



# Article Residential Energy Consumer Occupancy Prediction Based on Support Vector Machine

Dinh Hoa Nguyen 回

check for **updates** 

Citation: Nguyen, D.H. Residential Energy Consumer Occupancy Prediction Based on Support Vector Machine. *Sustainability* **2021**, *13*, 8321. https://doi.org/10.3390/su13158321

Academic Editor: Mouloud Denai

Received: 29 June 2021 Accepted: 22 July 2021 Published: 26 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). International Institute for Carbon-Neutral Energy Research (WPI-I<sup>2</sup>CNER), Institute of Mathematics for Industry (IMI), Kyushu University, Motooka 744, Nishi-ku, Fukuoka 819-0395, Japan; hoa.nd@i2cner.kyushu-u.ac.jp

**Abstract:** The occupancy of residential energy consumers is an important subject to be studied to account for the changes on the load curve shape caused by paradigm shifts to consumer-centric energy markets or by significant energy demand variations due to pandemics, such as COVID-19. For non-intrusive occupancy analysis, multiple types of sensors can be installed to collect data based on which the consumer occupancy can be learned. However, the overall system cost will be increased as a result. Therefore, this research proposes a cheap and lightweight machine learning approach to predict the energy consumer occupancy based solely on their electricity consumption data. The proposed approach employs a support vector machine (SVM), in which different kernels are used and compared, including positive semi-definite and conditionally positive definite kernels. Efficiency of the proposed approach is depicted by different performance indexes calculated on simulation results with a realistic, publicly available dataset. Among SVM models with different kernels, those with Gaussian (rbf) and sigmoid kernels have the highest performance indexes, hence they may be most suitable to be used for residential energy consumer occupancy prediction.

**Keywords:** energy consumer occupancy; consumer-centric energy systems and approaches; support vector machine; machine learning; artificial intelligence

# 1. Introduction

In residence and office buildings there is a strong correlation between the energy consumption and the occupancy of consumers. In particular, energy consumption peaks usually occur at time intervals during which consumers stay at their homes or offices [1–3]. Hence, consumer occupancy is one of important factors that has useful social implications. Examples include the establishment of more efficient energy management systems (EMSs) by means of better demand response (DR) programs [2], more economic operation of HVAC systems [4], energy saving buildings [5], of upgrading suggestions for energy systems [6], etc.

In the emerging paradigm shift to consumer-centric energy systems, consumer occupancy will be a critical index to be taken into account. In another context when a pandemic occurs, e.g., COVID-19, many people have to work from home, hence their energy consumption patterns will be significantly changed. This will definitely lead to variations on the load curve [7–11], which energy utilities have to reschedule their outputs in order to adapt to the changing demand. Therefore, occupancy analysis and prediction is a meaningful and practical area worth studying. Despite this, the occupancy of energy consumers is not directly known to utility companies, instead only the energy consumption data (electricity, gas, water, etc.) might be available.

Hitherto, existing methods for non-intrusive occupancy analysis and prediction are based on, either, data from various sensors deployed on-site [12–15], outputs of smart meters [1,16], artificial neural network (ANN) and machine learning (ML) approaches [13,14,17–20], or multivariate methods [21]. In [12], three  $CO_2$  sensors were deployed in a large-volume single-zone space to measure the flow of  $CO_2$  concentration in and out of the room, based on which prediction error minimization, ANN, and support vector machine (SVM) models were developed to count the number of occupants. Results of those models were also compared with a developed physical model, showing that their performances were better. Other types of sensors, e.g., low-resolution thermal imaging sensors [13], temperature and motion sensors [14,15], were employed to infer the occupancy information using different ML techniques, such as SVM, K-nearest neighbors, random forest, Bayes classification, and decision tree. Other ML methods were also utilized, e.g., long short-term memory (LSTM) [19], feed-forward neural network (FNN), extreme learning machine (ELM) [17], and hidden Markov model (HMD) [18]. For more details on the state of the art on both data-driven and analytical methods of occupancy detection and prediction, several existing review works are recommended, e.g., [3,20].

Power usage information provided from smart meters is another efficient way to deduce the energy consumer occupancy. A review on existing non-intrusive load monitoring datasets was performed in [22]. Binary occupancy, i.e., presence (1) and absence (0 or -1), can be detected from the outputs of smart meters [1,16] using several learning methods [16]. Similarly, water usage information can also be used for analysis and prediction of binary occupancy [23].

Although deploying many sensors for measurement of different factors (e.g., temperature,  $CO_2$  concentration, etc.) can provide more information about the presence of energy consumers, it will certainly increase the system capital cost, as well as the computational cost needed to process sensors' data. Bearing that in mind, this research aims to derive a cheap and lightweight approach for analysis and prediction of consumers' occupancy using only their energy consumption data for preservation of their private information and for saving system costs. These data can be collected from smart meters which are anticipated to be widely deployed at residential households in the near future along with other smart grid technologies.

Our proposed approach is based on the support vector machine (SVM), a supervised machine learning method, to obtain the binary classification of residential energy consumer occupancy. Distinct kernels (linear, polynomial, rbf, and sigmoid) and different time periods are investigated and compared to verify their performances in predicting the energy consumer occupancy. Accordingly, the contributions of this work are summarized below.

- Electricity consumption is used as the only feature for binary occupancy classification in SVM. This saves system costs since no additional sensors are deployed to collect other measurements on energy consumers. Additionally, the computational workload is reduced since fewer data need to be processed;
- A divide-and-average method to reduce the dimension of the data inputted to SVM, hence save computational time and cost. In this method, a high-dimension feature vector is divided into low-dimension vectors which are then summed up and averaged to attain the final feature vector for SVM;
- The proposed approach gives better performances compared to the existing result in the literature on the same dataset.

The rest of this paper is organized as follows. A brief introduction of SVM is given in Section 2. Then our SVM-based approach for occupancy analysis and prediction is presented and tested on a realistic dataset in Section 3. The paper is summarized and a few directions for future research are provided in Section 4.

## 2. Background on SVM

The purpose of SVM is to obtain a model for classification of data samples by learning from a given dataset. Conventionally, the learning goal of SVM is to derive separating hyperplanes to classify a given data set into different disjoint subsets and each of those subsets is assigned with a label. This is based on an assumption that the considering dataset is linearly separable, however this assumption does not hold for many realistic data sets. Hence, a technique called "kernel trick" was proposed to transform the considering dataset into another feature space in which it can be linearly separable. This gives rise to the use of kernel functions in non-linear SVM methods. Note that the conventional linear classification SVM models are a special case of kernel SVM models with the inner product being the linear kernel function. Therefore, in the following we will introduce backgrounds of kernel SVM methods for conciseness. Furthermore, we stick with the binary classification SVM since multi-class SVM methods can be generalized in a similar manner.

To begin, let  $\{(x_i, y_i)\}_{i=1,...,m}$  denote the dataset for training a SVM model, where  $x_i \in \mathcal{X} \subset \mathbb{R}^n$  are the feature vectors, and  $y_i \in \{1, -1\}$  are the labels associated with those feature vectors. Suppose that a feature map  $\phi(\cdot) : \mathcal{X} \to \mathbb{R}^p$  is selected for the data classification. Our aim then is to derive the parameters  $w \in \mathbb{R}^p$  and  $b \in \mathbb{R}$  for the SVM model so that the considering dataset can be linearly separated by hyperplanes  $w^T\phi(x_i) + b = 1$ , such that those  $x_i$  with  $y_i = 1$  lie on or above it, and  $w^T\phi(x_i) + b = -1$ , such that those  $x_i$  with  $y_i = -1$  lie on or below it. These conditions are equivalent to  $y_i(w^T\phi(x_i) + b) \ge 1$  for all i = 1, ..., m. The obtained SVM model hence can be used to predict for a test feature vector  $x \in \mathbb{R}^n$  by assigning its label  $y = \text{sgn}(w^T\phi(x_i) + b)$ , where sgn denotes the sign function.

Consequently, the determination of w and b can be handled in term of optimization problems. Let us start with the maximal (hard) margin classifier which is formulated as the following minimization problem.

min 
$$\frac{1}{2} \|w\|_2^2$$
 (1a)

s.t. 
$$y_i \left( w^T \phi(x_i) + b \right) \ge 1, i = 1, \dots, m$$
 (1b)

Let  $\alpha_i \in \mathbb{R}$ ,  $\alpha_i \ge 0$  be the Lagrange multipliers associated with the constraint (1b), the following Lagrangian is defined.

$$L(w,b,\alpha) \triangleq \frac{1}{2} \|w\|_2^2 - \alpha_i \Big[ y_i \Big( w^T \phi(x_i) + b \Big) - 1 \Big]$$
<sup>(2)</sup>

From the optimization theory, the infimum of this Lagrangian is achieved when the first-order conditions are satisfied, i.e., the partial derivatives of  $L(w, b, \alpha)$  with respect to w and b are vanished. That leads us to

$$w = \sum_{i=1}^{m} \alpha_i y_i \phi(x_i) \tag{3a}$$

$$0 = \sum_{i=1}^{m} \alpha_i y_i \tag{3b}$$

Next, substituting (3a) back to (2) gives us the infimum of the Lagrangian as follows.

$$\inf L(w,b,\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j)$$

$$\tag{4}$$

Denote  $K(x_i, x_j) \triangleq \phi(x_i)^T \phi(x_j)$ , then  $K(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$  is called a kernel. Let  $K \in \mathbb{R}^{m \times m}$  be a Gram matrix whose (i, j)-element is  $K(x_i, x_j)$ . This matrix must be a symmetric, positive semi-definite matrix due to the Mercer's theorem [24]. Accordingly, such a kernel is called a positive semi-definite (PSD) kernel. However, it is widely acknowledged that in practice some kernels which are not PSD but conditionally positive definite (CPD) also work well, e.g., the sigmoid kernel [25]. In the literature, the most common PSD kernels are: (i) polynomial:  $K(x_i, x_j) = (1 + x_i^T x_j)^q$ , where q is a positive integer; (ii) Gaussian or rbf:  $K(x_i, x_j) = e^{-\gamma ||x_i - x_j||^2}$ ,  $\gamma > 0$ , whereas one of the often used CPD kernel is the sigmoid kernel:  $K(x_i, x_j) = \tanh(\kappa x_i^T x_j + c), \kappa > 0, c < 0$ .

Now, we obtain the following dual optimization problem of Equation (1),

$$\max \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
(5a)

s.t. 
$$\sum_{i=1}^{m} \alpha_i y_i = 0$$
(5b)

$$\alpha_i \ge 0, i = 1, \dots, m \tag{5c}$$

As shown in [24], different SVM methods and models will finally end up with resolving a dual optimization problem similar to Equation (5), as follows.

$$\max \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
(6a)

s.t. 
$$\sum_{i=1}^{m} \alpha_i y_i = 0$$
 (6b)

$$0 \le \alpha_i \le C, i = 1, \dots, m \tag{6c}$$

The constraint (5c) is referred to as the box constraint in the SVM literature.

Lastly, the label of a test vector  $x \in \mathbb{R}^n$  is obtained by  $y = \text{sgn}(\sum_{i=1}^m \alpha_i y_i K(x_i, x) + b)$ , due to (3a). Using the Karush–Kuhn–Tucker (KKT) conditions, the following equation must also be satisfied for the found optimal values of the parameters w and b and of the Lagrange multiplier.

$$\alpha_i \Big[ y_i \Big( w^T \phi(x_i) + b \Big) - 1 \Big] = 0 \tag{7}$$

Thus, only the feature vectors  $x_i$  with which  $\alpha_i \neq 0$  can affect to the classification of a test vector, hence they are called supported vectors. These support vectors satisfy  $y_i(w^T\phi(x_i) + b) - 1 = 0$ , due to the condition (7), i.e., they lie on the hyperplanes on the feature space.

#### 3. Energy Consumer Occupancy Prediction

## 3.1. Electricity Consumption as a Learning Feature

To employ SVM for the analysis and prediction of energy consumer occupancy, a feature vector must be constructed based on the consumer electricity consumption. There may be multiple ways to do so, however in this research we directly use the consumer electricity consumption profile for constructing SVM feature vectors. The SVM feature vector length is determined by the electricity consumption data resolution and the time period for occupancy prediction. For instance, if the occupancy should be inferred each 15-min period and the electricity consumption data resolution is 1-minute, then the SVM feature vector length is 15. More specific will be illustrated in Section 3.3.

When the occupancy period to be validated is long (e.g., an hour), while the data resolution is high (e.g., one minute), the dimension of feature vectors is high, leading to high computation time and cost. However, if the intervals for consumer presence and absence inference are at least several times smaller than the validated occupancy period, such computational drawbacks can be eased by the following divide-and-average method.

First, the validated occupancy period, denoted by  $\mathbb{T}$ , is equally divided into a number, says *n*, of smaller time periods with length  $\mathcal{T}$ , i.e.,  $\mathbb{T} = n\mathcal{T}$ . Let  $\tilde{x}_1, \ldots, \tilde{x}_{\mathbb{T}}$  represent the electricity consumption during the period  $[1, \mathbb{T}]$ . Second, the electricity consumption is averaged during each time interval  $\mathcal{T}$  to obtain

$$\bar{x}_i \triangleq \frac{1}{\mathcal{T}} \left( \tilde{x}_{(i-1)\mathcal{T}+1} + \dots + \tilde{x}_{i\mathcal{T}} \right), \ i = 1, \dots, n$$
(8)

As the result, a new low-dimension feature vector of length *n* is constructed from the initial high-dimension feature vector  $[\tilde{x}_1, \ldots, \tilde{x}_T]^T$  of length  $\mathbb{T}$ , as follows.

$$x \triangleq [\bar{x}_1, \cdots, \bar{x}_n]^T \in \mathbb{R}^n \tag{9}$$

This method will be further illustrated through test cases in Section 3.3.

During the training process, SVM models are verified using the *k*-fold cross-validation method to assess their out-of-sample misclassification error. To this end, a summary of our SVM-based approach for predicting energy consumer occupancy is provided in Figure 1.



Figure 1. SVM-based approach for energy consumer occupancy prediction.

### 3.2. Realistic Dataset

In the current work, we employ the public-open dataset provided in [26]. More specifically, we utilize the data with one-minute resolution on the realistic electric consumption and occupancy profiles of two consumers in home A [26] for all simulations. Those data show great differences on the electric consumption and consumer occupancy between weekdays and weekends, and between different seasons. For example, in spring (Figure 2), during weekend home owners mostly stayed at home and only left for several hours in the evening. On the other hand, during weekdays they left home from the morning to the late afternoon (probably for working). Therefore, in this work we only focus our analysis on the electricity consumption and energy consumer occupancy during weekdays.



Figure 2. 1-min Resolution of a home electricity consumption and the modified 15-min resolution of its consumer occupancy, during the period of 1–7 April 2013, taken from [26].

## 3.3. Prediction Results

Performances of SVM models with different kernels will be compared through different indexes introduced in the following.

- Confusion matrix: a  $2 \times 2$  matrix for binary classification whose first row composes of true positive (TP) and false negative (FN), while its second row composes of false positive (FP) and true negative (TN). Here, TP means the prediction of consumer presence and the home is occupied, FN means prediction of consumer absence and the home is occupied, FP means prediction of consumer presence and the home is not occupied, and TN means prediction of consumer absence and the home is not occupied;
- Accuracy =  $\frac{TP+TN}{TP+TN+FP+FN}$ ; •
- Precision, or positive predictive value (PPV) =  $\frac{\text{TP}}{\text{TP}+\text{FP}}$ ; •
- True positive rate (TPR), or recall =  $\frac{\text{TP}}{\text{TP}+\text{FN}}$ ;
- True negative rate (TNR) =  $\frac{\text{TN}}{\text{TN}+\text{FP}}$ ;
- $F_1$ -score =  $\frac{2TPR \times PPV}{TPR + PPV}$ ; •
- Matthews correlation coefficient (MCC) =  $\frac{\text{TP} \times \text{TN} \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}};$ .
- Balanced accuracy =  $\frac{\text{TPR} + \text{TNR}}{2}$ .

The TNR, MCC and balanced accuracy are employed to better evaluate the performances of SVM models since the precision, TPR, and F<sub>1</sub>-score indexes focus only on the positive predictions, but not on the negative ones.

All simulations are performed in Matlab R2016b installed on a desktop computer equipped with Intel Core i7-6700K 4GHz CPU and 64GB RAM.

## 3.3.1. In Spring

It is worth noting that the occupancy profiles of two energy consumers in home A are almost identical, except a difference in the third weekday, as observed in Figure 2. The consumer occupancy profile displayed in green color has a more regular pattern, hence is considered for our analysis and prediction, for simplicity.

In the first simulation, we directly utilize the electricity consumption data in the first three weekdays, which are divided into periods of 15 min to train our SVM models. Different kernels are used, namely linear, polynomial, and radial basis function (rbf) kernels. In other words, the feature vectors fed to SVM models have length of 15, and the feature matrix has the dimensions of  $288 \times 15$ . The out-of-sample misclassification error for SVM model with linear, polynomial, rbf, and sigmoid kernel is 30.9%, 42.01%, 35.42%, and 35.76%, respectively, which are quite high.

Consequently, we use the above SVM models to predict the occupancy profile in the considering home for the last weekday. It then turns out that the result of linear kernel is totally wrong with 100% of consumer's presence, while the outcomes of polynomial and rbf kernels are much better, which are shown in Figure 3, because the residential electricity consumption data are not linearly separable. The sigmoid kernel, a CPD kernel [25], is also tested but its performance is not good, hence we do not show its result in Figure 3 to guarantee the figure clarity. Performance comparison of polynomial and rbf kernels is exhibited in Figure 4. As seen, the SVM model with polynomial kernel is worse than the SVM model with rbf kernel at some indexes but is better at some other indexes. In addition, the former is less accurate than the latter in predicting the presence of the energy consumer but is more accurate in predicting her absence.



Figure 3. 15-min Occupancy prediction in spring using SVM with different kernels.

	Index	Confusion matrix		Accuracy	Precision	TPR	F <sub>1</sub> -score	мсс	Balanced accuracy
	Kernel					TNR			
	polynomial	50	16	69.79%	79.37%	75.76%	77.52%	31.64%	66.22%
		13	17			56.67%			
		60	6	72.92%	75%	90.91%	82.18%	30.15%	62.12%
	וטו	20	10			33.33%			

Figure 4. Comparison of 15-min occupancy prediction in spring using SVM with different kernels.

However, both SVM models above with rbf and polynomial kernels misclassify the consumer occupancy at several intervals during her working time, i.e., her absence from home, as seen in Figure 3. This can be explained by some similarity of the electricity

consumption patterns when home owners were present and absent, as can be observed from Figure 2, leading to such wrong classification of the SVM models. Such similarity can be seen better, as shown in Figure 5, where the electricity consumption in each 15-min period is averaged and displayed for 3 training days and one test day. It is clear that during the presence of energy consumers at home the 15-min average electricity consumption are larger at several intervals, but are similar to that when they are not at home. This is obviously challenging for the occupancy classification.



Figure 5. 15-min Average electricity consumption and occupancy profile of one consumer.

Now, in the second simulation, we aim to predict the energy consumer occupancy in each hour interval. The same dataset for home A in Spring [26], above, is employed. Nevertheless, we do not go on the same route as in the first simulation, i.e., we do not use feature vectors having length of 60 containing one-minute electricity consumption data. Instead, our proposed divide-and-average is employed, in which the 15-min averages of electricity consumption, which was shown in Figure 5, will be utilized, resulting in feature vectors of length 4 and feature matrix of dimensions  $72 \times 4$  for the same dataset of three training days. Thus, we can significantly save the model training time and computational cost. In this scenario, the out-of-sample misclassification error when using the *k*-fold cross-validation for SVM model with polynomial, rbf, and sigmoid kernel is 2.78%, 5.56%, and 11.11%, respectively. Those errors are much smaller than that in the previous situation of 15-min occupancy prediction.

The results of the second simulation are depicted in Figure 6, where the sigmoid kernel is also used. The performance comparison of SVM models with different kernels in this case is shown in Figure 7. We can clearly observe that the SVM models with sigmoid and rbf kernels outperform that with polynomial kernel in this scenario in all performance indexes. On the other hand, the performances obtained with rbf and sigmoid kernels are slightly different in this case, but we note that with some other sets of parameters for the sigmoid kernel, their performances become identical.



Figure 6. 1-h Occupancy prediction in spring using SVM with different kernels.

Index Kernel	Confusion		Accuracy	Provision	TPR	E cooro	MCC	Balanced	
	(ernel	matrix		Accuracy	Precision	TNR	F <sub>1</sub> -score	witte	accuracy
	polynomial	12	5	70.83%	85.71%	70.59%	77.42%	38.74%	71.01%
		2	5			71.43%			
	sigmoid	14	3	83.33%	93.33%	82.35%	87.50%	63.91%	84.03%
		1	6			85.71%			
	rbf	15	2	83.33%	88.24%	88.24%	88.24%	59.66%	79.84%
		2	5			71.43%			

Figure 7. Comparison of one-hour occupancy prediction in spring using SVM with different kernels.

On the other hand, all models could not predict accurately the presence-absence switching times, though the sigmoid kernel is a bit better than the rbf kernel on tracking the switching time in this case. This can be explained by the fact that the number of hours for the presence-absence-presence pattern of the considering energy consumers in three training days are 9-8-7, 9-8-7, and 11-6-7, whereas for the test day is 10-7-7. Hence, the investigating SVM models are probably not sophisticated enough to capture such differences in the occupancy patterns, which requires further works to improve the occupancy predicting performance.

## 3.3.2. In Summer

For the same home as in the previous section, the electricity consumption profile and home owners' occupancy patterns in summer are much more irregular, as can be seen in Figure 8 and more clearly in Figure 9 for the same owner considered in the previous section for prediction. Particularly, the very short presence or absence of this home owner happened several times, while his or her long presence at home also occurred a few times. This makes the occupancy prediction even more challenging.



**Figure 8.** 1-min Resolution of a home electricity consumption and the modified 15-min resolution of its consumer occupancy, during the period of 8–14 July 2013 [26].



**Figure 9.** 15-min Average electricity consumption and occupancy profile of one consumer during 8–14 July 2013 [26].

In this case, the out-of-sample misclassification error when using the *k*-fold crossvalidation for SVM model with polynomial, rbf, and sigmoid kernel is 18.49%, 20.83%, and 35.42%, respectively, which are relatively high. The occupancy prediction results for the next weekday using polynomial, rbf, and sigmoid kernels are shown in Figure 10. Surprisingly, the prediction accuracy obtained with the sigmoid kernel is much better than that with the other two kernels. Such performance differences can be clearly seen in Figure 11, where all performance indexes of the SVM model with sigmoid kernel outperform that of the other SVM models with polynomial and rbf kernels. This is very interesting because its performance for the 15-min occupancy prediction in Spring in the previous section was worst than that utilizing the polynomial and rbf kernels.



Figure 10. 15-min Occupancy prediction using SVM for 12 July 2013.

Index	Confusion matrix		Accuracy	Precision	TPR	F <sub>1</sub> -score	мсс	Balanced accuracy
Kernel					TNR			
nohmomial	45	16	77.08%	88.24%	73.77%	80.36%	54.62%	78.72%
polynomial	6	29			82.86%			
cigmoid	51	10	89.58%	100%	83.61%	91.07%	80.64%	91.81%
sigitiolu	0	35			100%			
	47	14	83.33%	95.92%	77.05%	85.46%	68.68%	85.67%
rbi	2	33			94.29%			

**Figure 11.** Comparison of 15-min occupancy prediction using SVM with different kernels for 12 July 2013.

## 3.4. Comparison with Existing Results

In this section, we attempt to compare the performance of our proposed approach with that of an existing algorithm in the literature [1] conducted on the same dataset. Note that, the algorithm in [1] was also simple and threshold-based. The prediction results in [1] were evaluated through the confusion matrix, accuracy, TPR, precision, and F<sub>1</sub>-score indexes, and could be improved by combining different metrics (average power, standard deviation, and power range).

The comparison detail is provided in Figure 12. As seen, our proposed approach outperforms that in [1] for both cases of spring and summer times.

	This work	Ref. [1]	This work	Ref. [1]	
Index Period	Αссι	uracy	F <sub>1</sub> -score		
spring week	83.33%	73.27%	88.24%	N/A	
summer week	89.58%	79.09%	91.07%	82%	

Figure 12. Performance comparison for the same data of home A in Spring given in [26].

### 4. Conclusions and Future Works

This paper has proposed a SVM-based approach for analysis and prediction of binary occupancy for residential energy consumers. This approach is different from other existing ones in which only the electricity consumption is used as its input, hence system cost can be significantly reduced. In addition, a divide-and-average strategy has been proposed to decrease the dimension of the input feature vector. As a result, computational time and cost can be saved. Despite of its simplicity, the proposed approach' performance has been shown to outperform that of an existing method in the literature, performed on the same realistic dataset. Our test results also suggest that SVM models with rbf (Gaussian) kernel and sigmoid kernel give the highest performances.

There are several challenges to be addressed in the future research. First, the presenceabsence switching times are very difficult to track due to aperiodic occupancy patterns of residential energy consumers. Second, similarities on the electricity consumption patterns during the absence and presence periods make them hard for being distinguished. Furthermore, the occupancy and electricity consumption patterns can be completely different between weekends and weekdays, as well as between different seasons, which pose additional challenges to be resolved for the energy consumer occupancy analysis and prediction. Last but not least, novel kernels should be developed for achieving better prediction results on the binary occupancy of energy consumers.

Funding: This research was financially supported by JSPS Kakenhi Grant Number JP19K15013.

**Acknowledgments:** The author would like to thank Yoshinobu Kawahara at Institute of Mathematics for Industry, Kyushu University for discussions related to this research.

Conflicts of Interest: The author declares no conflict of interest.

#### References

- Chen, D.; Barker, S.; Subbaswamy, A.; Irwin, D.; Shenoy, P. Non-Intrusive Occupancy Monitoring using Smart Meters. In Proceedings of the BuildSys'13: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, Rome, Italy, 11–15 November 2013; pp. 1–8.
- 2. Chaney, J.; Owens, E.H.; Peacock, A.D. An evidence based approach to determining residential occupancy and its role in demand response management. *Energy Build.* **2016**, 125, 254–266. [CrossRef]
- 3. Rueda, L.; Agbossou, K.; Cardenas, A.; Henao, N.; Kelouwani, S. A comprehensive review of approaches to building occupancy detection. *Build. Environ.* **2020**, *180*, 106966. [CrossRef]
- 4. Dorokhova, M.; Ballif, C.; Wyrsch, N. Rule-based scheduling of air conditioning using occupancy forecasting. *Energy AI* **2020**, 2, 100022. [CrossRef]
- 5. Naylor, S.; Gillott, M.; Lau, T. A review of occupant-centric building control strategies to reduce building energy use. *Renew. Sustain. Energy Rev.* **2018**, *96*, 1–10. [CrossRef]
- Rana, A.; Perera, P.; Ruparathna, R.; Karunathilake, H.; Hewage, K.; Alam, M.S.; Sadiq, R. Occupant-based energy upgrades selection for Canadian residential buildings based on field energy data and calibrated simulations. *J. Clean. Prod.* 2020, 271, 122430. [CrossRef] [PubMed]
- 7. Santiago, I.; Moreno-Munoz, A.; Quintero-Jiménez, P.; Garcia-Torres, F.; Gonzalez-Redondo, M. Electricity demand during pandemic times: The case of the COVID-19 in Spain. *Energy Policy* **2021**, *148*, 111964. [CrossRef] [PubMed]
- Abu-Rayash, A.; Dincer, I. Analysis of the electricity demand trends amidst the COVID-19 coronavirus pandemic. *Energy Res.* Soc. Sci. 2020, 68, 101682. [CrossRef] [PubMed]

- Madurai Elavarasan, R.; Shafiullah, G.; Raju, K.; Mudgal, V.; Arif, M.; Jamal, T.; Subramanian, S.; Sriraja Balaguru, V.; Reddy, K.; Subramaniam, U. COVID-19: Impact Analysis and Recommendations for Power Sector Operation. *Appl. Energy* 2020, 279, 115739. [CrossRef]
- 10. Pecan Street. Shifting Energy Use Trends Due to COVID-19. 2020. Available online: https://www.pecanstreet.org/wp-content/uploads/2020/05/Covid-Webinar-May-2020-Slide-Deck-.pdf (accessed on 18 June 2021).
- 11. New York Independent System Operator. Recent Impacts on Hourly Load Patterns. 2020. Available online: https://www.nyiso. com/-/covid-19-and-the-electric-grid-load-shifts-as-new-yorkers-respond-to-crisis (accessed on 18 June 2021).
- 12. Zuraimi, M.; Pantazaras, A.; Chaturvedi, K.; Yang, J.; Tham, K.; Lee, S. Predicting occupancy counts using physical and statistical CO<sub>2</sub>-based modeling methodologies. *Build. Environ.* **2017**, *123*, 517–528. [CrossRef]
- Chidurala, V.; Li, Z. Occupancy Estimation Using Thermal Imaging Sensors and Machine Learning Algorithms. *IEEE Sens. J.* 2021, 21, 8627–8638. [CrossRef]
- Wang, C.; Jiang, J.; Roth, T.; Nguyen, C.; Liu, Y.; Lee, H. Integrated sensor data processing for occupancy detection in residential buildings. *Energy Build*. 2021, 237, 110810. [CrossRef]
- Sardianos, C.; Varlamis, I.; Chronis, C.; Dimitrakopoulos, G.; Himeur, Y.; Alsalemi, A.; Bensaali, F. A model for predicting room occupancy based on motion sensor data. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2–5 February 2020; pp. 394–399.
- Jin, M.; Jia, R.; Spanos, C.J. Virtual Occupancy Sensing: Using Smart Meters to Indicate Your Presence. *IEEE Trans. Mob. Comput.* 2017, 16, 4490–4501. [CrossRef]
- 17. Wei, Y.; Xia, L.; Pan, S.; Wu, J.; Zhang, X.; Han, M.; Zhang, W.; Xie, J.; Li, Q. Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. *Appl. Energy* **2019**, *240*, 276–294. [CrossRef]
- 18. Ryu, S.H.; Moon, H.J. Development of an occupancy prediction model using indoor environmental data based on machine learning techniques. *Build. Environ.* **2016**, *107*, 1–9. [CrossRef]
- 19. Feng, C.; Mehmani, A.; Zhang, J. Deep Learning-based Real-time Building Occupancy Detection Using AMI Data. *IEEE Trans. Smart Grid* **2020**, *11*, 4490–4501. [CrossRef]
- 20. Himeur, Y.; Ghanem, K.; Alsalemi, A.; Bensaali, F.; Amira, A. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Appl. Energy* **2021**, *287*, 116601. [CrossRef]
- 21. Liu, C.; Akintayo, A.; Jiang, Z.; Henze, G.P.; Sarkar, S. Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network. *Appl. Energy* **2018**, *211*, 1106–1122. [CrossRef]
- 22. Iqbal, H.K.; Malik, F.H.; Muhammad, A.; Qureshi, M.A.; Abbasi, M.N.; Chishti, A.R. A critical review of state-of-the-art non-intrusive load monitoring datasets. *Electr. Power Syst. Res.* **2021**, 192, 106921. [CrossRef]
- Fransson, V.; Bagge, H.; Johansson, D. A method to estimate absence in apartments based on domestic water use. *Build. Environ.* 2020, 180, 107023. [CrossRef]
- 24. Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods; Cambridge University Press: Cambridge, UK, 2000. [CrossRef]
- Lin, H.T. A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-Type Methods. 2005. Available online: https://home.work.caltech.edu/~htlin/publication/doc/tanh.pdf (accessed on 23 April 2021).
- Laboratory for Advanced System Software, University of Massachusetts Amherst, USA. NIOM Occupancy Dataset. 2017. Available online: http://traces.cs.umass.edu/index.php/Smart/Smart (accessed on 23 April 2021).