay area wherein a substantial number of vacant houses is observed. There is a residential district located in the north of the centre of the city in which the number of vacant houses tends to be high. A piece of coastal settlement exists along the coastline of the northern part of the city. Because of restricted land availability, the residential density of the settlement is high, and the number of vacant houses also indicates the same trend. The eastern part of the city is a mountainous area wherein small settlements are sparsely located. Although the number of vacant houses is not high in the mountainous area, the vacancy rate is still high because of low building density.

3. DATA AND METHOD

3.1 Closed municipal data

The closed municipal data that we utilised are resident registration, building registration and monthly water consumption information, which include pieces of information that are significant to improve the accuracy of the vacant housing detection model. Resident and building registration and water consumption information were retrieved as of April 2019, October 2018 and May 2019, respectively. The building registration data is half a year older than the other data. However, since changes in the features of building registration should be sluggish, the time lag of data collection does not affect analysis.

First, resident registration includes basic information such as residents who live in the jurisdiction. When a resident registers the place where they live, they fill out name, birthday, sex, family relationship and address. Since these pieces of information are of importance to jurisdiction, all of them are securely archived. Second, building registration particularly explains the

² Unlike the number of vacancies, vacancy rate expresses how vacant houses are distributed in an area. However, we assume that an increase in the density of vacant houses would have a

	RR	BR	WU	VH
(A) Number of original data	173,769	142,401	160,230	5,510
(B) Uniquely geocoded	141,719	96,451	122,035	5,510
Ratio of available data ((B)/(A))	81.6%	67.7%	76.2%	100.0%
(C) Available data per building	67,616	68,330	69,915	4,494
Ratio of data for analysis ((C)/(B))	47.7%	70.8%	57.3%	81.6%
$RR \setminus (BR \cup WU)$	3,229			
$BR \setminus (RR \cup WU)$	7,659			
$WU \setminus (RR \cup BR)$	5,160			
$RR \cap BR$	55,046			
$RR \cap WU$	55,414			
$BR \cap WU$	59,130			
$RR \cap BR \cap WU$	49,789			
$RR \cup BR \cup WU$	88,363			
(Total samples for analysis)				

Note: RR = Resident registration, BR = Building registration, WU = water use information, VH = Vacant housing field survey

Table 1. Summary of data processing

information on a building such as registration date, address, building use, structure, building age and floor area. This derives from a register that establishes the current condition and rights on a target property. In the case of Japan, buildings and lands are exclusively registered so we only use building registration. Third, monthly water consumption information literally indicates the amount of water consumed by a household. In addition to consumption, information on whether a hydrant is open or not is stored. In the case where water is drawn from a well, the amount of water consumed may be underestimated.

Data synthesis was based on a building unit extracted from a digital residential map by Zenrin Co. LTD. If any of the closed municipal data was absorbed to a building unit, we used it as a

negative effect on adjacent neighbours. We thus show the number of vacant houses to easily understand the agglomeration of vacancies.

Feature	Sample	Mean	S.D.	Min	Median	Max	
VH	Vacant house dummy	88,363	0.051	0.220	0	0	1
LU	Exclusive residential area dummy	88,363	0.114	0.318	0	0	1
	Residential area dummy	88,363	0.115	0.319	0	0	1
	Commercial area dummy	88,363	0.027	0.163	0	0	1
	Industrial area dummy	88,363	0.022	0.148	0	0	1
	Designated floor area ratio	24,613	203.427	58.444	50	200	600
RR	maximum age in a household	67,616	65.665	16.097	17	68	109
	minimum age in a household		42.559	26.438	0	43	108
	Number of inhabitants		2.932	1.769	1	3	19
	Resident registration dummy	88,363	0.765	0.424	0	1	1
BR	Type of housing dummy	68,330	0.973	0.163	0	1	1
	Wood structure dummy		0.712	0.453	0	1	1
	Steel structure dummy		0.161	0.367	0	0	1
	Reinforced concrete dummy		0.038	0.192	0	0	1
	Building age		28.790	14.432	0	30	94
	Floor area		110.983	47.669	10.07	104.49	999
	Floor number		1.909	0.360	1	2	6
WU	Average use of water in 2018	69,915	214.798	364.353	0	205	63373
	Hydrant is open		0.950	0.218	0	1	1
	Month of hydrant closed		3.753	23.394	0	0	349.5

Note: RR = Resident registration, BR = Building registration, WU = water use information, VH = Vacant housing field survey, LU = Land use regulation

Table 2. Descriptive statistics

sample data. When we added latitude and longitude to the closed municipal data, we used the CSV address matching service provided by Center for Spatial Information Science, the University of Tokyo³. Once we obtained the coordinates, we intersected them and the building polygons and specified the building in which the data is stored.

Table 1 illustrates how the closed municipal data are involved in the data set for analysis. From the original data source, approximately 70–90% of the data was geocoded. Out of the geocoded data, the ratio of the data for analysis ranged from 50% to 90%. Although this seems to be a low ratio for the used data, we only focus on detached housing and omit all apartments, factories, stores and so forth. Moreover, we conducted pre-processing to omit outliers so the number of inhabitants, building age and floor area ranged from 0 to 20, 0–100 and from 10 to 1,000, respectively.

To show the feature variables, descriptive statistics are summarised in Table 2. We further added land use zoning to the data set after the identification of urban, suburban and rural areas⁴. We roughly divided the land use zoning into 4 types: exclusively residential, residential, commercial and industrial zones. This category follows the Japanese City Planning Act and generally corresponds to the designated floor area ratio. We extracted the maximum and minimum ages in a household, number of inhabitants and a dummy of resident registration from the resident registration data. We included the dummy in resident registration because it signals that no residents are supposed to live in a house. Building registration is associated with structure dummies, building age, floor area and floor number. These data provide us with the durability and expected value of a building. Out of the water consumption information, we use the average consumption of water in 2018, a dummy whether a hydrant is open or not and the time length that the hydrant is closed.

In this study, all municipal data were modified in order to not specify individuals, so resident's names, birthdays and so forth are omitted from the data for analysis. Moreover, although we employ geospatial point data with coordinates which enable us to specify the any kinds of data resolution, we aggregate the attributes per 500-meter grid data, due to protection of personal information.

3.2 XGBoost: a machine learning method

We used eXtreme gradient boosting (XGBoost), which is developed by Chen and Guestrin (2016), for vacant housing detection. XGBoost is a type of ensemble learning method that generates many weak learners based on a decision tree, and node weights are readjusted through gradient boosting. Advantages of this method include high degree of accuracy and ability to deal with missing values although general performance is not high compared with other machine learning methods.

Let us denote the t -th decision tree as f_t , a simple form of error function is described as $\sum_{i=1}^l l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$, where x_i is the i -th feature and y_i is the i -th output. This means that the error function adds t -th decision tree $f_t(x_i)$ on the baseline of $(t - 1)$ -th decision trees, which contributes to the decrease in the error value of the function. However, the error function above may lead to an overfitting problem, so we need to modify the function by adding penalty terms. Error function $L^{(t)}$ at the t -th iteration is described as:

$$L^{(t)}(f_t) = \sum_{i=1}^l l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (1)$$

where γ and λ are the tuning parameters, T indicates the number of leaves in the tree, and w is a set of final prediction values. This means that the more the number of leaves and final prediction

³<http://newspat.csis.u-tokyo.ac.jp/geocode/>

⁴ Land use zoning information is retrieved from National Land Numerical Information: <https://nlftp.mlit.go.jp/ksj/>.

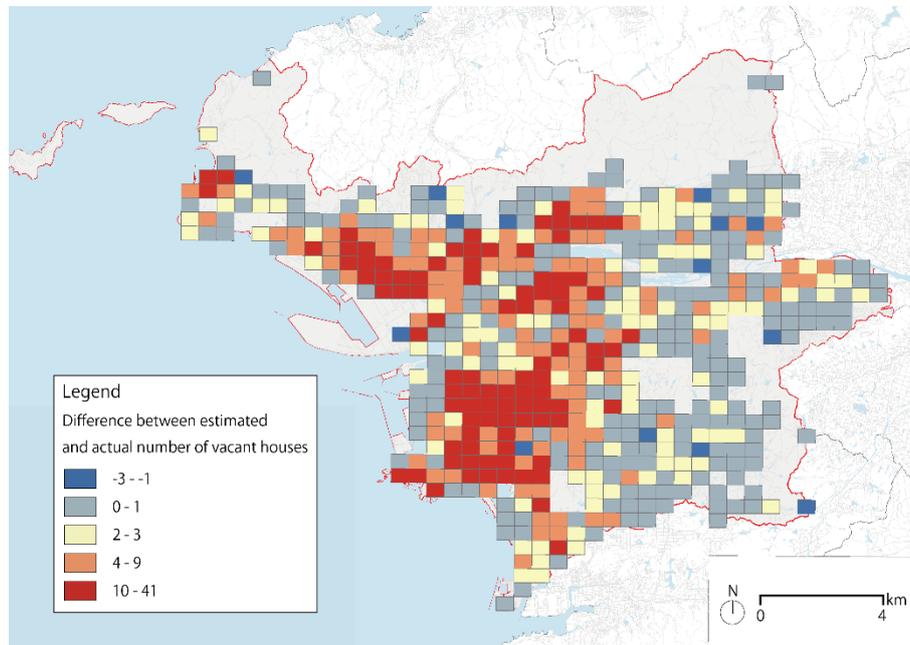


Figure 3. Geographical distribution of vacant houses subtracted from estimated ones per 500-meter grid cell

values increase, the more the value of the error function increases, which contributes to the prevention of overfitting.

In estimating $\sum_{i=1}^l l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i))$, XGBoost, since it is a kind of gradient booster, conducts a second-order approximation employing the Taylor series. As a consequence of the minimization of $L^{(t)}$, we obtain i -th optimal output value w_j^* .

XGBoost is able to deal with missing data for the following manner. It firstly computes the split of trees by ignoring the NA features, and then decides to allocate all the data with missing data so that the loss function is minimised. This algorithm is, therefore, beneficial when the missing data is not dispersed at random, because the allocation of the missing data is associated with reducing the value of the loss function. It is true that other methods such as Markov logic network are potentially able to handle missing data that are not at random, but XGBoost pursues high predictive power handling with missing data, which fits to our primary purpose.

Since XGBoost requires parameter tuning, we performed a grid search technique to identify the parameters. The maximum depth of a tree ranged from 2 to 8, the minimum child weight, which means the minimum sum of instance weight for the further tree partition step, was between 1 and 3, and the sub-sampling ratio of the training instances and columns were 0.5 and 1, respectively.

To determine the tuning parameters, we defined error ratio e_i by the following:

$$e_i = \frac{1}{N} \sum_{i=1}^l |y_i - [\hat{y}_i]| \quad (2)$$

where y_i is a vacancy dummy whether a building is vacant or not, \hat{y}_i is an estimated probability of vacancy, $|\cdot|$ indicates an absolute value in the parentheses, $[\cdot]$ is a function that returns 1 if the value is above 0.5 and otherwise 0, and N is the number of examples. Every time we changed the parameter set, a cross validation was conducted between the train and test data sets, and

validation was continued until the error ratio was converged. Consequently, we obtained the optimal values of the tuning parameter set.

4. RESULTS

We first conducted a grid search to determine the tuning parameters. We divided the analysing data into 70 percent train data and 30 percent test data. For each selected tuning parameter set, we calculated the error ratio and the optimal tuning parameter set is as follows: maximum depth = 6, minimum child weight = 2, sub-sampling ratio = 0.9 and sub-sampling of columns = 0.8. We also set the learning rate as equal to 0.01.

As a result of parameter tuning, we were able to check the model accuracy using the test data shown in Table 3. Overall, the accuracy rate was 95.4 percent, which we think is a fair number for predictive power. We also checked both true positive and false positive rates. True positive rate is the rate of a number of estimated vacancies that are also found as “vacant” to the total number of vacancies by field survey. In contrast, false positive rate is the rate of a number of estimated vacancies whose buildings are actually not vacant to the total number of “not vacant” buildings by field survey. These are the indicators that show the accuracy of the actual vacant housing detection and the extent of error. In this model, true positive rate is marked at 77.0 percent, while false positive rate was 3.7 percent. This means that the model could properly differentiate occupied dwellings from all other houses. Whereas it may be difficult for this model to perfectly detect an actual vacant house as a vacant one, the existence of confounding and hidden factors is assumed to make the improvement of the model more difficult.

We subsequently illustrated the geographic distribution of estimated vacant houses to check how spatial tendencies differ. We noticed that the estimated probabilities of vacant houses are attached to each building. We, however, visualised the number of vacancies per 500-meter grid because of protection of personal information. By our observation of Figure 3, if a geographical

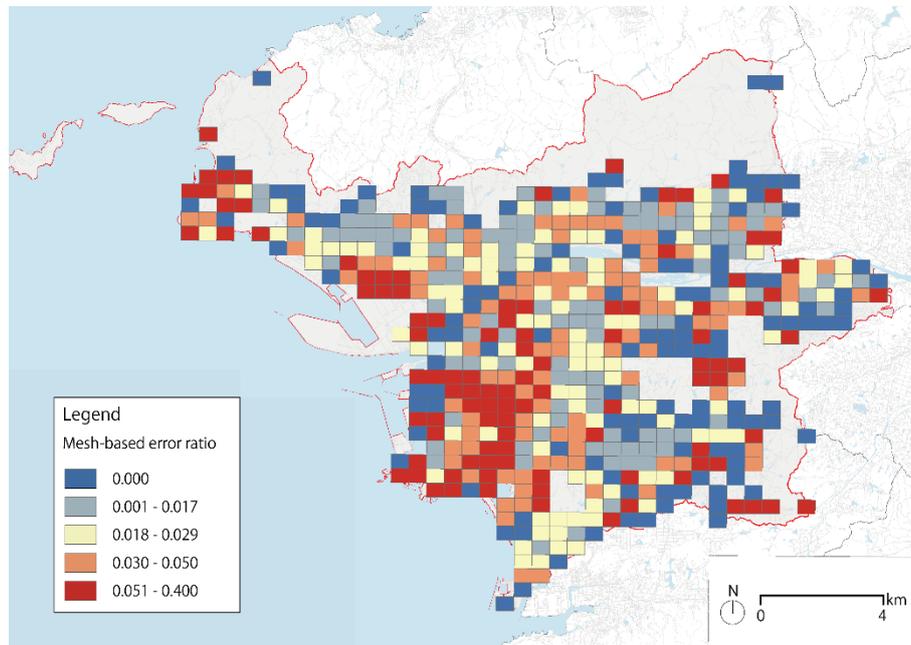


Figure 4. Geographical distribution of error ratio per 500-meter grid cell

Test data		Estimated number of vacant housing units		
		Yes	No	Total
Number of vacant housing units from field survey result	Yes	1,044	312	1,356
	No	918	24,235	25,153
	Total	1,962	24,547	26,509

Note: Accuracy rate = 95.4% (25,279/26,509); true positive rate = 77.0% (1,044/1,356); false positive rate = 3.7% (918/25,153)

Table 3. Result of cross validation

distribution of vacant houses is subtracted from the estimated ones, the number of vacant houses was likely to be higher in the estimated ones. This means that some of the occupied houses are incorrectly estimated as vacant ones. We, nevertheless, confirm that the estimation result provides us with both hot and cool spots of vacant houses. Since the values in Figure 3 are equally divided into 20 percentiles, more than 80 percent of the grid data are with below 10 errors. We, therefore, think that the estimated vacant housing distribution provides city planners with informative data to roughly grasp geographical tendencies.

To explore how the error is varied, we defined grid-based error ratio e_j by modifying equation [2]:

$$e_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |y_k - [\widehat{y}_k]| \quad (3)$$

where n_j is the number of housing units in the j -th grid cell. Figure 4 shows the distribution of error ratios per 500-meter grid cell. The high value of the error ratio indicated that the validity of the model in the grid cell was relatively low compared to others. Overall, since the thresholds were equally set at 20 percentile, the error ratios in the 80 percent of grid cells were below 5 percent, which confirmed that regardless of geographical differences, the accuracy of the model was high. According to

Figure 4, some areas in the city centre, the coastal settlement and the mountainous area are marked with high error ratios. Nevertheless, the difference between the actual and estimated number of vacant houses as shown in Figure 3 was relatively lower than the error ratio. It is conceivable that prediction accuracy is associated with the number of examples. It is exemplified that although the absolute error is one out of two examples, the error ratio is 0.5. One of the possible factors for the error would be the existence of farmhouses in the mountainous area, which are composed of distinctive features such as large lot size, wood structure, and so forth, making the prediction more difficult.

5. CONCLUDING REMARKS

Vacant housing detection is an urgent problem that needs to be addressed. It is a suitable example to promote utilisation of smart data that are stored in municipalities. We examined the estimation of geographical distribution of vacant houses employing the XGBoost technique. As a result of model estimation, we obtain the following findings:

We developed a model which enables us to handle missing data and non-linearity problems. Particularly, handling missing data is of importance for the practical use of closed municipal data because not all of the data are necessarily absorbed to a building. The XGBoost technique could solve these problems, since the method is non-parametric and able to consider missing values, employing a decision tree as a weak learner. However, the causal relationship between features and vacancy is unknown, and it needs to be addressed further.

Moreover, in the analysis, the model held an accuracy rate of 95.4 percent, which is high enough to detect vacant houses. The false positive rate was 3.7 percent, indicating that the model significantly detects houses in which residents live. However, the true positive rate was marked as 77.0 percent. Although the rate was not low to some extent, selection of features and further collection of extra samples may improve the true positive rate.

Geographic distribution of vacant houses further enabled us to check the difference between the actual and estimated number of vacant houses. Although the estimated number of vacant houses was likely to be higher than the actual ones, 80 percent of the grid cells keep the differences up to 10 and are below 5 percent of the error ratio.

For further improvement in promoting a smart city framework, we think of changing the ratio between training and test data sets, site specific training data extraction and extrapolation to other municipalities. In this study, we put 70 percent of the original data set as training data and the rest, which is 30 percent, as test data. While this study achieves high accuracy in vacant housing detection, it can be done by decreasing the ratio of training data. City officers can reduce the extent of field surveys if a small amount of supervised data is sufficient. Moreover, in case the model achieves high accuracy using the restricted area as training, field survey can be further simplified. By conducting a survey of specific areas in the city, we can extrapolate the geographic distribution of vacant houses. Lastly, once the model is built, we consider applying it to other municipalities. To confirm the extent of the model's prediction accuracy in other jurisdictions, a more generic model can be built, which would allow us to expand the model geographically. With the increasing use of closed municipal data, we expect that similar models will be developed for efficient execution of public works.

ACKNOWLEDGEMENTS

We are grateful to Wakayama city for provision of the original data set and the four anonymous referees for helpful comments. This work was supported by JSPS KAKENHI Grant Number JP20K14898.

REFERENCES

- Akiyama, Y., Ueda, A., Ouchi, K., Ito, N., Ono, Y., Takaoka, H., Hisadomi, K., 2020. Estimating the Spatial Distribution of Vacant Houses Using Public Municipal Data, in: Kyriakidis, P., Hadjimitsis, D., Skarlatos, D., Mansourian, A. (Eds.), *Geospatial Technologies for Local and Regional Development*. AGILE 2019. Lecture Notes in Geoinformation and Cartography. Springer, Cham, pp. 165–183. https://doi.org/10.1007/978-3-030-14745-7_10
- Alexander, F.S., 2005. *Land Bank Strategies for Renewing Urban Land*. J. Afford. Hous. Community Dev. Law 14, 140–169.
- Athey, S., 2017. Beyond prediction: Using big data for policy problems. *Science*. 355, 483–485. <https://doi.org/10.1126/science.aal4321>
- Audirac, I., 2018. Introduction: Shrinking Cities from marginal to mainstream: Views from North America and Europe. *Cities* 75, 1–5. <https://doi.org/10.1016/j.cities.2017.10.012>
- Baba, H., Asami, Y., 2017. Regional Differences in the Socio-economic and Built-environment Factors of Vacant House Ratio as a Key Indicator for Spatial Urban Shrinkage. *Urban Reg. Plan. Rev.* 4, 251–267. <https://doi.org/10.14398/urpr.4.251>
- Baba, H., Hino, K., 2019. Factors and tendencies of housing abandonment: An analysis of a survey of vacant houses in Kawaguchi City, Saitama. *Japan Archit. Rev.* 2, 367–375. <https://doi.org/10.1002/2475-8876.12088>
- Bernt, M., 2019. The emergence of “Stadtumbau Ost.” *Urban Geogr.* 40, 174–191. <https://doi.org/10.1080/02723638.2017.1332926>
- Blanco, H., Alberti, M., Olshansky, R., Chang, S., Wheeler, S.M., Randolph, J., London, J.B., Hollander, J.B., Pallagst, K.M., Schwarz, T., Popper, F.J., Parnell, S., Pieterse, E., Watson, V., 2009. Shaken, shrinking, hot, impoverished and informal: Emerging research agendas in planning. *Prog. Plann.* 72, 195–250. <https://doi.org/10.1016/j.progress.2009.09.001>
- Blumenstock, J., Cadamuro, G., On, R., 2015. Predicting poverty and wealth from mobile phone metadata. *Science*. 350, 1073–1076. <https://doi.org/10.1126/science.aac4420>
- Bontje, M., 2004. Facing the challenge of shrinking cities in East Germany: The case of Leipzig. *GeoJournal* 61, 13–21. <https://doi.org/10.1007/s10708-005-0843-2>
- Bourne, J., 2019. Empty homes: mapping the extent and value of low-use domestic property in England and Wales. *Palgrave Commun.* 5, 1–14. <https://doi.org/10.1057/s41599-019-0216-y>
- Chen, T., Guestrin, C., 2016. XGBoost, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. ACM Press, New York, New York, USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cheshire, P., Hilber, C.A.L., Koster, H.R.A., 2018. Empty homes, longer commutes: The unintended consequences of more restrictive local planning. *J. Public Econ.* 158, 126–151. <https://doi.org/10.1016/j.jpubecon.2017.12.006>
- Couch, C., Cocks, M., 2013. Housing Vacancy and the Shrinking City: Trends and Policies in the UK and the City of Liverpool. *Hous. Stud.* 28, 499–519. <https://doi.org/10.1080/02673037.2013.760029>
- Dol, K., Mazo, E.C., Llop, N.L., Hoekstra, J., Fuentes, G.C., Etxarri, A.E., 2017. Regionalization of housing policies? An exploratory study of Andalusia, Catalonia and the Basque Country. *J. Hous. Built Environ.* 32, 581–598. <https://doi.org/10.1007/s10901-016-9528-z>
- Du, M., Wang, L., Zou, S., Shi, C., 2018. Modeling the census tract level housing vacancy rate with the Jilin1-03 satellite and other geospatial data. *Remote Sens.* 10. <https://doi.org/10.3390/rs10121920>
- Fisher, L.M., Lambie-Hanson, L., Willen, P., 2015. The role of proximity in foreclosure externalities: Evidence from condominiums. *Am. Econ. J. Econ. Policy* 7, 119–140. <https://doi.org/10.1257/pol.20130102>
- Glaeser, E.L., Hillis, A., Kominers, S.D., Luca, M., 2016. Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy. *Am. Econ. Rev.* 106, 114–118. <https://doi.org/10.1257/aer.p20161027>
- Großmann, K., Bontje, M., Haase, A., Mykhnenko, V., 2013. Shrinking cities: Notes for the further research agenda. *Cities* 35, 221–225. <https://doi.org/10.1016/j.cities.2013.07.007>
- Han, H.S., 2014. The Impact of Abandoned Properties on Nearby Property Values. *Hous. Policy Debate.* <https://doi.org/10.1080/10511482.2013.832350>

Hattori, K., Kaido, K., Matsuyuki, M., 2017. The development of urban shrinkage discourse and policy response in Japan. *Cities* 69, 124–132. <https://doi.org/10.1016/j.cities.2017.02.011>

Hoekstra, M.S., Hochstenbach, C., Bontje, M.A., Musterd, S., 2018. Shrinkage and housing inequality: Policy responses to population decline and class change. *J. Urban Aff.* 00, 1–18. <https://doi.org/10.1080/07352166.2018.1457407>

Hollander, J.B., Hartt, M.D., Wiley, A., Vavra, S., 2018. Vacancy in shrinking downtowns: a comparative study of Québec, Ontario, and New England. *J. Hous. Built Environ.* 33, 591–613. <https://doi.org/10.1007/s10901-017-9587-9>

Howlett, M., 2009. Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Can. Public Adm.* 52, 153–175. https://doi.org/10.1111/j.1754-7121.2009.00070_1.x

Mallach, A., Haase, A., Hattori, K., 2017. The shrinking city in comparative perspective: Contrasting dynamics and responses to urban shrinkage. *Cities* 69, 102–108. <https://doi.org/10.1016/j.cities.2016.09.008>

Nelle, A., Großmann, K., Haase, D., Kabisch, S., Rink, D., Wolff, M., 2017. Urban shrinkage in Germany: An entangled web of conditions, debates and policies. *Cities* 69, 116–123. <https://doi.org/10.1016/j.cities.2017.02.006>

Oswalt, P., 2006. *Shrinking cities, Volume 1, International research*. Hatje Cantz Verlag, Ostfildern-Ruit, Germany.

Radzimski, A., 2016. Changing policy responses to shrinkage: The case of dealing with housing vacancies in Eastern Germany. *Cities* 50, 197–205. <https://doi.org/10.1016/j.cities.2015.10.005>

Schilling, J., Logan, J., 2008. Greening the rust belt: A green infrastructure model for right sizing America's shrinking cities. *J. Am. Plan. Assoc.* 74, 451–466. <https://doi.org/10.1080/01944360802354956>

Wheaton, W.C., 1990. Vacancy, Search, and Prices in a Housing Market Matching Model. *J. Polit. Econ.* 98, 1270–1292. <https://doi.org/10.1086/261734>

Whitaker, S., Fitzpatrick, T.J., 2016. LAND BANK 2.0: AN EMPIRICAL EVALUATION. *J. Reg. Sci.* 56, 156–175. <https://doi.org/10.1111/jors.12206>