Exploring household emission patterns and driving factors in Japan using machine learning methods

Peipei Chen[a], Yi Wu[b], Honglin Zhong[c], Yin Long[c,d*], Jing Meng[a]

[a] Bartlett School of Sustainable Construction, University College London, London WC1E 7HB, UK

[b] University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, UK

[c] Institute of Blue and Green Development, Shandong University, Weihai 264209, China

[d] Graduate School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan

Highlights:

- Multi-level household emission patterns are recognized by machine learning models.
- Demographic structure, age and appliances are key factors to vary emission levels.
- Customized emission reduction paths are proposed to guide energy-saving behaviours.
- Household appliance usage obtains the largest reduction potential to mitigate emissions.

Abstract: Given by the ambitious GHG mitigation targets set by governments worldwide, household is playing an increasingly important role for reaching listed reduction goals. Consequently, a deep understanding of its emission patterns and the corresponding driving factors are of great importance for exploring the untapped potential of household. However, how to accurately capture household emission features still demand further support from both data and method development. To bridge this knowledge gap, we try to use machine learning technology, which is well linked to the micro-level household survey data, to identify key determinants that could explain the household home-energy consumption and associated emissions. Here, we investigate the household CO_2 emissions based on a representative survey which covers 31,133 households in Japan. Six types of machine learning process are employed to find key factors determining to different household emission patterns. Results show that demographic structure, average age and electricity-intensive appliances (electric water heaters, electric heaters, etc.) are most significant driving factors that explain differences in household emissions. Results also further verified that differences in driving factors can be observed in identifying various household emission patterns. The results of study provide vital information for the customized decarbonization pathways for households, as well as discussing further energy-saving behaviours from data-oriented method.

Keywords: Household carbon emission; Factor analysis; Japan; Machine learning; Decarbonization

* Corresponding author: Dr. Yin Long; longyinutokyo@gmail.com

1. Introduction

Increasing concentrations of greenhouse gas (GHG) emissions have become a global concern as climatic issues keep rising [1,2], and governments worldwide have developed corresponding adapting or mitigating measures aiming at optimizing energy structure [3,4], introducing green technology [5,6], promoting decarbonized consumption behaviours [7,8], etc. Japan, as the world's third largest economy, released 1145 million tons of carbon dioxide (Mt CO_2) in 2016 [9], which is equal to the gross emissions of Australia, Brazil, Netherlands and the Philippines at the same year. To response, the Japanese government has submitted a series of ambitious reduction targets it has sought to actively promote [10]. For example, in 2016, Japan submitted the Intended Nationally Determined Contribution (NDC) and committed to reduce its GHG emissions by 26% from the level of 2013 by 2030 [11]. In addition, the Global Warming Countermeasures Plan in Japan also aims to reduce GHG emissions by 80% by 2050 to limit temperature increase to 1.5˚C as a long-term goal [12]. More recently, Japan committed to adopt a carbon neutral goal for 2050 and reach net-zero GHG emissions by the target year. Although all the sectors are encouraged to engage in emissions mitigation, the household sector is with the highest expectation, which is designed to reduce nearly 40% until 2030. This special focus on the household sector is not unique in Japan [13], but also worldwide. For example, US households account for 42% of total carbon emissions from fossil fuel combustion [14], and China's household sector energy-related CO_2 emissions were equivalent to the gross emissions of Brazil and Australia in 2016 [15,16], both of whom will need to make significant efforts in household sector to meet the national overall reduction targets.

Tackling emissions mitigation requires efforts to optimize energy consumption behaviours and find new ways to reduce carbon emissions in the household sector. An increasing number of studies have been conducted to analyse the driving factors of household emissions [17–22], including aspects of geographical location, demographic structure, household floor space, economics, etc. Field or experiment-based behavioural analysis was popularized after Daniel Kahneman won the 2002 Nobel Prize [23], especially behavioural interventions to energy conservation [24,25]. Additionally, subjective variables such as awareness and motivation [26], and institutionally relevant variables such as social norms have also been introduced into the analysis [27]. In more recent studies, regression analysis and decomposition analysis have been the main research methods.

In terms of regression analysis, Ordinary Least Squares (OLS) is the most widely applied model [28,29], but when fitted with too many variables, it will most likely be undermined by problems of overfitting and multicollinearity [30]. Besides, there is Logistic Regression (LR) method to find driving factors of household energy consumption [31]. As for decomposition analysis, such as Index Decomposition Analysis (IDA) [19,32,33], Structural Decomposition Analysis (SDA) [34] and IPAT model [35], they model mostly national or sub-national levels and cannot deal with household-level survey data in a bottom-up manner, because although households are randomly sampled and surveyed, it is still hard to do a bottom-up mapping from household emissions to the macro-level emissions. Moreover, factor analysis based on the decomposition method is usually summarized into changes of scale or intensity without granularity. As such, more detailed discussions about the influence of household characteristics on household emission patterns are needed. As mentioned, methods applied in existing factor research are regression and decomposition analysis [30,36]. However, due to data and methodological limitations, previous studies have rarely provided reliable recommendations for household emissions reduction from a data science perspective and have lacked policy insights for identifying and guiding high-emitting households. Here we use data science to attempt to analyse the Japanese micro questionnaire in a way that avoids the previous experience-based bias.

Survey questions for households can be fully converted into numerous characteristics describing a household's energy consumption and carbon emission patterns, however, as the number of household features increases and the survey data becomes sparse, they are not supported by traditional research methods. According to machine learning method, more information can be found, it overcomes the shortcomings of traditional research methods. For example, not all characteristics are significantly explainable and only a fraction of them can capture the differences in household emission patterns. In such cases, machine learning method can help us select key factors and identify the emission patterns at the household-level [28,37,38]. In order to infer the overall picture of household carbon emissions, machine learning methods are also useful to estimate based on the recognized patterns and rules from surveyed households when given strong driving factors. But, limited by the data quality, few households' energy survey support the basic data demand of machine learning, only limited interests tries have been made in household electricity consumption of Hong Kong [39], household transportation energy consumption in Delaware Valley region [40] and consumption-induced environmental impacts in Switzerland [37]. The conclusions derived from literature review are shown in **Table 1**.

Table 1

Overview of driving factors research method.

Method	Description	Specific method	Literature
Regression analysis	• The traditional method of doing factor analysis. • Suitable for studies with a small number of factors. • Tends to lead to overfitting and multicollinearity problems.	OLS	Meangbua et al [18] Yagita et al [21] Li et al [26] Zhang et al [28] Wang et al [29]
		LR	Fuks et al [31]
Decomposition analysis	• Suitable for macroeconomic data. • Inability to focus on intra-household characteristics.	IDA (LMDI, AMDI)	Ang [32] Shigetomi et al [19] Liu et al [41]
		SDA	Yuan et al [20] Meng et al [34]
		IPAT	Wang et al [35]
Machine learning	• Suitable for survey data with many factors. • High requirements for data quality.	Random Forest	Froemelt et al [37]
		Neural network	Guo et al [38]
		Support Vector Machine	Wang et al [39]
		Artificial neural network	Shams Amiri et al [40]

To bridge this knowledge gap, this study firstly investigates the importance of factors from aspects of housing conditions, economic income, demographic structure, household appliances, and energy use habits on household carbon emissions based on the household survey data in Japan. We use machine learning methods including LASSO, Decision Tree, Random Forest and XGBoost to precisely identify driving factors, which fit well with our multi-dimensional data. Results provide clues to answer three questions, i.e. 1) which household characteristics have a significant impact on carbon emissions per capita; 2) which household characteristics are the main driving factors of emission patterns according to energy use; 3) what are the differences in household characteristics under different emission patterns. The originality of this study lies in the analysis of Japanese household energy survey data through a machine learning approach, which enriches the research methodology on household energy consumption and carbon emissions. Besides, this study also provides valuable insights for long-term emission reductions from the household sector and specific suggestions for households with different emission modes in Japan, for customized energy supply management and services including new energy vehicles and photovoltaic promotion, which might also be valuable for the other countries that face the same issues such as household dilution and ageing.

2. Methodology and data

The survey includes data from demographic information, energy consumption to living habits of Japan in the year 2015, 2017 and 2018 [21,42], which portray a comprehensive picture of household energy consumption. Based on the survey data, this study starts with data reprocessing and conversion, e.g., data cleaning, removal of outliers and quantifying survey answers. In addition, household characteristics are selected and restructured by aspects of housing, economy, demography, transportation, and behaviour. For each aspect, carbon emissions are accounted using a bottom-up model. Then, we apply the LASSO (Least Absolute Shrinkage and Selection Operator) model for feature selection. Based on the selected features, we use various machine learning methods to estimate the importance of each factor, and according to the results, one optimal algorithm is applied to further explore the driving factors for household emission patterns. An overview of the methodology is shown in **Fig. 1**.

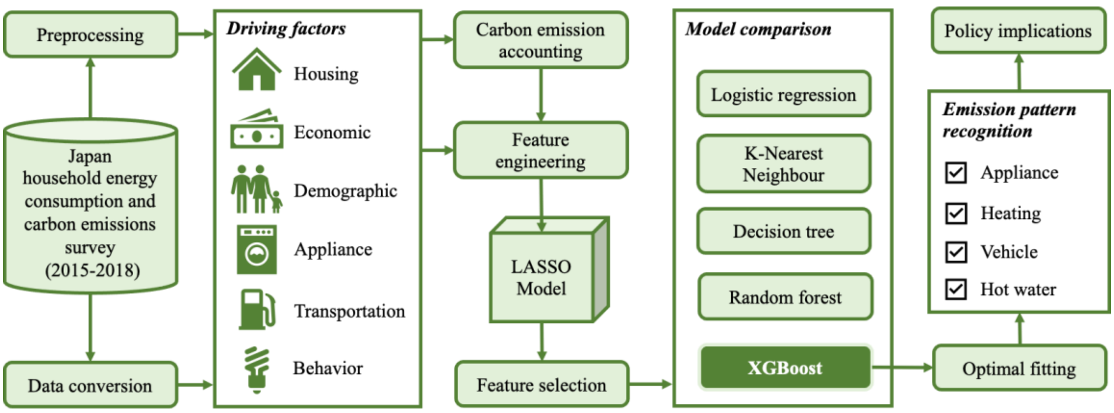


Fig.1. Research framework of household emission patterns and driving factors

2.1 Carbon emission released from household energy consumption

The dataset applied by this study adopts a bottom-up method to calculate household-level emissions, in which we measure and convert the energy consumed by each household for electric appliances, heating, transportation, and hot water supplying into CO_2 emissions [13,43,44]. Then, the gross emission is aggregated based on the subsets. Detailed electric appliances and vehicles therein are listed in the **Appendix Table 1**. Emissions of household-level appliances (TVs, refrigerators, washing machines, air conditioners, etc.) obtained from various appliances and time of use in the survey are quantified as followed:

$$CEA = \sum_a (PA_a \times TA_a \times e_a) \quad (1)$$

where CEA denotes the CO_2 emissions of appliances, PA_a denotes the unit power of appliance type a, TA_a is the service time of appliance a, and e_a denotes the electricity emission factor.

The emission intensities (see **Appendix Table 2**) in our equations and emissions correction methods are from Ministry of the Environment (MOE), Government of Japan, which can be directly driven from survey [40].

Here, household emissions from heating partly come from electric appliances (air conditioners, electric heaters, electric blankets, etc.) and the rest is generated by fuel-powered appliances using gas, kerosene and so on. Therefore, carbon emissions from heating are calculated by:

$$CEH = \sum_a (TH_a \times e_a) \tag{2}$$

where CEH denotes the CO_2 emissions of heating, a denotes the type of energy used for heating (electricity, city gas, liquefied natural gas (LNG), kerosene), TH_a denotes the consumption of energy a, and e_a denotes the emission factor of energy a.

Similarly, emissions from transportation and hot water for households are calculated by equation (3) and (4) where CEV and CEW denote the CO_2 emissions of vehicles and hot water, TV_a and TW_a are the energy consumption of vehicle fuels a (e.g., gasoline, light oil) and hot water (e.g., electricity, city gas, LNG, kerosene) respectively, and e_a denotes the emission factor of energy a.

$$CEV = \sum_a (TV_a \times e_a) \tag{3}$$

$$CEW = \sum_a (TW_a \times e_a) \tag{4}$$

2.2 LASSO-based feature selection

Linear regression models are widely used to estimate the effects of covariates on a given dependent variable. But for models with numerous covariates, classic linear regression models will easily lead to issues such as overfitting and multicollinearity [30,46]. Another issue that is often overlooked is data sparsity in household-level factor models, which refers to the fact that only a small fraction of independent variables plays significant roles in the model [30]. For instance, when many households' electric appliances are included and only frequently used appliances have sufficient values, then the data will be sparse, which means we need to filter out insignificant covariates. Therefore, we use the prevailing LASSO model to do the feature selection.

The LASSO model is specifically designed to screen for significant variables [42]. It is a compression estimation method that reduces the set of variables (in descending order). It can compress insignificant variables by allocating 0 to their coefficients through a penalty function for the purpose of variable selection.

The objective function of the LASSO model is a loss function with a penalty term, which optimizes the intercept and coefficients of covariates:

$$\frac{1}{2m}\left[\sum_{i=1}^{m}(y_i - w_0 - \sum_j^k w_j x_{ij})^2 + \lambda \sum_j^k |w_j|\right] \tag{5}$$

where m is the number of samples, k is the number of parameters, y_i is the dependent variable for household i, x_{ij} is the observation of covariate j for household i, w is the coefficient of independent variable. $\lambda \sum_{j=1}^{k} |w_j|$ is the penalty term in the form of L1-norm. λ is the penalty term weight that controls the number of covariates entering the model. We optimize the value of λ by iteratively grid-searching the parameter space from 0.001 to 0.1.

2.3 Household carbon emissions pattern recognition

Finally, all the households are classified into four types based on their emission by different energy usages, which are appliance-, heating-, vehicle- and hot water-oriented households. In this regard, we simplify the question by building four independent binary classification models corresponding to four emission modes. On top of that, we also estimate the importance of household characteristics in determining per capita emissions. Then, we compare the performance of various machine learning models using the same metrics to identify the key features that lead to the mode of household emissions pattern. Four models are applied as followed:

(1) Logistic Regression model

Since the household emission pattern recognition is binary question, we firstly use logistic regression model to detect the key features that make a household choose to use more energy in one of the orientations [43,44].

(2) KNN model

KNN (K-Nearest Neighbour) classifier is an analogy-based learning method that searches for the k training samples closest to the given test samples [45]. The KNN algorithm is easy to understand and to implement as there is only one free parameter for model tuning [46]. This method also requires a distance measure to assign classes based on samples' attributes where Euclidean distance is typically used [47].

(3) Tree-based model

Tree-based models are typical, but useful machine learning methods to incorporate nonlinearity in models. Classification tree models are well-known to be effective in dealing with categorical classification problems [48]. In our case, we start with Decision Tree (DT) model, Random Forest (RF), and then use eXtreme Gradient Boosting (XGBoost) model to identify key features.

Decision Tree is a supervised learning method where a tree represents a segmentation of samples created by a set of rules. The most common tree algorithms including chi-squared automatic interaction detection (CHAID), classification and regression trees (CART) [49], C4.5 [50], and C5.0 [51]. Hereby, we use CART and GINI index for criterion selection.

The main drawback of DT is that it tends to overfit [52]. To address it, Brieman (1996) proposed the bootstrap aggregating method, also called the "bagging" method. Random Forest is an algorithm derived from bagging, which uses dropout procedure to decorrelate regression trees and produce stable forecast [53,54]. Random Forest votes on the hundreds of decision trees generated, and finally the highest scoring classification is the final output. In addition, it is user-friendly because it has only two parameters (the number of variables in the random subset at each node and the number of trees in the forest) and is usually not too sensitive to their values [55], which makes it easy to apply in an empirical analysis.

XGBoost is a state-of-end-art tree boosting-based algorithm, and it is a gradient boosted tree (GBDT) model that integrates many tree models together to form a strong classifier [56,57]. Besides, XGBoost is a novel and sparsity-aware algorithm, which can address the data sparsity issue in our data. XGBoost tends to have better accuracy when making predictions compared to linear models. Feature importance gives an intuitive picture of the importance of features, but there is no way to tell how the features relate to the final prediction. We therefore use the SHapley Additive exPlanation (SHAP) value to solve this problem [58].

(4) Model comparison

Here, data are divided into the training and test set, which obtains 70% and 30% data, respectively. And models are firstly trained by the training set (70%) and then validated with the test set (30%). During the validation process, the following metrics are adopted: precision rate, recall rate, F1 score, and accuracy score. The precision rate (%) is the ratio of observations that have been predicted correctly [8], which is used for model comparison in this study.

3. Results

3.1 Japan's household carbon emissions

For better understanding the emission features of different population groups, we firstly aggregated households into groups by different socio-economic characteristics (i.e., income group, household type, average age, and region) (**Fig. 2**).

Firstly, the results show that higher income are observed with higher emission, which is consist with previous findings from China and Japan [38,64], as shown in **Fig 2a**. The carbon emissions per capita of households with an income of 0–1-million-yen are around 1.8 tons of carbon dioxide by per capita per year (tCO_2/cap/year), while this number rise nearly to 3.0 tCO_2/cap/year for 5-million-yen/year group. However, the per capita emissions and income of households are not completely linear. For example, the per capita emissions of 2–3-million-yen/year income groups are lower than those of 1–2-million-yen/year income groups. Therefore, apart from income, there are also other household factors that impacting household emissions. In addition, the household average age and per capita emissions show a linear relationship in **Fig 2b**, where the higher the average age, the higher emissions per capita. The highest emissions per capita in the age level 5 (60-75 years old) is also in line with the high emissions per capita of single elderly and married couples in the household type.

In terms of household type (**Fig 2c**), household are divided into seven types, i.e., single living alone, single with kids, single elderly (elder than 60 years old), big family (more than 3 people), couple, couple elderly and couple with kids. Among varied groups, single elderly has the highest emissions, which is close to 3.0 tCO_2/cap/year, almost twice as much as couples with kids' group. Japan's population is ageing severely and the number of elderly people living alone is increasing every year [60,61]. Elder population has higher demand for heating who use more lamp oil and kerosene, so their per capita carbon emissions are the highest. Second is the group non-early living alone (below 60 years old), with carbon emission per capita about 2.6 tCO_2/cap/year. Living alone means all the energy consumed, such as for heating, air conditioner, vehicle, are not shareable with anyone else, which results into higher energy use by per capita unit. In the contrary, larger families such as couples, families with kids, and big families (more than three people in a household) have lower carbon emissions per capita.

Regarding regions in **Fig 2d**, Hokkaido' household emission per capita are the highest, and Kinki's is the lowest, at half that of Hokkaido. Additionally, the emissions are higher in colder regions. And there are some areas where the population commutes more frequently, called commuter prefectures, such as Tohoku's Fukushima, Iwate and Aomori prefectures, which also have higher carbon emissions per capita than other prefectures. This also reflects social trends in Japan, where commuter prefectures have a greater need to control emissions in terms of transportation and population immigration [19].

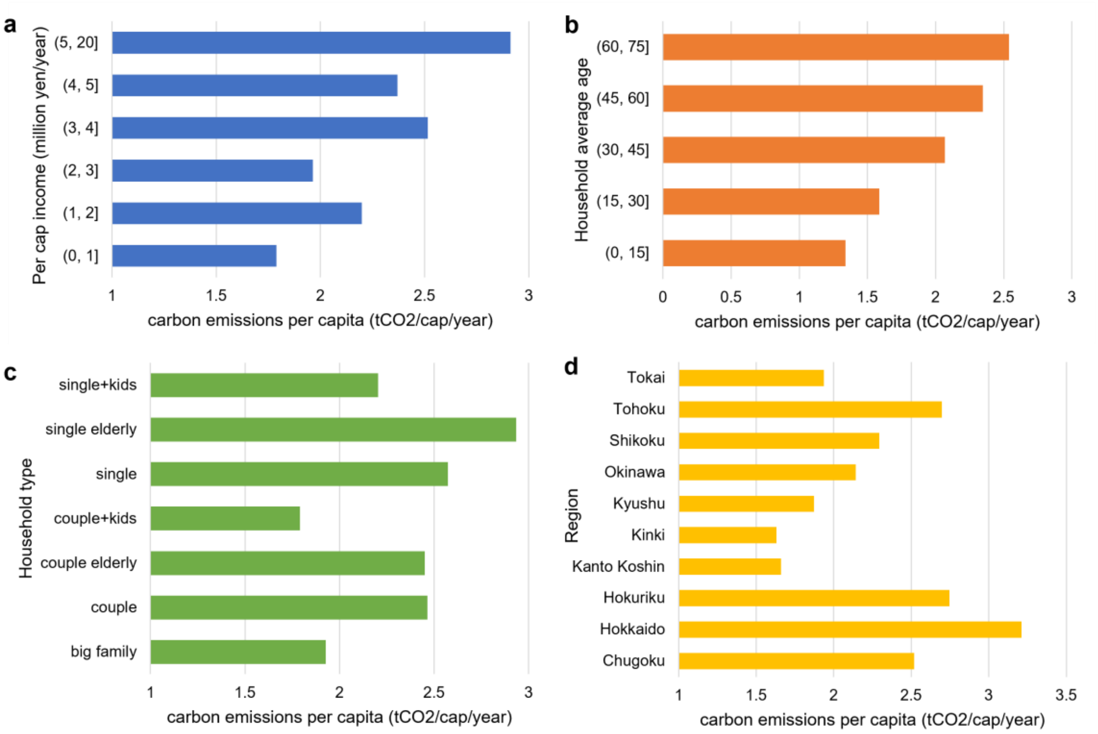


Fig. 2. Household carbon emissions from different household characteristics

3.2 Household carbon emissions driving factors

The driving factors and coefficients with high contribution to household carbon emissions per capita are obtained through LASSO model. Among them, household appliances and demographic structure have the highest contribution (**Table 2**).

In terms of household appliances, the contribution of electric water heaters, kerosene water heaters, heat-storage heater, and fridge are positive and very high. This indicates that the total emissions of households with these appliances will be higher than those without them, and that the more applicants there are, the higher carbon emissions per capita of the household. In addition, the use of some electric appliances can reduce the overall emissions of the household, such as solar water heater and stoves used woody fuel.

Within demographic structure groups, emissions per capita are generally lower in big families, such as those with many members and couples with children, while emissions per capita are higher in smaller households, such as single elderly, or single living alone. Among them, the single elderly in Japan has higher requirements for heating and hot water than other groups [62], so the carbon emissions per capita of that household will be higher.

From a transportation perspective, the number of vehicles and motorbikes, the frequency of driving have a significant positive effect on the carbon emissions per capita of a household, while the number of new energy vehicles (NEVs) has a negative effect. This also echoes Japan’s NEV policy [4,63]. In Japan, the market share of NEVs has been consistently high in recent years, where reducing vehicle emissions is a key part in the Japan roadmap for reducing emissions [64].

With respect to behaviour, the more heating devices are used per day, the more time is spent watching TV and the more days per week are spent at home during the day, the higher are the emissions per capita of a given household. In addition, the ratio of energy-saving behaviours has a negative effect on emissions per capita. Specifically, the more energy-saving behaviours, the lower the emissions per capita.

Regarding housing conditions, the more rooms with heating devices and city class, the higher the emissions per capita. This is since there are more rooms and corresponding appliances, especially heating devices. The higher the demand for energy consumption, the higher the carbon emissions. Households with double-layer window frames and glass have better warm/cool effects and lower carbon emissions per capita.

Table 2

Driving factors related to household carbon emissions through LASSO model.

Classification	**Variable**	**Coefficient**	**Classification**	**Variable**	**Coefficient**
Appliance	Electric water heater	0.696	Demographic	If single elderly	0.624
	Kerosene water heater / bath	0.326		If single	0.494
	Heat-storage heater	0.194		Number of children	0.053
	Fridge number	0.182		If couple	0.034
	Clothes dryer (gas)	0.087		Elderly number	-0.034
	Dishwasher	0.080		If couple with children	-0.225
	Stoves used woody fuel	-0.099		If big family	-0.251
	Solar water heater	-0.173	Transportation	Number of vehicles	0.200
Behaviour	Heating time	0.109		Driving frequency	0.114
	Hot water using time	0.059		Number of motorcycles	0.044
	TV using time	0.036		Number of renewable energy vehicles	-0.257
	Daytime at home	0.017	Housing	Heating room	0.040
	Saving rate	-0.004		City class	0.038
	AC using time	-0.007		Double-layer window	-0.048

3.3 Pattern recognition of emissions for different purposes

We classify household carbon emissions into appliance-driven emissions, heating-driven emissions, vehicle-driven emissions, and hot water-driven emissions based on energy use. Then we identify households whose emissions are mainly used for these purposes, and divide them into four patterns: appliance-, heating-, vehicle- and hot water-oriented households, accounting for 57%, 8%, 32% and 3% of the total number of households, respectively. Specifically, an appliance-oriented household means that the household has highest emissions in terms of appliances, i.e., household emissions are appliance-oriented.

Table 3

Accuracy comparison among different machine learning classifiers.

Model	Appliance-oriented (57%)	Heating-oriented (8%)	Vehicle-oriented (32%)	Hot water-oriented (3%)
Decision Tree	0.61	0.68	0.72	0.78
KNN	0.62	0.70	0.76	0.76

Logistic Regression	0.63	0.72	0.74	0.76
Random Forest	0.68	0.76	0.78	0.83
XGBoost	0.70	0.76	0.79	0.85

To identify the pattern features of these four types of households, this study used classifiers to calculate the importance of the feature selection factors, and the accuracy of the model, as shown in **Table 3**. In this study, six classification models were selected, and the accuracy of XGBoost was found to be the highest. Therefore, the following feature analysis is based on the classification results of XGBoost. The learning curves of XGBoost for the classification of the four household types are shown in **Appendix Figure 1**. It can be found that the accuracy of the model steadily increases on the test set and gradually decreases on the training set as the training sample increases, effectively proving that the model is not overfitted and has an accuracy rate of over 70% during the training process. Considering that few models have been trained on the emission characteristics of households for different purposes, our model is robust enough to identify a range of valid factors.

In terms of energy consumption, this study has represented the contribution of households' characteristics to the emissions per capita of different households (**Fig. 3**).

For appliance-driven emissions in **Fig. 3a**, variables such as the number of fridges, air conditioners, TV, and type of lighting are significantly associated with whether a household would be a potentially high-emitter. These appliances are either used frequently, or are high energy consumers, which is consistent with the findings of previous studies [38,70]. For heating-driven emissions, variables such as kerosene heater, service time of heaters, number of rooms with heaters, heat-storage heaters have high impact on emission patterns as shown in **Fig. 3b**. This is related to the duration of use of heating devices and the energy sources used. For example, electricity is more low carbon than kerosene or gas, and electric blankets take less time to use than heaters. For appliance- and heating-oriented household emission patterns, the differences in terms of driving factors are small.

For hot water-driven emissions in **Fig. 3c** and vehicle-driven emissions in **Fig. 3d**, the importance of the drivers varies widely. In terms of hot water, the number of electric water heater is the variable that ranks first in importance and is much higher than the other variables, which means that there is a greater probability that the household with an electric water heater is hot water-driven emission pattern. For vehicle, total driving distance is the most important variable, followed by driving frequency and number of vehicles. This suggests that households with a high number of vehicles and high use are likely to be vehicle-oriented emission households.

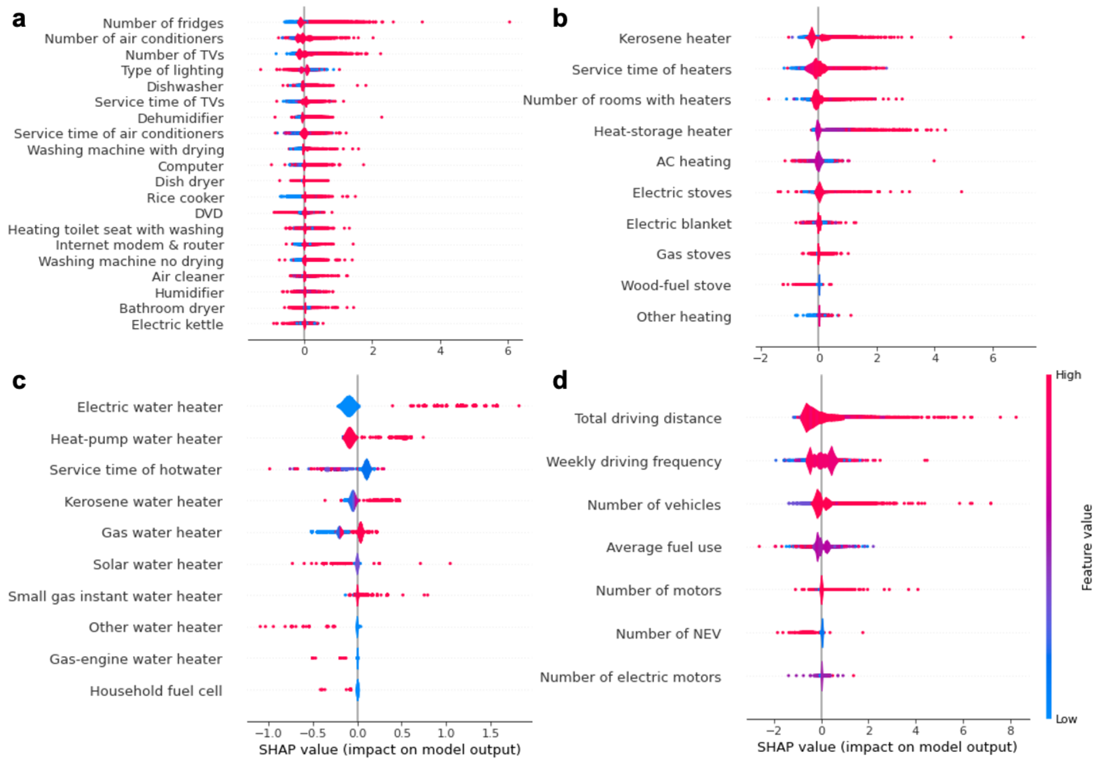


Fig.3. Importance of factors (SHAP values).

Caption: a: appliance-driven emissions, b: heating-driven emissions, c: hot water-driven emissions, d: vehicle-driven emission. Each row represents a feature, and the horizontal coordinate is the SHAP value, indicating the impact of that feature. A dot represents a sample, with redder colours indicating larger values for the feature itself, and bluer colours indicating smaller values for the feature itself.

As can be seen in **Fig. 4**, drivers that determine high and low emissions for different household uses are not consistent, and the differences between the factors are more pronounced. We have discussed the differences in the driving factors by categories as follows, and the differences of household characteristics are listed in **Appendix Table 3**.

(1) Appliance-oriented households

Among appliance-oriented emission households (**Fig.4a**), the most obvious differences were the number and use time of air conditioner. The next significant variables were energy-consuming appliances such as bathroom dryer, dish dryer, and heating toilet seat. The number of small household appliances such as Internet modem & router, electric clothes dryer, gas stove, and air cleaner did not show significant differences between the two groups. In general, appliance-driven emissions are always a larger piece of household emissions. To effectively reduce those emissions, we can consider expanding the use of low-power, green appliances on the one hand, and encouraging households to install household photovoltaics to use renewable energy to power these energy-consuming appliances on the other hand, ultimately reducing emissions.

(2) Heating-oriented households

Households with high heating emissions (**Fig.4b**) have a distinctive feature from other households in terms of the service time and number of heaters, while some households are still using traditional heaters fuelled by kerosene. It is worth noting that not many households use air conditioners for heating, and such households are often not the ones with the most emissions on the heating side. The choice and condition of appliances are related to the household's economic situation, but outdated heating methods, such as the use of kerosene for heating, pose safety risks in addition to generating more carbon emissions. More policies are needed to improve heating conditions for low-income residents, while helping to reduce their emissions.

(3) Vehicle-oriented households

Weekly driving frequency, total driving distance and the number of vehicles is key factors in determining whether a household is potentially a vehicle-oriented one (**Fig.4c**). Interestingly, although the penetration of NEVs is currently low (0.9% in Japan, 2019 [66]), households with low vehicle-driven emissions do not have significantly more NEVs than those with high vehicle-driven emissions. Therefore, efforts to replace fuel cars with NEVs need to be further strengthened so that NEVs can cover very high frequency and intensive vehicle use to reduce emissions from vehicles. Otherwise, even with NEVs, the transportation habits of these households will still result in larger vehicle-driven emissions, i.e., a rebound effect in energy consumption [67,68]. The existence and extent of the rebound effect, which diminishes the effectiveness of policies, is therefore one of the most important factors that must be considered in the implementation of energy policies [69].

(4) Hot water-driven emissions

The proportion of hot water-oriented households is small, accounting for only 3% of the total. Regarding their driving factors, the number of electric water heater has the biggest impact (**Fig.4d**). The more electric water heaters in a household, the more likely it is to produce greater emissions in terms of hot water supply. It is worth noting that households with gas water heaters do not produce significant emissions in terms of hot water supply. Therefore, improving equipment use and promoting more energy-efficient hot water supply methods will be the key means to achieve emissions reduction in the future.

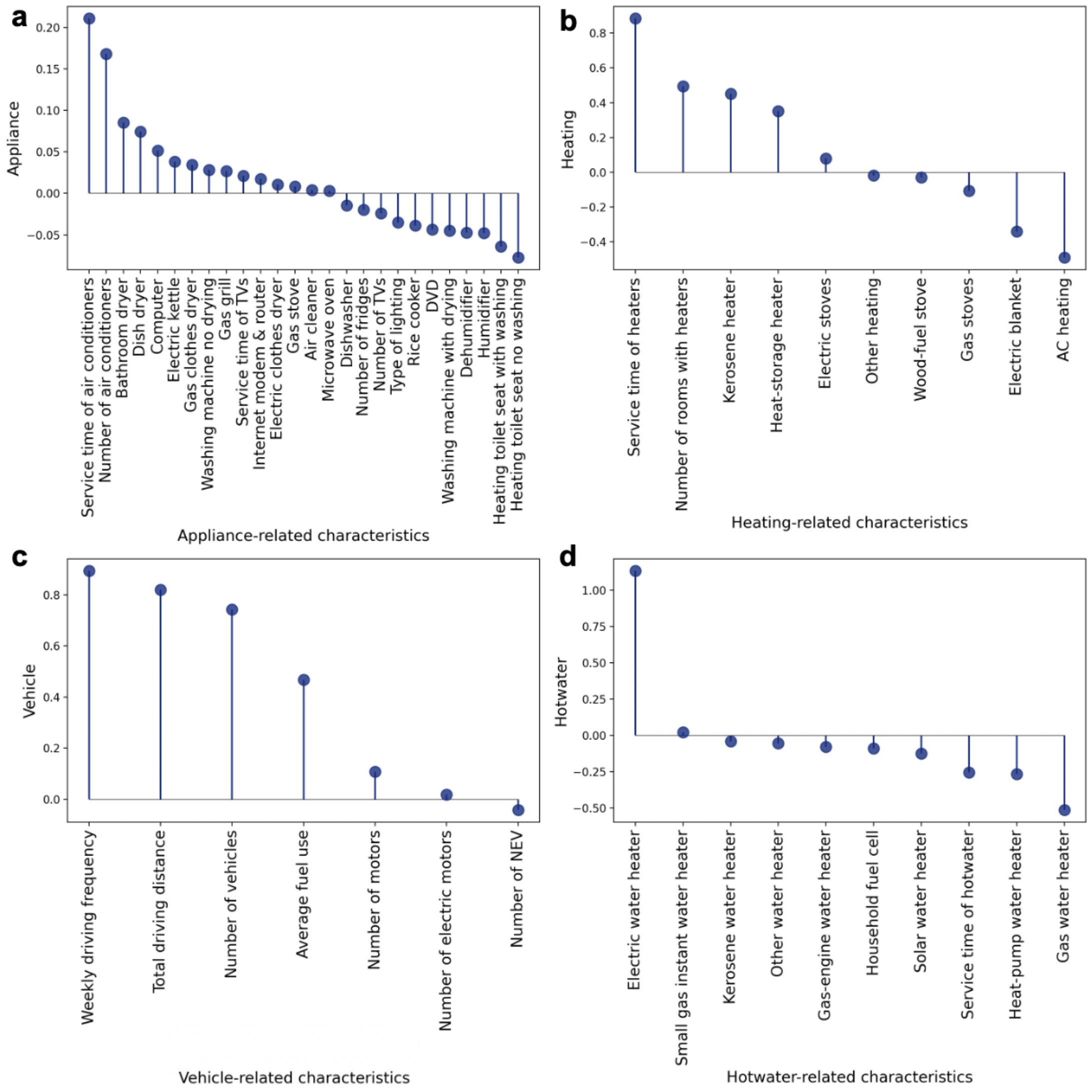


Fig. 4. Normalized differences in household characteristics across True/False groups for various emission patterns.

This study also describes the differences between common characteristics such as city class, housing conditions, and demographic structure, as shown in **Fig. 5**. The results show that the effects of the same variables on emissions for different uses are inconsistent. In terms of city class, the differences on appliance-, heating- and vehicle-oriented households are more significant, and households in larger cities are more likely to be appliance-oriented households, with heating- and vehicle-oriented not being the most dominant.

For housing conditions, significant and opposite differences are shown for appliance and heating-oriented households. The smaller the house, the more households tend to emit in terms of appliance use, while households with larger space and more rooms are more likely to be heating-oriented households.

There is no significant difference in income between households with appliance- and vehicle-driven emissions, but households with higher emissions of heating and hot water have significantly lower per capita income than other households. This indicates that the emissions of relatively low-income households are mainly related to maintaining their daily living.

About demographic characteristics, small households are more likely to be appliance-and heating-oriented emissions, and large families are more likely to be transportation- and hot water-oriented emissions. Households with elder people or those without children have greater needs for appliances and heating, a trend that is even being amplified over time, while households with fewer elder people or with children have greater needs for transportation. This significant difference is generally determined by the transportation needs of young people who need to commute daily and is amplified by the transportation needs of families with children. Further, different household types are analysed in depth for the mechanisms that determine their high and low emissions for different uses of the household. Of these, elderly groups (including single living one and couples) have higher electric and heating emissions. Since the use of electric appliances and heating are the basic needs of residents, they respond to the characteristics of the elderly's immediate need for emissions in these two areas. In addition, single residents have high emissions mainly in the use of appliances, while large families generate higher emissions in the supply of hot water.

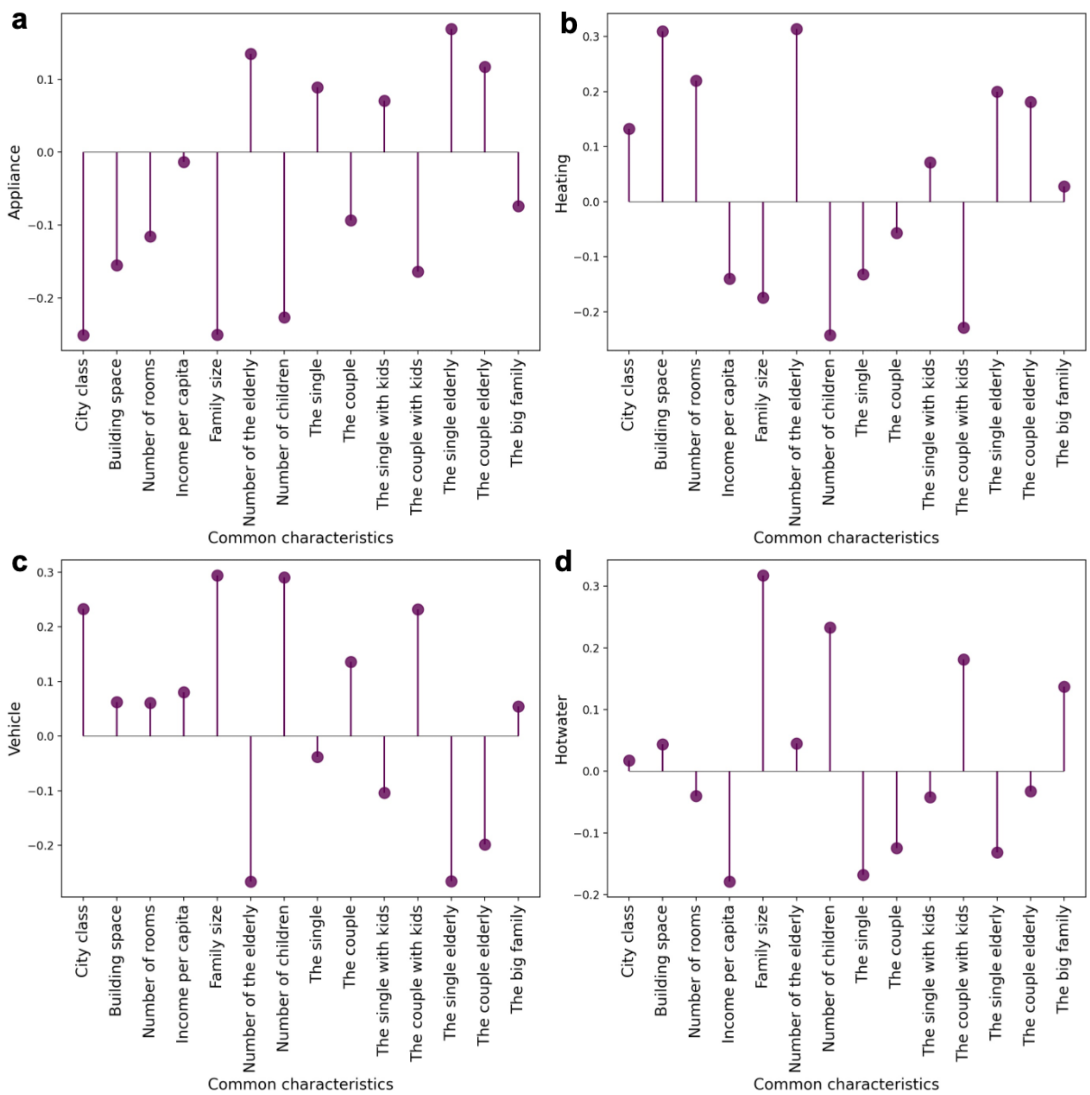


Fig.5. Normalized differences between household common characteristics across True/False groups for various emission patterns.

4. Discussions

Our findings show that high-emitting households can be identified using their characteristics. Small-sized households from large cities with high incomes are more likely to emit more than average households; if they also have intensive needs for commuting, high energy consumption, and heavy usages of appliances, such households are more inclined to have significantly higher emissions. The results associated with household lifestyles are much consistent with findings in the literature [19] where lifestyle shift is found to contribute to increasing emissions in Japan. Regarding decarbonization in household sector, specific emission reduction pathways can be proposed for these high-emission households based on the driving factors.

Based on the results of our study, the significant identified drivers for household emissions can be helped to point out corresponding emission reduction policies in terms of household energy consumption behaviours. This is also known as the Nudge intervention [70], which refers to the tendency of consumers to follow mainstream trends in decision making according to economics, behavioural science and psychology. Currently, Nudge intervention has already had a considerable influence on government decision-making process in developed countries such as Japan, the UK, and the USA. In this case, factor analysis can be an important method to provide policy implications by summarizing observational household energy use behaviours. Besides, the identification of key factors that explain the highest emission source for each household also provides a way for policymakers to develop customized low-carbon campaigns for different household category. Such customized campaigns are designed to raise the sense of self-efficacy rather than the mandatory economic incentives, which makes factor identification a cost-effective intervention to reduce emission.

Based on the Nudge intervention, Japan has also made practical policy implementation on the consumer side. For example, Japan has been implementing the Voluntary Energy-Saving Labelling Program and the Dictionary of Energy Savings in Households, which aim to inform consumers of the energy consumption of a product relative to the most efficient products in the same class [71]. This study provides the basis for these public policies, combined with Nudge intervention to provide insights into household emission reduction pathways.

Furthermore, our findings are consistent with those in the existing researches. Although carbon emissions per capita are lower in large families, it does not mean they have much lower emissions in total. High-emitting households could reduce emissions by choosing low-carbon lifestyles, such as using energy-efficient appliances, reducing the time spent on non-essential heating or cooling, and owning fewer vehicles. Policies can be developed to support these lifestyle choice improvements through economic incentives, educational guidance and infrastructure, which is, in this case, consistent with the findings in previous study [19,77,78]. However, the collection of information on household appliances from the survey data in this study enriches the scope of analysis of previous studies. Following the basic energy policies of Japan, as the part of "3E+S" system (Safety, Energy Security, Economic Efficiency, Environment), the energy efficiency is highlighted in household sector. Specifically, Japan's Ministry of Economy, Trade and Industry has suggested to improve energy efficiency in household sector by promoting high-efficiency hot water system, new energy vehicles, LED, etc., expecting to have 100% coverage of LED and 90% coverage of high-efficiency hot water system by 2030 [74]. In this case, it's anticipated to lower the final energy demand in household sector by 14 million kl by 2030. On the basis of the current plan, our results also suggest to expand the coverage of high-efficiency electric appliance to air conditioners, dryers, and heaters. In this case, both support measures and regulatory measures are required. Given the fact that households demand differently on electric appliance, differentiated regulatory and support measures can be developed [75]. For instance, an ecological points system could be developed for high-efficiency equipment and put carbon footprint labels on electric appliance to make consumers aware of the ecological impact.

It is also important to ensure that elderly households receive adequate heating and cooling in an ageing Japan, where older people spend more time at home and have higher heating and cooling needs than average households [76]. Also, as elder people are more inclined to own old-fashioned appliances and less likely to be adopt energy-saving behaviours, public guidance becomes particularly important for them. A suggested way to persuade them could be to use a trusted person close to them [21,82], and to make use of the energy efficiency program to upgrade their electric appliances to help abate additional emissions.

5. Conclusion and policy implications

In this study, we synthesized and identified the drivers of carbon emissions per capita of Japanese households from a pool of potential factors throughout several machine learning methods. Then we classified four categories of emission patterns by picking out the highest emission source of households and used machine learning models to independently identify the driving factors. Demographic features and specific electric appliance usages (including heating and hot water) were confirmed to be significant. As the effects of demographic change and lifestyle shift on household emissions have been recognized, our analysis takes a deep dive into the implications by mixing household demographic and behavioral characteristics [19,67].

The results show that single elderly households have the highest carbon emissions per capita in Japan. The higher the average age of households, the higher their emissions per capita. Our findings also identify the correlation between household emission trends and population trends in Japan. Additionally, high-emission households are mainly clustered in high-income households or households in north eastern Japan. Secondly, driving factors such as the number of electric water heaters, kerosene water heaters, electric heaters, and refrigerators explain the higher household emissions per capita the most. Thirdly, electricity-intensive appliances (e.g., air conditioners and heated toilets) and longer usage time have significant effects on appliance-driven emissions. Moreover, household driving frequency and the number of vehicles have a greater impact on household emissions per capita than the fuel types of vehicles.

For small families in big cities, who are typically single families, ensuring their living needs are the minimum standard. To further reduce emission by promoting energy conservation, we suggest raising their awareness of energy-saving behaviours, such as turning off the power when leaving the house and taking public transportation to reduce the use of personal vehicles. We also suggest small families to gradually adjust the scale of energy use from a whole house to the right scale according to their needs, e.g., avoiding excessive heating or idle use of appliances. Since the COVID-19 brings about a work style reform, people in big cities may transfer to telework and other flexible work styles in the future which is expected to benefit carbon emission reduction by saving electricity from not working at office [78].

For higher-income households, they are more concerned about life quality, thus, to reduce their excessive emissions, customized energy use plans and energy efficiency updates are needed. Besides, the government could give targeted low-carbon education to such households, for instance, suggestions related to uses of energy-efficient appliances and purchase on low-carbon products. In addition, in prefectures with intensive demand on commuting, wider popularization of NEVs can gradually replace traditional fuel and gasoline vehicles to significantly reduce emissions from commuting. NEVs is expected to achieve between 50% and 70% of sales target by 2030, led by the METI [84]. However, so far both unaffordable price (or limited subsidy on new purchased NEVs) and insufficient infrastructure (e.g. EV charging station) for further promotion on NEVs are hindering consumers from adopting them. Despite technology development, more supportive policies and incentivization are needed from the government side to encourage households choose NEVs.

Our findings provide insights into how to identify household emission patterns using numerous household characteristics, and thus identify high-emission households segment and its emission-intensive behaviours. Customized energy supply management and services can be developed based on household emissions. For example, for households with electricity-intensive appliances, an efficient way could be promoting energy-efficient appliances. Through education and effective campaigns, consumers can easily recognize and adopt eco-labels and energy-saving products. Besides, customized policy implementation should be taken into consideration, for instance, in prefectures with intensive demand on commuting, customized incentives for households to accept public transportation or NEVs could be provided [63].

The identification of different appliances' contribution to higher household emissions is important and useful to provide a better understanding of which appliance obtain positive effect on low-carbon transition and how the usage of appliances drives higher household emissions. As for household who uses coal and wood for cooking and heating, improving electrification or subsidizing them to adopt electricity or natural gas can reduce emissions. Furthermore, for low-income households in colder prefectures, photovoltaic promotion might be effective to reduce environmental degradation and poverty [30,85]. Besides, throughout the energy efficiency program, high-efficiency appliances are expected to significantly lower final energy demand in household sector (METI, 2018) by expanding the coverage of LED, high-efficiency hot water system, and other equipment. Also, household photovoltaic is another choice for residents to offset their own electricity use since FIT-based contracts between households and utilities have started to expire from 2019 [80]. From a broader perspective, our findings can reconcile the low-carbon transition suggestions for households with Japan's priority areas for carbon neutrality by 2050, including next-generation PV and lifestyle shift business, high-efficiency electric appliance, and new energy vehicles for households [81].

Of course, there are several limitations in the current study. For example, the temporal effect of household characteristics on their emission patterns were not considered in the model but we assumed that in the short run, households would not significantly change their energy use behaviours. Besides, the applied models were not carefully tuned for optimal parameters since the ultimate purpose of this study is to evaluate the key factors rather than optimize model accuracy, such as the heterogeneity of emission factors across regions. Also, as the accuracy and the complexity increasing, the future study can further reveal the spatial-temporal disparity of household emission features.

Acknowledgment

This dataset is driven from "Energy-Related CO_2 Emission Reduction Technology Evaluation/Commissioned Survey", commissioned by the Japanese Ministry of the Environment. All the data' cleaning, processing, analysis is conducted by committee member (Dr. Yin Long). Jing Meng appreciates the support from National Natural Science Foundation of China (NSFC) Grant No. 72173133.

Reference

[1] Raftery AE, Zimmer A, Frierson DMW, Startz R, Liu P. Less than 2 °c warming by 2100 unlikely. Nature Climate Change 2017;7:637–41. https://doi.org/10.1038/nclimate3352.

[2] O'Neill BC, Carter TR, Ebi K, Harrison PA, Kemp-Benedict E, Kok K, et al. Achievements and needs for the climate change scenario framework. Nature Climate Change 2020;10:1074–84. https://doi.org/10.1038/s41558-020-00952-0.

[3] Yan X, Fang YP. CO2 emissions and mitigation potential of the Chinese manufacturing industry. Journal of Cleaner Production 2015;103:759–73. https://doi.org/10.1016/j.jclepro.2015.01.051.

[4] Xue M, Lin B Le, Tsunemi K. Emission implications of electric vehicles in Japan considering energy structure transition and penetration uncertainty. Journal of Cleaner Production 2021;280:124402. https://doi.org/10.1016/j.jclepro.2020.124402.

[5] Du K, Li J. Towards a green world: How do green technology innovations affect total-factor carbon productivity. Energy Policy 2019;131:240–50. https://doi.org/10.1016/j.enpol.2019.04.033.

[6] Azadi M, Northey SA, Ali SH, Edraki M. Transparency on greenhouse gas emissions from mining to enable climate change mitigation. Nature Geoscience 2020. https://doi.org/10.1038/s41561-020-0531-3.

[7] Froemelt A, Dürrenmatt DJ, Hellweg S. Using Data Mining to Assess Environmental Impacts of Household Consumption Behaviors. Environmental Science and Technology 2018;52:8467–78. https://doi.org/10.1021/acs.est.8b01452.

[8] Song K, Anderson K, Lee SH. An energy-cyber-physical system for personalized normative messaging interventions: Identification and classification of behavioral reference groups. Elsevier Ltd, 2020. https://doi.org/10.1016/j.apenergy.2019.114237.

[9] International Energy Agence. Data and Statistics 2020. https://www.iea.org/data-and-statistics?country=JAPAN&fuel=CO2 emissions&indicator=TotCO2 (accessed April 29, 2021).

[10] Lau LC, Lee KT, Mohamed AR. Global warming mitigation and renewable energy policy development from the Kyoto Protocol to the Copenhagen Accord - A comment. Renewable and Sustainable Energy Reviews 2012;16:5280–4. https://doi.org/10.1016/j.rser.2012.04.006.

[11] Japanese Ministry of Environment (MOE). Submission of Japan's Nationally Determined Contribution (NDC) 2020. http://www.env.go.jp/en/headline/2442.html (accessed April 28, 2021).

[12] Kuriyama A, Tamura K, Kuramochi T. Can Japan enhance its 2030 greenhouse gas emission reduction targets? Assessment of economic and energy-related assumptions in Japan's NDC. Elsevier Ltd, 2019. https://doi.org/10.1016/j.enpol.2019.03.055.

[13] Long Y, Yoshida Y, Fang K, Zhang H, Dhondt M. City-level household carbon footprint from purchaser point of view by a modified input-output model. Applied Energy 2019;236:379–87. https://doi.org/10.1016/j.apenergy.2018.12.002.
[14] Lee S, Lee B. The influence of urban form on GHG emissions in the U.S. household sector. Energy Policy 2014;68:534–49. https://doi.org/10.1016/j.enpol.2014.01.024.
[15] Li H, Zhao Y, Wang S, Liu Y. Spatial-temporal characteristics and drivers of the regional residential CO2 emissions in China during 2000–2017. Journal of Cleaner Production 2020;276:124116. https://doi.org/10.1016/j.jclepro.2020.124116.
[16] Dou Y, Zhao J, Dong X, Dong K. Quantifying the impacts of energy inequality on carbon emissions in China: A household-level analysis. Energy Economics 2021;102:105502. https://doi.org/10.1016/J.ENECO.2021.105502.
[17] Tian J, Andraded C, Lumbreras J, Guan D, Wang F, Liao H. Integrating Sustainability Into City-level CO2 Accounting: Social Consumption Pattern and Income Distribution. Ecological Economics 2018;153:1–16. https://doi.org/10.1016/j.ecolecon.2018.06.019.
[18] Meangbua O, Dhakal S, Kuwornu JKM. Factors influencing energy requirements and CO2 emissions of households in Thailand: A panel data analysis. Energy Policy 2019;129:521–31. https://doi.org/10.1016/j.enpol.2019.02.050.
[19] Shigetomi Y, Matsumoto K, Ogawa Y, Shiraki H, Yamamoto Y, Ochi Y, et al. Driving forces underlying sub-national carbon dioxide emissions within the household sector and implications for the Paris Agreement targets in Japan. Applied Energy 2018;228:2321–32. https://doi.org/10.1016/j.apenergy.2018.07.057.
[20] Yuan R, Rodrigues JFD, Behrens P. Driving forces of household carbon emissions in China: A spatial decomposition analysis. Journal of Cleaner Production 2019;233:932–45. https://doi.org/10.1016/j.jclepro.2019.06.110.
[21] Yagita Y, Iwafune Y. Residential energy use and energy-saving of older adults: A case from Japan, the fastest-aging country. Energy Research and Social Science 2021;75:102022. https://doi.org/10.1016/j.erss.2021.102022.
[22] Dong K, Hochman G, Timilsina GR. Do drivers of CO2 emission growth alter overtime and by the stage of economic development? Energy Policy 2020;140:111420. https://doi.org/10.1016/J.ENPOL.2020.111420.
[23] Daniel Kahneman - Facts 2002. https://www.nobelprize.org/prizes/economic-sciences/2002/kahneman/facts/ (accessed August 23, 2021).
[24] Karlin B, Davis N, Sanguinetti A, Gamble K, Kirkby D, Stokols D. Dimensions of Conservation: Exploring Differences Among Energy Behaviors. Environment and Behavior 2014;46:423–52. https://doi.org/10.1177/0013916512467532.
[25] Andor MA, Fels KM. Behavioral Economics and Energy Conservation – A Systematic Review of Non-price Interventions and Their Causal Effects. Ecological Economics 2018;148:178–210. https://doi.org/10.1016/j.ecolecon.2018.01.018.
[26] Li J, Zhang D, Su B. The Impact of Social Awareness and Lifestyles on Household Carbon Emissions in China. Ecological Economics 2019;160:145–55. https://doi.org/10.1016/j.ecolecon.2019.02.020.
[27] Gifford R, Nilsson A. Personal and social factors that influence pro-environmental concern and behaviour: A review. International Journal of Psychology 2014;49:n/a-n/a. https://doi.org/10.1002/ijop.12034.

[28] Zhang W, Robinson C, Guhathakurta S, Garikapati VM, Dilkina B, Brown MA, et al. Estimating residential energy consumption in metropolitan areas: A microsimulation approach. Energy 2018;155:162–73. https://doi.org/10.1016/j.energy.2018.04.161.

[29] Wang Z, Han B, Lu M. Measurement of energy rebound effect in households: Evidence from residential electricity consumption in Beijing, China. Renewable and Sustainable Energy Reviews 2016;58:852–61. https://doi.org/10.1016/j.rser.2015.12.179.

[30] Shi X, Wang K, Cheong TS, Zhang H. Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data. Energy Economics 2020;92:104942. https://doi.org/10.1016/j.eneco.2020.104942.

[31] Fuks M, Salazar E. Applying models for ordinal logistic regression to the analysis of household electricity consumption classes in Rio de Janeiro, Brazil. Energy Economics 2008;30:1672–92. https://doi.org/10.1016/j.eneco.2007.09.006.

[32] Ang BW. Decomposition analysis for policymaking in energy:: which is the preferred method? Energy Policy 2004;32:1131–9. https://doi.org/10.1016/S0301-4215(03)00076-4.

[33] Wang J, Dong X, Dong K. How renewable energy reduces CO2 emissions? Decoupling and decomposition analysis for 25 countries along the Belt and Road. Https://DoiOrg/101080/0003684620211904126 2021;53:4597–613. https://doi.org/10.1080/00036846.2021.1904126.

[34] Meng J, Yang H, Yi K, Liu J, Guan D, Liu Z, et al. The Slowdown in Global Air-Pollutant Emission Growth and Driving Factors. One Earth 2019;1:138–48. https://doi.org/10.1016/j.oneear.2019.08.013.

[35] Wang C, Wang F, Zhang X, Yang Y, Su Y, Ye Y, et al. Examining the driving factors of energy related carbon emissions using the extended STIRPAT model based on IPAT identity in Xinjiang. Renewable and Sustainable Energy Reviews 2017;67:51–61. https://doi.org/10.1016/j.rser.2016.09.006.

[36] Shigetomi Y, Matsumoto K, Ehara T, Ochi Y, Ogawa Y, Shiraki H, et al. Socio-Demographic Drivers of Residential CO2 Emissions in the 47 Prefectures of Japan. SSRN Electronic Journal 2018. https://doi.org/10.2139/SSRN.3156008.

[37] Froemelt A, Buffat R, Hellweg S. Machine learning based modeling of households: A regionalized bottom-up approach to investigate consumption-induced environmental impacts. Journal of Industrial Ecology 2020;24:639–52. https://doi.org/10.1111/jiec.12969.

[38] Guo P, Lam JCK, Li VOK. Drivers of domestic electricity users' price responsiveness: A novel machine learning approach. Applied Energy 2019;235:900–13. https://doi.org/10.1016/j.apenergy.2018.11.014.

[39] Wang A, Lam JCK, Song S, Li VOK, Guo P. Can smart energy information interventions help householders save electricity? A SVR machine learning approach. Environmental Science and Policy 2020;112:381–93. https://doi.org/10.1016/j.envsci.2020.07.003.

[40] Shams Amiri S, Mottahedi S, Lee ER, Hoque S. Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption. Computers, Environment and Urban Systems 2021;88:101647. https://doi.org/10.1016/j.compenvurbsys.2021.101647.

[41] Liu X, Wang X, Song J, Wang H, Wang S. Indirect carbon emissions of urban households in China: Patterns, determinants and inequality. Journal of Cleaner Production 2019;241:118335. https://doi.org/10.1016/j.jclepro.2019.118335.

[42] Ken-ichiro N. Understanding the Energy Expenditures by Construction Periods using Micro Data of Survey on Carbon Dioxide Emissions from Residential Sector and Machine Learning. Journal of Japan Society of Energy and Resources 2021;42:175–84. https://doi.org/10.24778/JJSER.42.3_175.

[43] Long Y, Yoshida Y, Zhang H, Zheng H, Shan Y, Guan D. Japan prefectural emission accounts and socioeconomic data 2007 to 2015. Scientific Data 2020;7:233. https://doi.org/10.1038/s41597-020-0571-y.

[44] Yan Y, Zhang H, Long Y, Zhou X, Liao Q, Xu N, et al. A factor-based bottom-up approach for the long-term electricity consumption estimation in the Japanese residential sector. Journal of Environmental Management 2020;270:110750. https://doi.org/10.1016/j.jenvman.2020.110750.

[45] Ministry of the Environment. Household CO2 Emission Reality Survey (Household CO2 Statistics) 2019. http://www.env.go.jp/earth/ondanka/ghg/kateiCO2tokei.html (accessed September 18, 2021).

[46] Xu G, Ren X, Xiong K, Li L, Bi X, Wu Q. Analysis of the driving factors of PM2.5 concentration in the air: A case study of the Yangtze River Delta, China. Elsevier B.V., 2020. https://doi.org/10.1016/j.ecolind.2019.105889.

[47] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 1996;58:267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[48] Xiao R, Liu Y, Huang X, Shi R, Yu W, Zhang T. Exploring the driving forces of farmland loss under rapidurbanization using binary logistic regression and spatial regression: A case study of Shanghai and Hangzhou Bay. Elsevier B.V., 2018. https://doi.org/10.1016/j.ecolind.2018.07.057.

[49] Wang Z, Zhang B, Yin J, Zhang Y. Determinants and policy implications for household electricity-saving behaviour: Evidence from Beijing, China. Energy Policy 2011;39:3550–7. https://doi.org/10.1016/j.enpol.2011.03.055.

[50] Bishop C. Pattern Recognition and Machine Learning. Springer; 2006.

[51] Yesilbudak M, Sagiroglu S, Colak I. A novel implementation of kNN classifier based on multi-tupled meteorological input data for wind power prediction. Energy Conversion and Management 2017;135:434–44. https://doi.org/10.1016/j.enconman.2016.12.094.

[52] Donohoo BK, Ohlsen C, Pasricha S, Xiang Y, Anderson C. Context-aware energy enhancements for smart mobile devices. IEEE Transactions on Mobile Computing 2014;13:1720–32. https://doi.org/10.1109/TMC.2013.94.

[53] Suthaharan S. Decision Tree Learning, Springer, Boston, MA; 2016, p. 237–69. https://doi.org/10.1007/978-1-4899-7641-3_10.

[54] Gordon AD, Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Biometrics 1984;40:874. https://doi.org/10.2307/2530946.

[55] Quinlan J. C4. 5: programs for machine learning. USA: Morgan Kaufmann; 2014.

[56] Tso GKF, Yau KKW. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy 2007;32:1761–8. https://doi.org/10.1016/j.energy.2006.11.010.

[57] Sutton CD. Classification and Regression Trees, Bagging, and Boosting. vol. 24. Elsevier; 2005. https://doi.org/10.1016/S0169-7161(04)24011-1.

[58] Breiman L. Random forests. vol. 45. Springer; 2001. https://doi.org/10.1023/A:1010933404324.
[59] Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 2012;67:93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002.
[60] Liaw A, Wiener M. Classification and Regression by randomForest. Computer Science 2007.
[61] Luckner M, Topolski B, Mazurek M. Application of XGboost algorithm in fingerprinting localisation task. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10244 LNCS, Springer Verlag; 2017, p. 661–71. https://doi.org/10.1007/978-3-319-59105-6_57.
[62] Chen T, Guestrin C. XGBoost: A scalable tree boosting system, 2016. https://doi.org/10.1145/2939672.2939785.
[63] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 4768–4777.
[64] Zhang J, Yu B, Wei YM. Heterogeneous impacts of households on carbon dioxide emissions in Chinese provinces. Applied Energy 2018;229:236–52. https://doi.org/10.1016/J.APENERGY.2018.07.110.
[65] Enrico D'AMBROGIO. Japan's ageing society. 2020.
[66] Cabinet Office. White Paper on Aging Society (Full edition) 2018. https://www8.cao.go.jp/kourei/whitepaper/w-2019/zenbun/01pdf_index.html (accessed September 28, 2021).
[67] Long Y, Yoshida Y, Meng J, Guan D, Yao L, Zhang H. Unequal age-based household emission and its monthly variation embodied in energy consumption – A cases study of Tokyo, Japan. Applied Energy 2019;247:350–62. https://doi.org/10.1016/j.apenergy.2019.04.019.
[68] Watabe A, Leaver J, Ishida H, Shafiei E. Impact of low emissions vehicles on reducing greenhouse gas emissions in Japan. Energy Policy 2019;130:227–42. https://doi.org/10.1016/j.enpol.2019.03.057.
[69] Japanese Ministry of Environment (MOE). Japan's National Greenhouse Gas Emissions in Fiscal Year 2016 2016. https://www.env.go.jp/en/headline/2355.html (accessed April 23, 2021).
[70] Matsumoto S. How do household characteristics affect appliance usage? Application of conditional demand analysis to Japanese household data. Energy Policy 2016;94:214–23. https://doi.org/10.1016/J.ENPOL.2016.03.048.
[71] Statista. Japan: electric vehicle market share 2019 2019. https://www.statista.com/statistics/711994/japan-electric-car-market-share/ (accessed April 24, 2021).
[72] Zhou M, Liu Y, Feng S, Liu Y, Lu Y. Decomposition of rebound effect: An energy-specific, general equilibrium analysis in the context of China. Applied Energy 2018;221:280–98. https://doi.org/10.1016/J.APENERGY.2018.03.074.
[73] Jin T, Kim J. A new approach for assessing the macroeconomic growth energy rebound effect. Applied Energy 2019;239:192–200.

https://doi.org/10.1016/J.APENERGY.2019.01.220.
[74] Lin B, Zhu P. Measurement of the direct rebound effect of residential electricity consumption: An empirical study based on the China family panel studies. Applied Energy 2021;301:117409. https://doi.org/10.1016/J.APENERGY.2021.117409.
[75] Stankuniene G, Streimikiene D, Kyriakopoulos GL. Systematic Literature Review on Behavioral Barriers of Climate Change Mitigation in Households. Sustainability 2020;12:7369. https://doi.org/10.3390/su12187369.
[76] Iwata K, Katayama H, Arimura TH. Do households misperceive the benefits of energy-saving actions? Evidence from a Japanese household survey. Energy for Sustainable Development 2015;25:27–33. https://doi.org/10.1016/j.esd.2014.12.005.
[77] Shigetomi Y, Kanemoto K, Yamamoto Y, Kondo Y. Quantifying the carbon footprint reduction potential of lifestyle choices in Japan. Environmental Research Letters 2021. https://doi.org/10.1088/1748-9326/abfc07.
[78] Koide R, Lettenmeier M, Kojima S, Toivio V, Amellina A, Akenji L. Carbon Footprints and Consumer Lifestyles: An Analysis of Lifestyle Factors and Gap Analysis by Consumer Segment in Japan. Sustainability 2019, Vol 11, Page 5983 2019;11:5983. https://doi.org/10.3390/SU11215983.
[79] Ministry of Economy T and I. Japan’s Energy 2018 2018. https://www.enecho.meti.go.jp/en/category/brochures/pdf/japan_energy_2018.pdf (accessed September 30, 2021).
[80] Ministry of Economy. Energy Conservation Policies of Japan. 2011.
[81] Day R. Low carbon thermal technologies in an ageing society - What are the issues? Energy Policy 2015;84:250–6. https://doi.org/10.1016/j.enpol.2014.11.017.
[82] Chapman A, Shigetomi Y. Developing national frameworks for inclusive sustainable development incorporating lifestyle factor importance. Journal of Cleaner Production 2018;200:39–47. https://doi.org/10.1016/j.jclepro.2018.07.302.
[83] Ministry of the Environment. Annual Report on the Environment in Japan 2018. 2018.
[84] METI. Long-Term Goal and Strategy of Japan’s Automotive Industry for Tackling Global Climate Change. Tokyo: 2018.
[85] IEA. National Survey Report of PV Power Applications in JAPAN 2019. 2020.
[86] FIT. Based Contracts between Household Consumers and Electricity Companies Start to Expire Sequentially 2019. https://www.meti.go.jp/english/press/2019/1101_001.html (accessed September 30, 2021).
[87] Ministry of Economy T and I. Japan’s 14 Priority Areas for Carbon Neutrality by 2050 2020. https://www.meti.go.jp/english/policy/energy_environment/global_warming/roadmap/report/20201212.html (accessed September 30, 2021).

Appendix

Appendix Table 1. Variables and definitions

Classification	Variable	Definition
Housing	Prefecture	Household in 47 prefectures.
	City area	Household in 10 areas including Hokkaido, Tohoku, Tohoku,

		Tohoku, Tokai, Tokai, Chugoku, Shikoku, Kyushu and Okinawa.
	City class	City class of household, with 1 – large cities (prefectural capital and cities >100,000 population); 2 – medium cities (population 10,000-100,000); 3 – small cities (population <10,000).
	Build type	Build type of household, with 1 – detached; 2 – amalgamated.
	Build period	Build period of household, with 1 – before 1970; 2 – (1971-1980); 3 – (1981-1985); 4 – (1986-1990); 5 – (1991-1995); 6 – (1996-2000); 7 – (2001-2005); 8 – (2006-2010); 9 – after 2011.
	Build space	Household floor space (m^2).
	Room number	Number of rooms.
	Double-layer window	Whether the household has double window, with 0 – all windows have; 2 – on some of windows; 3 – no.
	Heat room number	Number of heating rooms.
Economic	Income	Annual household income, with 1 – under 2.5 million yen; 2 – (2.5-5) million yen; 3 – (5-7.5) million yen; 4 – (7.7-10) million yen; 5 – (10-15) million yen; 6 – (15-20) million yen; 7 – above 20 million yen
	Income per capita	Annual household income per capita.
	Labour number	Number of people in employment.
Demographic	Residential size	Number of occupants in the household.
	Average age	Average age of occupants in the household (everyone's age recorded in our survey is a range, e.g., 20-29 years old, we took the middle value and averaged the ages of everyone).
	Number of the elderly	Number of elderly people (age>65) in the household.
	Number of children	Number of children in the household.
	The single	Whether the occupant in household is single, with 0 – no; 1 – yes.
	The couple elderly	Whether the occupants in household are elderly couple, with 0 – no; 1 – yes.
	The single elderly	Whether the occupant in household is elderly single, with 0 – no; 1 – yes.
	The big family	Whether the household is a big family (number of occupants≥4), with 0 – no; 1 – yes.
	The couple	Whether the occupants in household are couple, with 0 – no; 1 – yes.
	The single with kids	Whether the occupants in household are single with kids, with 0 – no; 1 – yes.
	The couple with kids	Whether the occupants in household are couple with kids, with 0 – no; 1 – yes.
Appliance	TV	Number of TVs.
	Fridge	Number of fridges.

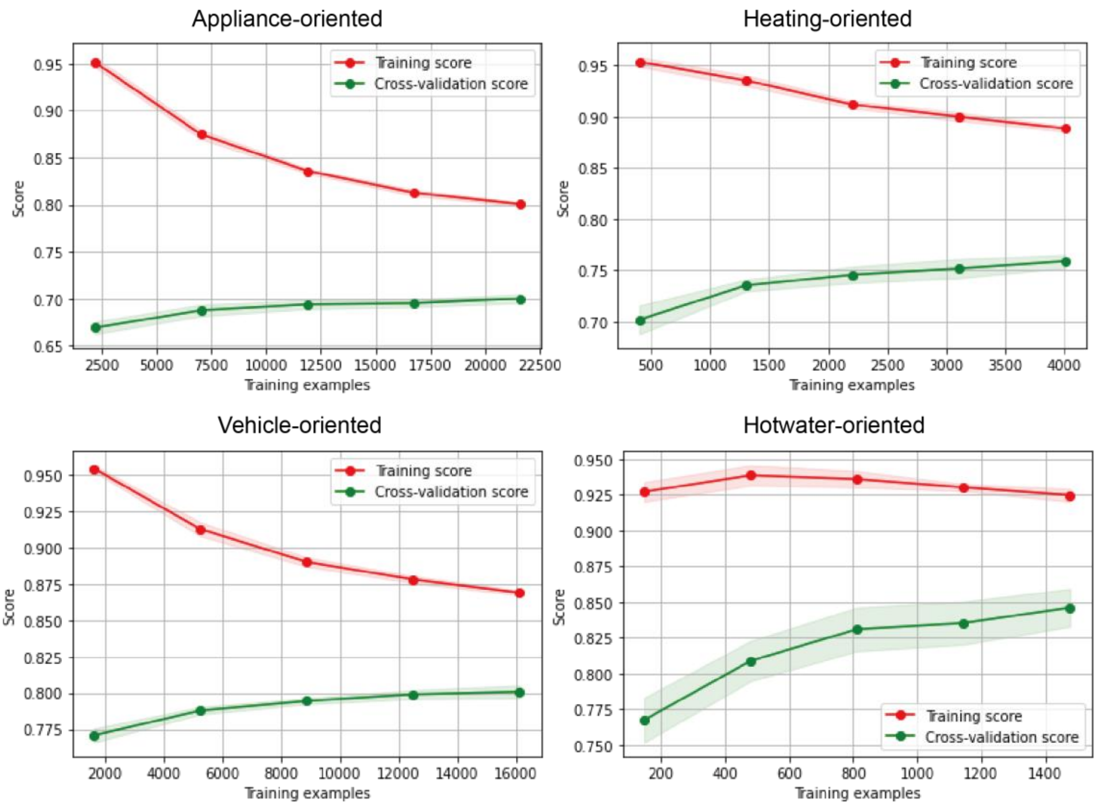
	Air conditioner	Number of air conditioners.
	Washing machine (no drying function)	Number of washing machines (no drying function).
	Washing machine (with drying function)	Number of washing machines (with drying function).
	Clothes dryer (electric)	Number of clothes dryers (electric).
	Clothes dryer (gas)	Number of clothes dryers (gas).
	Bathroom Dryer	Number of bathroom dryers.
	Dishwasher	Number of dishwashers.
	Dish dryer	Number of dish dryers.
	Microwave oven	Number of microwave ovens.
	Gas grill	Number of gas grills.
	Rice cooker	Number of rice cookers.
	Gas stove	Number of gas stoves.
	Electric kettle	Number of electric kettles.
	Wash the toilet with warm water	Number of wash the toilet with warm water.
	Heating toilet seat (without hot water washing function)	Number of heating toilet seats (without hot water washing function).
	Humidifier	Number of humidifiers.
	Dehumidifier	Number of dehumidifiers.
	Air cleaner	Number of air cleaners.
	Computer	Number of computers.
	DVD	Number of DVDs.
	Internet modem, router	Number of internet modem & routers.
Heating	Air conditioner	Number of air conditioners.
	Electric stoves	Number of electric stoves.
	Electric carpet	Number of electric carpets.
	Electric heat storage heater	Number of electric heat storage heaters.
	Gas stoves	Number of gas stoves.
	Kerosene stoves	Number of kerosene stoves.
	Stoves used woody fuel	Number of stoves used woody fuels.
	Solar heating	Number of solar heating.
	Central heating	Number of central heating.
	Floor heating	Number of floor heating.
Vehicle	Vehicle source	Type of vehicle sources.
	Vehicle number	Number of vehicle numbers.
	Motorcycle number	Number of motorcycle numbers.
	Electric motorcycle number	Number of electric motorcycle numbers.
	Driving frequency	Driving frequency of the household.
	Driving distance	Total driving distances of the household.
	Vehicle fuel	Type of vehicle fuel.
Behavior	Daytime at home	Whether the occupants in household are at home during the day on weekdays, with 1 – every day per week; 2 – three or

		four days per week; 3 – one or two days per week; 4 – almost none.
	Energy saving rate	The percentage of energy saving behaviors in the questionnaire.
	TV time	Time spent watching TV every day, with 1 – less than 1 hour; 2 – (1-2) hours; 3 – (2-4) hours; 4 – (4-8) hours; 5 – (8-12) hours; 6 – (12-16) hours; 7 – (16-20) hours; 8 – more than 20 hours.
	AC time	Time spent using air conditioner on summer every day, with 1 – less than 2 hours; 2 – (2-4) hours; 3 – (4-8) hours; 4 – (8-12) hours; 5 – (12-16) hours; 6 – (16-20) hours; 7 – (20-24) hours; 8 – 24 hours.
	Transportation saving	Whether to adopt energy-saving behavior in transportation, with 0 – no; 1 – yes.
	Light saving	Whether to adopt energy-saving behavior in light, with 0 – no; 1 – yes.
	Heating time	Time spent using heating on winter every day, with 1 – less than 2 hours; 2 – (2-4) hours; 3 – (4-8) hours; 4 – (8-12) hours; 5 – (12-16) hours; 6 – (16-20) hours; 7 – (20-24) hours; 8 – 24 hours.
	Hot water time	Time spent using hot water in kitchen every week, with 1 – every day; 2 – (5-6) days/week; 3 – (3-4) days/week; 4 – (1-2) days/week; 5 – almost none.

Appendix Table 2. Intensity of electricity, city gas, LP gas, gasoline, and light

Energy type	Heat conversion coefficient	CO_2 emission coefficient
Electricity	3.6 MJ/kWh	0.688 Kg-C/kwh in Hokkaido 0.525 Kg-C/kwh in Tochigi, Gunma, Ibaraki, Saitama, Chiba, Tokyo, Kanagawa, Yamashi 0.516 Kg-C/kwh in Nagano
City gas	44.8 MJ/m^3	0.0136 t-C/GJ
LP gas	50.8 MJ/kg	0.0161 t-C/GJ
Kerosene	36.7 MJ/ℓ	0.0185 t-C/GJ
Gasoline	34.6 MJ/ℓ	0.0183 t-C/GJ
Light oil	37.7 MJ/ℓ	0.0187 t-C/GJ

Appendix Figure 1. Learning curve of training examples of the XGBoost model



Appendix Table 3. The differences of household characteristics classification

Appliance-oriented household	2015		2017		2018	
Variable	No	Yes	No	Yes	No	Yes
Household space	114.77	101.93	106.11	99.02	107.79	99.30
Household size	2.967	2.691	2.907	2.523	2.849	2.467
If couple	0.114	0.086	0.123	0.095	0.131	0.101
Elderly number	0.732	0.831	0.751	0.876	0.788	0.942
Child number	0.653	0.482	0.648	0.414	0.584	0.359
Bathroom dryer	0.179	0.217	0.182	0.219	0.195	0.225
City class	1.895	1.734	1.910	1.729	1.925	1.721
Heating-oriented household	2015		2017		2018	
Variable	No	Yes	No	Yes	No	Yes
Electric heat storage heater	0.046	0.466	0.045	0.282	0.038	0.390
Stoves used woody fuel	0.008	0.008	0.020	0.011	0.008	0.007
Heating time	3.607	5.322	3.602	4.926	3.508	4.941
Heating room	1.857	2.549	2.167	2.942	2.181	2.906
Household income	205.47	180.68	225.22	204.67	228.48	205.31
Household space	105.37	129.16	100.71	117.15	101.80	119.16
Vehicle-oriented household	2015		2017		2018	
Variable	No	Yes	No	Yes	No	Yes
Renewable energy vehicle	0.006	0.003	0.012	0.007	0.012	0.008
Driving frequency	3.641	5.665	3.303	5.416	3.387	5.617
Driving distance	5955	14075	6491	14520	6652	15931

Vehicle number	1.177	1.801	1.049	1.725	1.058	1.800
Electric motorcycle number	0.005	0.006	0.012	0.013	0.003	0.005
Vehicle fuel	7.577	10.238	6.105	8.562	5.864	8.514

Hot water-oriented household	2015		2017		2018	
Variable	No	Yes	No	Yes	No	Yes
Electric water heater	0.075	0.690	0.091	0.430	0.088	0.377
Hot water time	1.802	1.491	1.772	1.498	1.839	1.385
Household income	204.10	189.39	224.27	191.30	227.63	188.36
Household size	2.791	3.145	2.679	3.203	2.628	3.082
If single	0.102	0.063	0.104	0.036	0.097	0.065
If couple elderly	0.148	0.149	0.160	0.135	0.180	0.169
If big family	0.169	0.206	0.142	0.191	0.134	0.203

Note: Please refer to the definition in Appendix Table 1 for the units of variables here.