DEVELOPMENT OF EXTRAPOLATION MODEL FOR ESTIMATION OF VACANT HOUSE DISTRIBUTION USING MUNICIPALLY OWNED DATA AND MACHINE LEARNING

Kento Tomita[1] and Yuki Akiyama*[2]
[1]Graduate School of Integrative Science and Engineering, Tokyo City University
1-28-1, Tamazutsumi, Setagaya, Tokyo 158-0087, Japan,
Email: g2281640@tcu.ac.jp
[2]Associate Professor, Faculty of Architecture and Urban Design, Tokyo City University
1-28-1, Tamazutsumi, Setagaya, Tokyo 158-0087, Japan,
Email: akiyamay@tcu.ac.jp

KEY WORDS: Vacant house, Municipal data, Machine learning, Extrapolation, Field survey

ABSTRACT: In recent years, the number of vacant houses in Japan has continued to increase nationwide, and understanding their distribution has become an important task for local governments. However, the main method for surveying the distribution of vacant houses is basically field survey, which re-quires a lot of time, labor, and budget. Therefore, there is a need for a quick and inexpensive method to determine the distribution of vacant houses using data held by local governments, such as water consumption information of each household. Many previous studies have conducted to understand and estimate the distribution of vacant houses using various statistical and spatial information. However, very few methods have been addressed with a view to social implementation and horizontal deployment. In this study, we developed a machine learning method to estimate the spatial distribution of vacant houses across an entire city, utilizing data held by municipalities. In addition, we extrapolated the model to other municipalities with similar geographical conditions, and evaluated the accuracy, advantages, and problems associated with the extrapolation. The results showed that our method has high accuracy in predicting non-vacant houses, and buildings predicted as non-vacant with a high degree of accuracy can be excluded from the field survey. It means that we can identify buildings that should be excluded from field surveys in advance, and the burden of conducting such surveys for vacant houses in municipalities will be significantly decreased. In addition, the model was able to correctly predict approximately 83% of the municipalities in which it was extrapolated, as well as the municipalities for which it was created, regarding vacant houses.

1. INTRODUCTION

In recent years, the number of vacant houses has continued to increase nationwide due to Japan's declining and aging population. According to the results of the Housing and Land Survey in 2018 by the Statistics Bureau of the Ministry of Internal Affairs and Communications, one of Japan's administrative agencies, the number of vacant houses in Japan reached approximately 8.5 million and the vacant house rate reached 13.6% in 2018 . It has been pointed out not only in Japan but also internationally that the increase in the number of inadequately managed vacant houses affects the attractiveness and vitality of the entire community through deterioration of the landscape and public safety, deterioration of sanitation, and the risk of collapse in the event of an earthquake or other disaster (H. Yoo et al., 2011). For example, Miyake et al. (2012) surveyed residents about their experiences of being inconvenienced by vacant houses and the nature of the inconvenience, and found that they caused concern to nearby residents, especially in terms of hygiene and public safety. Awazu (2014) found that the presence of a vacant house in a mismanaged condition, especially within 100 m to 300 m, significantly depresses land prices. Furthermore, outside Japan, Stacy (2017) found that reducing vacant houses in the United States reduces crime. Sun et al. (2013) indicated that in China, this distorts the relationship between supply and demand in the real estate market and increases housing prices.

Therefore, understanding the spatial distribution of vacant houses is one of the most important tasks for local governments in Japan. In May 2015, the "Act on Special Measures concerning Promotion of Measures for Vacant Houses" (hereinafter referred to as the "Act on Special Measures for Vacant Houses") came into effect. Under the Act, local governments throughout Japan are obligated to make efforts to conduct surveys on the distribution of vacant houses. However, the current method of surveying the distribution of vacant houses is mainly based on field surveys: visual inspection of the exterior, which is time-consuming, labor-intensive, and expensive. Therefore, it is important to grasp the distribution of vacant houses quickly and inexpensively by using data already possessed by local governments, existing statistics, and open data.

1.1 Literature review

According to Mashita and Akiyama (2020), studies related to vacant houses in Japan can be divided into two major categories: study on "survey methods" aimed at understanding the qualitative status quo and obtaining original information, and study on "analysis methods" aimed at analyzing quantitative information and clarifying the relationship between various pieces of information. The most frequently encountered "survey methods" for understanding the spatial distribution of vacant houses include the field surveys based on visual inspection of exteriors, as well as interviews and questionnaire surveys to understand the perspectives of stakeholders concerning vacant houses. Although these methods can reliably identify the distribution of vacant houses on each building, they are difficult to apply to a wide-area survey of an entire municipality because the survey targets only a very limited area. On the other hand, there has been an increasing number of studies in recent years on "analytical methods" for understanding and estimating the distribution status of vacant houses by utilizing various statistical and spatial information. For example, Kanamori et al. (2015 conducted a multiple regression analysis to predict vacant houses in each prefecture based on the surplus rate of housing derived from the Housing and Land Survey and Housing Starts Statistics, and regional characteristics such as population and household composition. However, this method is limited to the distribution of vacant houses at the prefectural and spatial macro level. Yamashita et al. (2015) used water supply shutoff data to identify the distribution of vacant houses over a wide area. However, the method defines all properties with "closed" or "inactive" water service as vacant houses, and it is pointed out that the basis for this definition is not clear and that it is difficult to identify vacant houses based on water service data alone.

Under these circumstances, studies have been conducted to develop a rapid and inexpensive method for estimating the distribution of vacant houses using data owned by both local governments and the private sector. For example, the authors of this paper used the results of field surveys for vacant house conducted by local governments as training data. By utilizing various data owned by the local governments from a period close to when field surveys of vacant houses were conducted (hereinafter referred to as "municipally owned data"), they constructed a machine learning model to estimate the probability of each building being vacant (Akiyama et al., 2021; Mizutani et al., 2022; Tomita et al., 2022). This made it possible to estimate the number of vacant houses and the percentage of vacant houses in each subregion, such as a 250-meter mesh or city block. Municipally owned data is often held in common by municipalities throughout Japan and is continuously updated and managed, so it is expected that methods using these data will help municipalities reduce the burden of field surveys of vacant houses. However, in some cases, it is impossible to build a model within a municipality because the municipality does not have the data from the vacant house field survey, which is the training data, or has not conducted a vacant house field survey in the past. This issue could be solved by extrapolating models from municipalities that are able to build models to municipalities that have difficulty building machine learning models.

1.2 Purpose of this study

In this study, we first applied and advanced the method we have developed so far (Tomita et al., 2022) and focused on municipalities that have conducted on-site vacant house surveys across the entire city. We constructed a database for estimating vacant house distribution by combining two types of municipally owned data - basic resident registries and water consumption data - with open data from national censuses, all of which are held by municipalities across Japan. Subsequently, using this database, we develop a method through machine learning to estimate the spatial distribution of non-vacant and vacant houses throughout the city. Finally, the developed model extrapolated to other municipalities with similar geographic conditions to verify the accuracy, benefits, and challenges of the extrapolation.

1.3 Study area

The study areas are the entire Maebashi city, Gunma prefecture, and Wakayama city, Wakayama prefecture (Figure 1). Each of the cities is a capital city of prefecture with a population of approximately 300,000 and is a typical Japanese regional city with diverse geographical characteristics, such as a central city including business and shopping districts, old and new residential areas, industrial areas, and rural areas. In this study, an estimation model was created in Maebashi city, where vacant house surveys are conducted in the entire area, and Wakayama city, where vacant house surveys are conducted only in some residences, was used as the extrapolation municipality for the model to be verified.

2. DEVELOPMENT OF A DATABASE OF VACANT HOUSES

In this study, we used data owned by local governments, building data owned by private companies and the national census which is open data in Japan to create variables that explain the probability of vacant houses. While municipally owned data contains information for each building, open data is often aggregated by subregion, such as city blocks and grids. Therefore, we first spatially integrated these data based on the location of each of them, and to create a "vacant house database" (hereinafter referred to as "VHD") for both municipalities, which is used to analyze data for estimating the distribution of vacant houses.

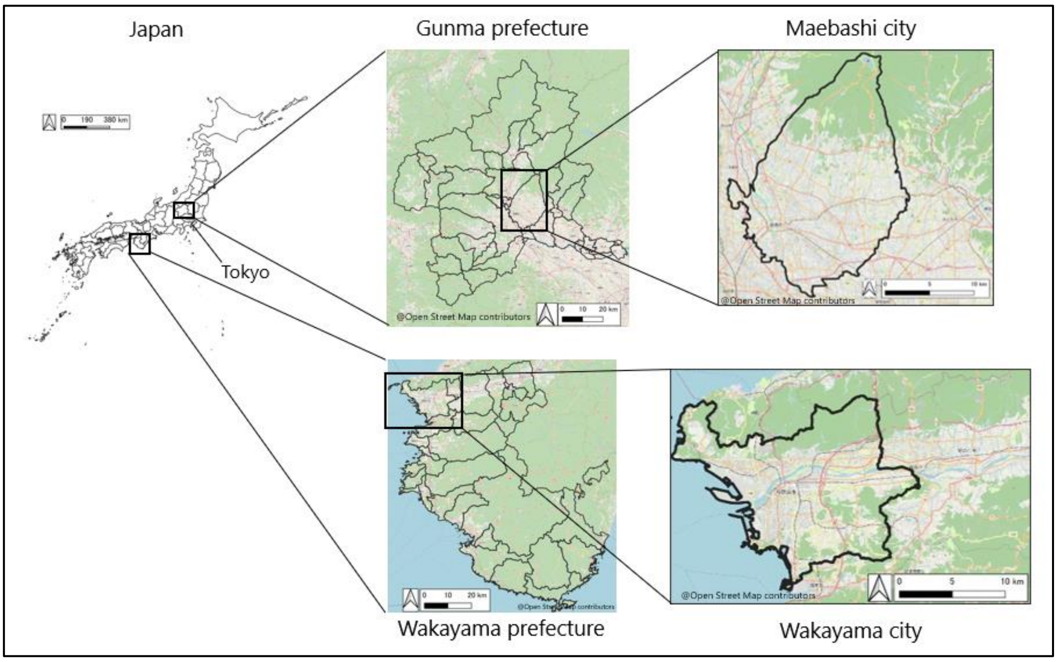

Figure 1. Location of Maebashi city and Wakayama city

2.1 Building Data (Commercial data)

Building data is the digital residential map in 2016 provided by Zenrin Co. Ltd. This map contains a building polygon data with attributes such as address, area, number of floors, and building use for each building. Because this study focused on detached houses, only building data for detached houses were used. Maebashi city had 91,445 detached houses out of 194,269 buildings. Similarly, Wakayama City had 107,286 detached houses out of 204,137 buildings.

2.2 Basic Resident Register (Municipally owned data)

The Basic Resident Register (hereinafter referred to as "BRR") is a database that contains the address, age, gender, etc. of each resident, and is used by local governments to perform various administrative processes related to residents. We used the BRR of all residents (337,595) as of March 31, 2017 in Maebashi city, and of all residents (367,415) as of March 31, 2019 in Wakayama city. Resident names and personal numbers have been removed in advance to protect personal information. The explanatory variables used in this study were the maximum and minimum age in the household, the number of people in the household, the number of people in the household by young population, working-age population, and old population, and the percentage of males in the household.

2.3 Water Consumption Data (Municipally owned data)

Water Consumption Data (hereinafter referred to as "WCD") contains water IDs, bi-monthly water consumption (5-year data from FY2014 to FY2019 for Maebashi city and 2-year data from FY2017 to FY2018 for Wakayama city) and addresses for all taps. In this study, it is necessary to keep the number of variables in the database used for the model and the database for extrapolation the same. Considering Wakayama city, which only has data for two years, water usage data for Maebashi city was taken from the fiscal year 2016 when the vacant house field survey was conducted, as well as the previous year, 2015. The maximum water usage for each year was aggregated and used as an explanatory variable for the prediction.

2.4 National Census (Open data)

The Population Census is the most important national statistical survey of all people and households living in Japan and is conducted every five years to clarify the actual status of the population and households in the country. The Ministry of Internal Affairs and Communications (MIC), one of Japan's administrative agencies, releases the results for all of Japan as open data. In this study, we used the Population Census aggregated by city block unit of Maebashi city and Wakayama city in 2015. The data contains the total number of each item and the total number of each category for each city block. In this study, the percentage of each category per total number of items was calculated and assigned as a characteristic of the area, which was used as an explanatory variable for the forecast. The total number of variables is 119, including basic tabulations of population, employment status, household structure, etc., and tabulations of population and employment

Table 1. List of explanatory variables used to estimate vacant house

Data	Variable	Number of variables
Water consumption data (WCD)	Maximum consumption for each year (2 years)	2
Basic resident register (BRR)	Number of people in household	7
	Maximum age of residents, Minimum age of residents	
	Number of household members by age (3 categories)	
	Percentage of males in the household	
National Census	Aggregate data on population, employment status, household structure, etc.	119
	Aggregation of population and employment status by place of work or school attendance, etc.	
Vacant house field survey	Evaluation (Vacant / Non-vacant)	1

status by place of employment and place of commuting to school.

2.5 Vacant house survey results (Municipally owned data)

This is the result of a citywide vacant house distribution field survey conducted by Maebashi city in 2016 for detached houses. The data includes the condition and location (longitude and latitude) of vacant houses. The total number of data is 6,101. The status of vacant houses also includes. It is the degree of damage ranging from houses in good condition that are marketable, to houses so damaged that they require removal. However, since the aim of this study is to estimate the spatial distribution of all vacant houses, including those in various states, the detailed condition of the vacant houses was not considered. Instead, two dummy variables, "vacant house" and "non-vacant house," were used as the target variables. On the other hand, the data from the vacant house distribution survey conducted by Wakayama city in 2017 reflects the results of the field survey after filtering for houses with zero water consumption. As in Maebashi city, the status and location of vacant houses are included. The total number of data is 3,607 for detached houses. The results of the survey on the distribution of vacant houses in Wakayama city were used to verify the accuracy of the extrapolation model.

2.6 Construction of VHD

Since the BRR and WCD contain address information, it converted to the location information (latitude and longitude) by address geocoding. Next, using GIS, (2.2) through (2.5) were integrated with the building data through a spatial integration process, and a VHD, which is data for analysis to estimate the distribution of vacant houses, was developed for both municipalities. At least one or more explanatory variables are required for the machine learning described below, and some municipally owned data may be missing for some buildings. BRR of Maebashi city was linked to about 78.1% (71,455) of the buildings, and WCD to about 77.9% (71,243). BRR of Wakayama city was linked to about 69.3% (74,349) of the buildings, and WCD to about 70.0% (75,196). Buildings that could not be linked to these data were treated as missing values. The results of the survey on vacant houses were excluded from the correct answers if they were linked to buildings that were not detached houses. As a result, about 65.6% (3,943) of the 6,011 data in Maebashi city were integrated with buildings, and about 58.4% (3,219) of the 5,510 data in Wakayama city were integrated with buildings. On the other hand, for the national census, it is possible to provide each information for all buildings. Table 1 shows the list of explanatory variables in VHD.

3. MACHINE LEARNING TO ESTIMATE THE PROBABILITY OF VACANT HOUSES

3.1 Machine learning model in this study

In this study, a decision tree-based machine learning model called XGBoost (eXtreme Gradient Boosting) was employed to estimate the estimated probability of vacant house per building ("vacant house rate"). XGBoost is a method proposed by Chen and Guestrin (2016) and is one of the methods called gradient boosting. This section describes in detail how "gradient boosting" works. First, a decision tree is created to improve the objective variable calculated from the objective variable and the predictions, the model is trained, and the predictions are calculated. Next, the error between the predictions and the objective variable is calculated, and a new decision tree is created and trained to fill in the error. This is repeated for a specified number of trees. As the trees are created, the weights of the decision trees created become smaller as the predictions of the model match the target variables. The sum of the weights of the leaves to which the data to be predicted belong in each decision tree is the predicted value (Figure 2). For example, if the highest age in the

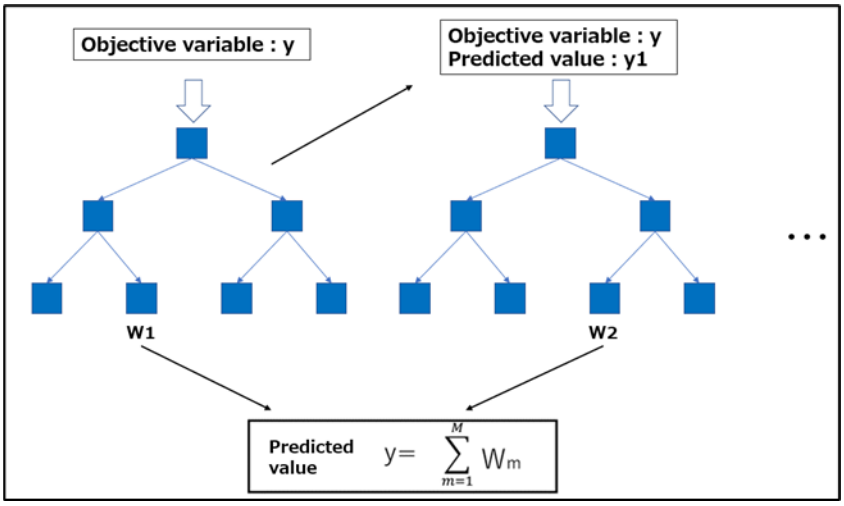

Figure 2. Calculation flow of predictive value by XGBoost

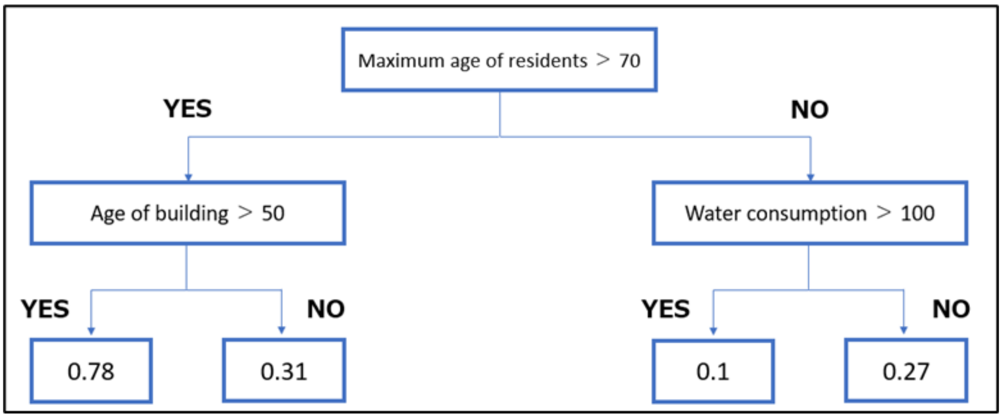

Figure 3. Image of decision tree in XGBoost

household is 75, the water consumption is 300 m3, and the age of the house is 40 years, an estimated vacant house rate of 0.31 is obtained from the tree structure in Figure 3. From there, decision trees are successively created, and the final predicted value, i.e., the estimated vacant house rate, is calculated by summing the results of each decision tree.
The reason for adopting this method is that it is more advantageous than other learning models in this study because of its high estimation accuracy and its ability to handle missing values. XGBoost has also been used to classify data containing missing values in previous studies in various fields, for example, in the analysis of gene expression data of cancer cells (Latief et al., 2020), online commerce behavior evaluation (Yang, 2021), and used house price prediction (Peng et al., 2019).

3.2 Construction of a machine learning model

In this study, an estimation model was constructed based on the BHD of Maebashi city. First, the original BHD of Maebashi city was divided into training data and test data. The reason for dividing the data is that if the entire data is used for training data, a trained model adapted only to the training data is created, and if the model is used to predict unknown data, the accuracy of the model will drop significantly. In this study, training and test data were split at a ratio of 8:2. The results of the survey of vacant houses in the correct response data are unbalanced data with a large difference in the amount of data with vacant house ratings (vacant houses) and without ratings (non-vacant houses). Therefore, under sampling was performed so that the ratio of non-vacant house: vacant house in the training data was 3:1. In addition, since XGBoost requires parameter tuning, tuning was performed using "Optuna (Akiba et al., 2019). The results after tuning are shown in Table 2. Finally, cross-validation was implemented and trained to calculate the estimated percentage of vacant houses per building.

3.3 Evaluation of the model accuracy

To verify the estimation accuracy of the learned model, we calculated the estimation results for a total of 16,707 test data items. The correctness rate (the percentage of correctly estimated vacant or non-vacant houses) was 0.8225 for the test data. Table 3 shows the estimation results for the test data. The proportion of actual non-vacant houses that were correctly

Table 2. Result of parameter tuning

Hyperparameter	value
max_depth	5
min_child_weight	3
gamma	2.1813
subsample	0.7670
colsample_bytree	0.7274
learning_rate	0.0687

Table 3. Verification results for 16,707 test data

		Estimated value	
		Non-vacant house	Vacant house
True value	Non-vacant house	14,452	3,050
	Vacant house	195	594

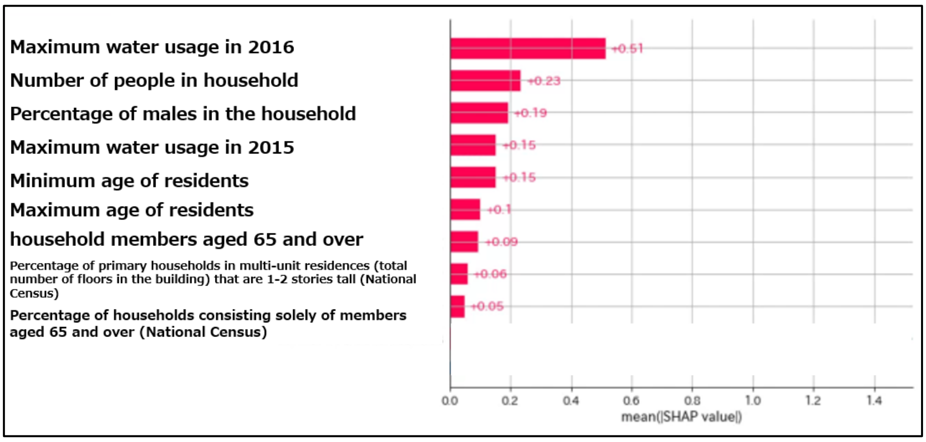

Figure 4. Importance rating of explanatory variables (top 9 variables)

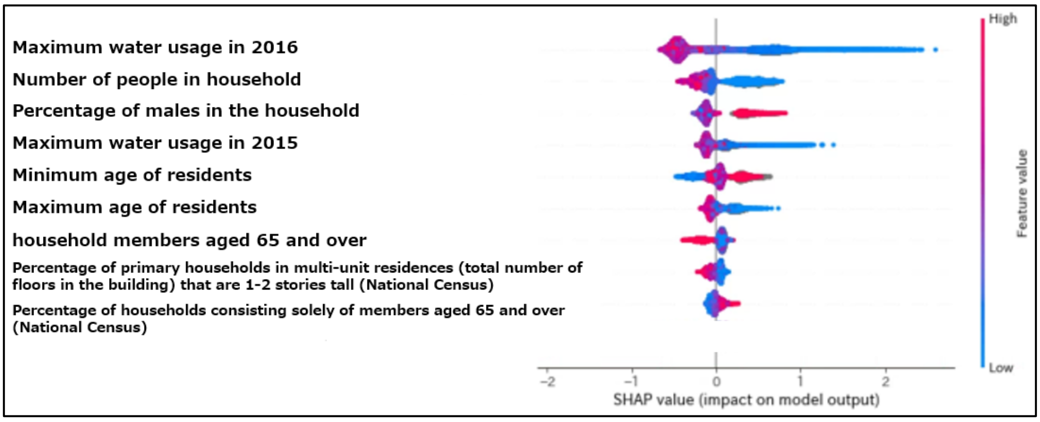

Figure 5. Beaswarm diagram per explanatory variable (top 9 importance items)

predicted as non-vacant (specificity) was 0.8257, and the proportion of actual vacant houses that were correctly predicted as vacant (recall) was 0.7528. Although these values were lower than those in the previous study by Tomita et al., 2022, due to the increase in sample data, the predictions were still made at a high level for both non-vacant and vacant houses. In particular, the ability to predict non-vacant houses with a high degree of accuracy will enable municipalities to reduce the number of non-vacant houses from the list of buildings to be surveyed when they conduct on-site inspections, which is expected to make a significant contribution to reducing the burden on municipalities in their vacant house survey work.

3.4 Feature Importance and Interpretation of the Constructed Model

For the constructed model, we conducted verification to understand the importance of the features and how each explanatory variable affects the predicted values. The method called "SHAP (SHapley Additive exPlana-tions)" was used for the above verification. This is an open-source library that applies the Shapley Value of cooperative game theory to machine learning and is one of the methods used to interpret machine learning models (Scott and Su-In., 2017). By using SHAP, it is possible to understand in detail not only the importance of each variable, but also the impact of each explanatory variable on the predictions derived by the model. In other words, it is a very good interpretation method in that it allows for both micro- and macro-interpretation in a consistent manner. Figure 4 shows a histogram of the nine explanatory variables in order of importance, and Figure 5 shows a beeswarm diagram for each of the top nine explanatory variables in terms of importance. A beeswarm diagram is used to determine the correlation between the SHAP value (the contribution of each explanatory variable) and the size of the explanatory variable values. In this study, a larger positive SHAP value indicates a greater probability of the house being vacant, while a larger negative SHAP value indicates a smaller probability of the house being vacant. The color of each point indicates the relationship between the size of the values of the explanatory variables: red indicates a larger value and blue indicates a smaller value. Gray represents buildings with missing values. For example, in the case of "maximum water consumption in each year," the lower the water consumption (blue distribution), the larger the SHAP value (positively pushing up the predicted value). In other words, the lower the water consumption, the higher the probability that the building is vacant.
Figure 4 shows that the importance of the municipality-owned data such as water consumption, age, number of household members, and number of household members aged 65 and over increased in the top 9 items, and this was especially true for water consumption close to the year of the field survey. Figure 5 shows that in addition to water consumption, the BRR tended to be more vacant in buildings with fewer household members and a higher male ratio, and in buildings with a higher minimum age of the household members. Furthermore, the gray scatterplot for the minimum age indicates missing data. In other words, buildings with missing BRR are more likely to be vacant houses.

4. EXTRAPOLATION OF THE MODEL

The probability of vacant house per building in Wakayama city was estimated by extrapolating the data of Wakayama city using the model constructed in Maebashi city. Table 4 shows the results for all 107,286 detached buildings in Wakayama city. The correct response rate was 0.6907, which was not as high as that of Maebashi, but with a reproducibility rate of 0.8341 (2685 out of 3219 cases), the Maebashi model predicted that most of the buildings found vacant in the Wakayama city field survey were vacant houses in the Maebashi city model. However, as mentioned in Section 2.5, it should be noted that the field survey of vacant houses conducted by Wakayama city did not cover the entire area, but only those houses with zero water consumption, and thus may have missed some vacant houses with non-zero water consumption. For example, if a tap in a water main in a vacant house is left slightly open, or if a water main is leaking due to aging, a situation could arise in which water consumption does not reach zero.
Furthermore, analysis of the data (FP) for 32,639 buildings that were predicted to be vacant but were not vacant revealed that 24516 buildings were missing both water and residential data. This is believed to be partly because, in this study, municipal data that could not be linked to buildings through address matching was analyzed as missing. Data that should have been included was excluded, thereby contributing to the reduction in prediction accuracy. Figure 6 shows the relationship between water consumption in 2017 and 2018 and the percentage of vacant houses for the 32,639 data points (FP). Among these buildings, there were a number of buildings with extremely low water consumption and missing basic resident records, as well as several buildings that had a certain amount of water consumption in 2017 but almost zero in 2018; the percentage of vacant houses was high in these buildings. Photographs of the exteriors of buildings that met

Table 4. As a result of extrapolating the Maebashi model, we estimated 102,876 buildings in Wakayama City.

		Estimated value	
		Non-vacant house	Vacant house
True value	Non-vacant house	71,428	32,639
	Vacant house	534	2,685

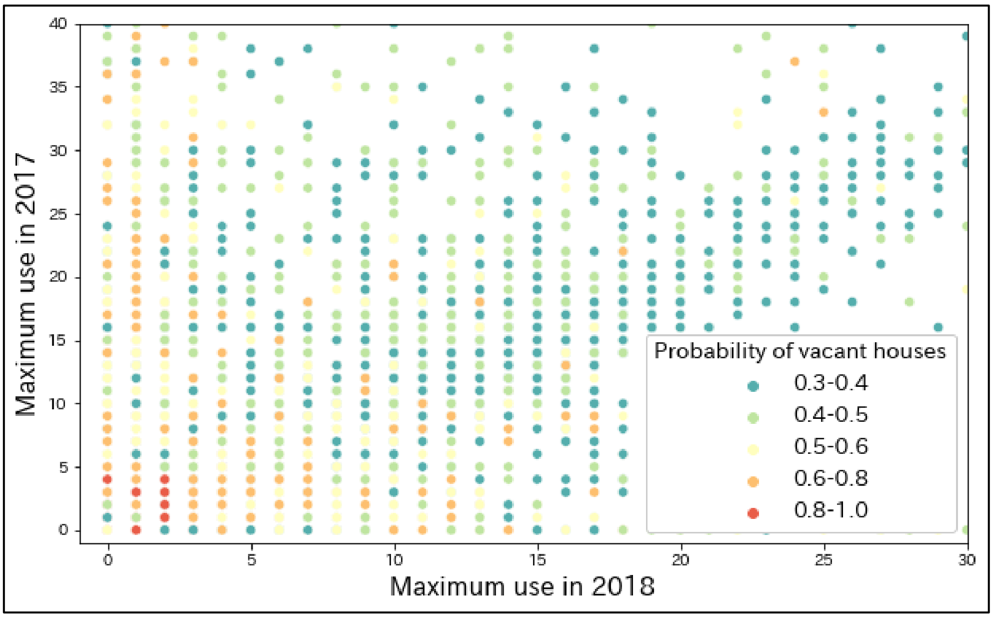

Figure 6. Relationship between annual fluctuations in water consumption and probability of vacant houses

Figure 7. Example of the exterior of a building with a high probability of vacant house that falls under FP

these conditions confirmed that they were buildings with signs advertising for tenants and damaged buildings with overgrown grass and trees that had not been cared for a long time (Figure-7). Therefore, buildings that are predicted to be vacant but are non-vacant houses are also worthy of investigation, especially if the estimated probability of vacant house is high.

5. CONCLUSION AND FUTURE RESEARCH

In this study, we developed a method for estimating the spatial distribution of vacant and non-vacant houses across an entire city by creating a vacant house database using mainly data owned by municipalities and implementing machine learning that uses actual vacant house distribution information as training data. The developed model was extrapolated to other municipalities with similar geographical conditions, and the extrapolated municipality's vacant houses were confirmed. As a result, it was confirmed that the spatial distribution of vacant and non-vacant houses could be estimated with relatively high accuracy not only in the municipality where the model was created but also in the extrapolated municipality. By visualizing the effects of variables on the model, it is possible to understand the relationship between regional characteristics and vacant houses. These indices are expected to contribute to urban development policies of local governments in Japan, such as urban and regional planning.

Future development will first improve the model used in this study and then extrapolate it to municipalities with different geographical conditions and population sizes to compare what similarities and differences can be found. Based on the estimation results of the model, we will verify the certainty of the estimated model by conducting field surveys. Furthermore, it is not always possible to pinpoint a building by address alone, as there are cases in which buildings in Japan have completely identical addresses, even if they are different buildings. This had a detrimental effect, and as pointed out in Chapter 4, some buildings could not be linked to municipal data by address matching. On the other hand, Japan is currently promoting a project called "Real Estate ID" that will grant a unique ID to each land and building based on the building and land registration information from 2022. Therefore, it is expected that by using this real estate ID to construct the VHD (Vacant House Database) in the future, the accuracy of linking between municipal datasets can be improved. Additionally, by using the new VHD constructed with the real estate ID to develop the model, we aim to improve the estimation accuracy and will conduct comparative evaluations with traditional methods, including differences in accuracy.

Acknowledgements:

This study was supported by the Digital Urban Spatial Information Research and Development Unit, Advanced Research Laboratories, Tokyo City University. This study was also conducted as part of the Joint Research Project (No. 880) on CSIS, the University of Tokyo. The municipality owned data: basic resident register, water consumption data, and results of field survey on vacant houses were provided us by the Future Policy Division of Maebashi city. In addition, the municipality owned data of Wakayama city were provided us by Wakayama city. The authors would like to express their gratitude to the following organizations for them.

References from Journals:

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., 2019, “Optuna: A Next-generation Hyperparameter Optimization Framework,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* 2623-2631.

Akiyama, Y., Baba, H., Ohno, S. and Takaoka, H.,2021, “Sophistication of monitoring method for spatial distribution of vacant houses using machine learning. (A study on the estimation method of spatial distribution of vacant houses using municipal public data (Part3)),” *J. Archit. Plann*., AIJ, Vol.86, No.786,2136-2146, doi: https://doi.org/10.3130/aija.86.2136.

Awazu, T.,2014, “The study related with the external effect and the effect of countermeasures for unmaintained vacant houses and buildings,” *Urban Housing Sciences,* Vol. 87, pp. 209-217, doi: https://doi.org/10.11531/uhs.2014.87_209.

Chen, T. and Guestrin, C., 2016: Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.

Kanamori, Y., Ariga, T. and Matuhashi, K., 2015, “Factor analysis and estimation of rate of vacant dwellings” *Journal of the City Planning Institute of Japan*, Vol.50 No.3, pp. 1017-1024, doi: https://doi.org/10.11361/journalcpij.50.1017.

Latief, M.A., Bustamam, A. and Siswantining, T., 2020 "Performance Evaluation XGBoost in Handling Missing Value on Classification of Hepatocellular Carcinoma Gene Expression Data," *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, pp.1-6, doi: 10.1109/ICICoS51170.2020.9299012.

Lundberg, SM. and Lee, S-I., 2017, “A Unified Approach to Interpreting Model Predictions,” P*roceedings of the 31st Interna-tional Conference on Neural Information Processing Systems*, 4768-4777.

Masuda, T. and Akiyama, Y.,2020, “*Recent trends in vacant house research in Japan,*” geospatial, Vol. 13-1, pp. 1-26, doi: https://doi.org/10.24586/jags.13.1_1.

Miyake, R., Koizumi, H. and Okata, J., 2012, “A basic study on the stable maintenance of vacant lands and houses in detached housing estates in a suburban area - A case study of housing estate in Sakura city, Chiba prefecture -,” *Journal of the City Planning Institute of Japan*, Vol.47 No.3, pp. 493-498, doi: https://doi.org/10.11361/journalcpij.47.493.

Mizutani, K., Tomita, K., Akiyama, Y., Baba, H. and Yachida, O.,2022, “Developing a Prediction Model for Vacant House Which Consider the Geographical Characteristics Around Each Building,” *Proceedings of the 31st Annual Conference on Geographic Information System Association*, C-4-4.

Peng, Z., Huang, Q., and Han, Y., 2019, "Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm," *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*, pp.168-172, doi: 10.1109/ICAIT.2019.8935894.

Stacy, C. P., 2017, “The effect of vacant building demolitions on crime under depopulation,” *J Regional Sci,* Vol. 58, Issue 1, pp.100-115, doi: org/10.1111/jors.12350.

Sun, WH., Xu, SC., and Li, LS., 2011, “Reducing Vacant Houses Is More Important than the Implementation of Green Buildings,” *Advanced Materials Research*, Vol. 280, pp. 250-254, doi: org/10.4028/www.scientific.net/amr.280.250.

Tomita, K., Mizutani, K., Akiyama, Y., Baba, H. and Yachida, O.,2022, Estimation of the Spatial Distribution for Vacant Houses Using Municipal Data and National Census - Case Study in Maebashi City, Gunma Prefecture -," *Proceedings of the 66th Annual Conference on Civil Engineering and Planning*, #7273.

Yamashita, S. and Yamashita, A.,2015, "A study on the pattern of vacant houses in a local core city," *Urban Planning Papers*, Vol. 50, No.3, pp. 932-937, doi: https://doi.org/10.11361/journalcpij.50.932.

Yang, Y., 2021,"Market Forecast using XGboost and Hyperparameters Optimized by TPE," *2021 IEEE International Confer-ence on Artificial Intelligence and Industrial Design (AIID)*, pp. 7-10, doi: 10.1109/AIID51893.2021.9456538.

Yoo, H., and Kwon Y., 2019. "Different Factors Affecting Vacant Housing According to Regional Characteristics in South Korea," Sustainability, Vol.11, Issue. 24, doi: 10.3390/su11246913.

References from Other Literature:

Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 2022. 2018 Housing and Land Survey Tabulation Results. Retrieved September 1, 2022, from https://www.mlit.go.jp/common/001314574.pdf.

Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 2022. Real Estate ID Rule Guidelines. Retrieved September 1, 2022, from https://www.mlit.go.jp/common/001314574.pdf.