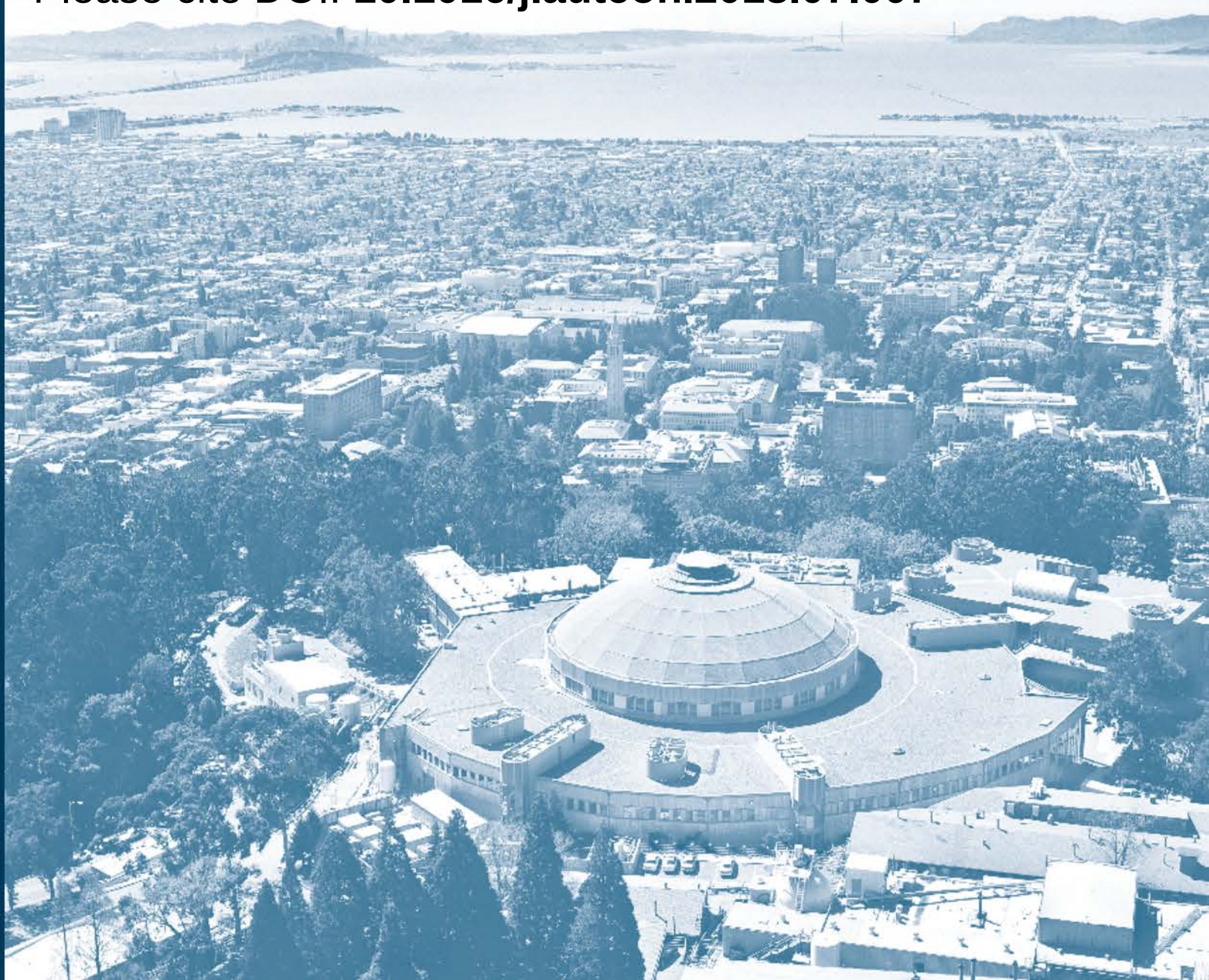
Lawrence Berkeley National Laboratory

Occupancy prediction through machine learning and data fusion of the environmental sensing and Wi-Fi sensing in buildings

Wei Wang, City University of Hong Kong
Jiayu Chen, City University of Hong Kong
Tianzhen Hong, Lawrence Berkeley National Laboratory

October 2018

Please cite DOI: **10.1016/j.autcon.2018.07.007**



Occupancy prediction through machine learning and data fusion of the environmental sensing and Wi-Fi sensing in buildings

Wei Wang [a,b], Jiayu Chen [a,*], Tianzhen Hong [b,*]

[a] Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621, AC1, Tat Chee Ave, Kowloon, Hong Kong

[b] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Abstract: In recent years, many occupancy studies have used environmental sensor data (such as carbon dioxide [CO_2], air temperature, and relative humidity) or Wi-Fi data to predict building occupancy information. However, the value of a data fusion approach that uses both environmental sensing and Wi-Fi sensing to predict occupancy remains an open question. To answer this question, this study conducted an on-site experiment in one office room in City University of Hong Kong. Three feature-based occupancy models using machine learning algorithms—k-nearest neighbors (kNN), support vector machine (SVM), and artificial neural network (ANN)—were selected to learn and predict occupancy information. In the model input, the study tested three data groups: environmental parameters only, Wi-Fi data only, and a combination of both. To assess the three occupancy models, the mean average error (MAE), mean average percentage error (MAPE), and root mean squared error (RMSE) indices were utilized. Results showed that when only the environmental parameter data were applied to learn occupancy, the ANN-based occupancy model was more suitable and accurate, and so will be with the combination of environmental parameters and the Wi-Fi data. The SVM model is more suitable and accurate in learning occupancy information with Wi-Fi data. The ANN model is more suitable environment dataset and the combination. On the other hand, the combination of datasets cannot improve accuracy significantly during three days when compared with

* Corresponding author. tel.: +852 3442 4696;

e-mail addresses: jiaychen@cityu.edu.hk (Jiayu Chen)

* Corresponding author. tel.: +1 (510)486-7082;

e-mail addresses: thong@lbl.gov (Tianzhen Hong)

occupancy results under environmental parameters, and decreases the accuracy under the Wi-Fi dataset. However, the combination of both datasets can improve robustness of occupancy prediction.

Keywords: occupancy prediction, environmental sensing, Wi-Fi sensing, data fusion, machine learning

1. Introduction

The term occupancy behaviors refers to building occupants' presence, movement, and interactions with buildings, which have significant impacts on building performance [1–3]. As geometrical location, envelope, and shapes of buildings are optimized, the influence of occupant behavior on building performance increases [4,5]. By observing how occupants behave inside buildings and interact with building technology, Zhou et al. simulated building energy performance using quantitative occupant use patterns in centralized and decentralized air-conditioning systems [6]. Wang and Chen [7,8] applied indoor positioning systems to obtain occupancy distribution information and simulated the impact on energy-saving in building air-conditioning control systems. The results showed about 22% of building energy could be saved using this approach. A further study [9] in an occupancy-based multi-zone outdoor air control model revealed a saving of about 23.6% of ventilation energy cost during work days. Therefore, occupant behavior becomes an inspiring objective for not only efficient HVAC controls but also developing building energy models [10–14] .

In recent years, the relationship between occupancy behavior and indoor environmental parameters, including lighting use, CO_2 concentration, air temperature, and relative humidity (RH), has been established and proven to be useful [15–17]. Occupancy estimation with environmental sensing of multiple parameters is a significant trend [18–20]. Zhu et al. estimated office occupancy with environmental sensing via non-iterative local receptive fields in time and frequency domains with a dataset including CO_2, air RH, air temperature, and air pressure. Becerik-Gerber et al. studied a fusion of light, sound, motion, CO_2, temperature, relative humidity, passive infrared sensors (PIR), door switch sensors, and applied autoregressive moving average model (ARMA), Nerual Network, Markov Chain, and Logit Regression to model occupancy profiles [21].

On the other hand, Wi-Fi signals are widely available in modern buildings and play an important role in the interactions between occupants and building services. When buildings are in use, indoor envrionmental parameters, such as temperature, RH, and CO2 will vary with an occupant's body movements and actions, and the Wi-Fi signal utilization variation will vary as occupants use computers and mobile devices. In Mohamed' study [22], the effectiveness and possibility of using Wi-Fi technologies to detect and predict building occupancy in classroom level were studied and it found Wi-Fi counts predicted actual occupancy level more accurately than CO_2 concentration by investigating Person's correlation between Wi-Fi connections, CO_2 concentration and the number of occupants, respectively. However, this study didn't further research how to use Wi-Fi counts or CO_2 concentration to infer the number of occupants and how to fuse both datasets in predicting occupants. Moreover, Chen et al. demonstrated that Wi-Fi connections and disconnections have a positive relationship with energy load variation [23]. Balaji [24] showed that Wi-Fi connections can determine building occupancy profiles with an 83% accuracy. Wang et al. also explored the Wi-Fi probe-based occupancy study to sense the Wi-Fi signal request and response, which resulted in more than 80% accuracy of occupancy detection [25].

However, a data fusion of Wi-Fi signal and environmental parameters in improving accuracy of occupancy prediction has not been explored. Inspired by and built upon the environmental sensing and Wi-Fi sensing in occupancy studies, this study proposes a data fusion-based machine learning approach that employs both environmental sensor data and Wi-Fi data to predict office building occupancy. We also deepen our exploration of data-driven forecast by examining the effectiveness of occupancy prediction using three machine learning algorithms.

The main contributions of this study are as follows:

(1) It proposed a fusion framework for data-driven occupancy modeling with the combination of environmental sensing (temperature, relative humidity, and CO_2 concentration) and Wi-Fi probe technology.

(2) It conducted analytics to determine if Wi-Fi data can improve performance of environmental parameter-based occupancy prediction, and vice versa.

(3) It evaluated the performance of ANN, kNN, and SVM algorithms in predicting occupancy features using experimental data.

This study tested also the performance of three machine learning techniques in occupancy prediction using different datasets, and provided guidelines for selecting the appropriate machine learning algorithms.

This work is organized as follows: Section 2 reviews and discusses some related works. Section 3 includes the overview of this study and feature-based machine learning techniques for occupancy modeling. The experiment, ground truth, and assessment indices are introduced in Section 4. Sections 5 and 6 show the results and discuss occupancy profile features. Section 7 offers conclusions.

2. Related works

Energy consumption of the building is highly related to the occupant's behavior, the way how the energy efficiency is improved should be focused on occupancy behavior [26–28]. As the main topic of the International Energy Agency (IEA) Energy in Buildings and Communities (EBC) Programme Annex 66, occupant behavior (OB) has been regarded as having one of the most significant impacts on building energy consumption and energy-saving technology [29]. Annex 66 summarized detailed occupancy modeling issues, including occupancy sensing and model evaluation. By observing OB with a single environmental parameter or technology, many researchers have utilized sensor ssystems, such as CO_2 concentration [30–32], PIR [33], lighting sensors [34,35], and radio frequency identification (RFID) [12,36]. On the other hand, energy consumption of devices in building can be a good indicator of occupancy. Diaz and Jimenez measured the power consumption of computers and CO_2 concentration, to imply occupancy information, respectively and the results suggested that CO_2 concentration is informative and expected to serve as a good indicator of occupancy [37]. Gul and Patidar proposed a study of understanding of energy consumption and occupancy in a multi-purpose academic building [13].

A recent development has been the use of multiple environmental sensors for occupancy estimation. The environmental parameters usually include indoor air temperature, RH, air pressure, and indoor CO_2 concentration. Recently, some studies have focused on using one or more of those parameters to model OB. For example, Pedersen et al. applied an occupancy detection method using air temperature, humidity, CO_2, volatile organic compounds (VOC), passive infrared sensor (PIR), and noise sensors. The experiment was conducted in a simple test room and in a three-room dorm, to detect two occupancy room statuses—occupied or vacant—resulting in a maximum accuracy of 98% and 78%, respectively in the two study areas [38]. Soh [39,40] studied occupancy estimation has been studied using air temperature, RH, CO_2, and air pressure.

Several algorithms, such as machine learning algorithms and probabilistic models, with data fusion of occupancy sensing technologies have been discussed individually; namely, Markov chain, artificial neural network (ANN), k-nearest neighbors (kNN), support vector machine (SVM), Classification and Regression Trees (CART), extreme learning machine (ELM), and linear discriminant functions (LDA). For example, Jiang et al Applied ELM to estimate and predict the occupancy information using CO_2 concentrations and verified the model in an office room. Page et al. [41] also used the Markov models to model occupancy behaviors by considering Boolean occupancy status (occupied or vacant). Peng et al. using kNN based machine learning technique for occupancy prediction in office areas and also propose multi-learning processes for demand-driven cooling control [42]. Lin et al estimated the number of people in a crowded scenes using SVM model to classify the featured area pf the head-like contour. [43] Szczurek et al. studied the performance of three environmental parameters—air temperature, RH, and CO_2—individually and as a combined three-sensor array in occupancy determination. They also compared the kNN algorithm with LDA when occupancy classification was required, and found kNN to be more efficient [44]. Based on machine learning techniques, Ryu and Moon developed one occupancy prediction model using CO_2, first order shifted of difference of CO_2, indoor CO_2 moving average and rate of change, and indoor and outdoor CO_2 ratio as indoor environmental data features [45]. Yang and Becerik-Gerber installed three sensor boxes, including light, sound, motion, CO_2, temperature, RH, PIR, and door-switch sensors in three typical offices, and applied cameras to collect occupancy ground truth. Based on sensing data, ARMA, Neural Network (NN), Markov Chain, and Logit Regression algorithms were used to model occupancy profiles [21]. Two data-driven decision tree and hidden Markov model (HMM) algorithms proved well suited to predict occupancy. With a fusion of light sensor, Candanedo and Feldheim also evaluated a method using temperature, humidity, and CO_2 sensors to predict occupancy with different statistical classification models: LDA, CART, and Random Forest (RF). They obtained about 97% accuracy when using only two of the environmental parameters with the LDA model in a one-day measurement. To estimate occupancy in a large area, Dong et al. [16] applied one information technology-enabled sustainability test-bed (ITEST) for occupancy prediction with a wireless ambient-sensing system, a wired CO_2 sensing system, and a wired indoor air quality sensing system. The experiment was conducted in a large-scale open office area, and it resulted in an average of 73% accuracy.

As Wi-Fi signals are widely used in buildings, Wi-Fi offers a more efficient, affordable, and convenient option to enhance indoor occupants' communication service, especially in commercial and residential buildings. There are many Wi-Fi enabled devices (such as laptops, smart phones, tablets, and wearables), and it is assumed that one occupant owns at least one device to connect to Wi-Fi networks. Indoor users request and connect to the Wi-Fi signal and receive the response from Wi-Fi signal. When buildings are in use, the indoor environment is influenced by occupants and Wi-Fi signal utilization. Chen [23] et al. found that Wi-Fi signal connections and disconnections have a positive relationship with building energy variation. Wi-Fi connections and disconnections also can be utilized as indicators of building occupancy [46]. Many studies based on Wi-Fi technology have been applied in occupancy patterns and energy efficiency studies [17, 36–38]. Balaji et al. used Wi-Fi connections to determine building occupancy profiles with an 83% accuracy [24]. To investigate occupancy patterns and improve building energy efficiency through Wi-Fi, Wang and Shao conducted one 24-hour monitoring over 30 days in a library and applied a rule mining approach, finding that 26.1% of the total energy cost can be saved [50]. In a Wi-Fi based occupancy-driven lighting control study, Zou et al. demonstrated 93.09% and 80.27% of energy saving compared to static scheduling and a PIR-based lighting control scheme, respectively [51]. An experiment at the Massachusetts Institute of Technology (MIT) identified occupants through walls with Wi-Fi signals by sensing the Wi-Fi signal distribution varying with occupants' movements in indoor spaces [52]. In that study, Wi-Fi signals distributed through indoor space like the air surrounding them and were reflected by the human body.

The approaches above employed environmental sensors or a Wi-Fi signal to estimate occupancy using various data-driven algorithms. In this study, we investigated whether using the performance of the Wi-Fi signal to estimate office occupancy could be improved by adding environmental sensors, or vice versa. We proposed a data fusion method of

Wi-Fi signal and environmental parameter sensing, and applied machine learning algorithms, including ANN, SVM, and kNN. To evaluate their performance, we conducted an on-site experiment in a large office room.

3. METHODOLOGY

3.1 Overview of this study

Figure 1 illustrates the overall approach. This study applied Wi-Fi probe technology to collect Wi-Fi data by recording a Wi-Fi connection probe request and response from user devices. During the experiment, the data collected from Wi-Fi signals contained the media access control (MAC) address of devices, received signal strength indicator (RSSI), and frequency of the MAC address. The data were uploaded into one server through the Wi-Fi network. The collected data were then uploaded to a server and downloaded to local storage with an application programming interface (API). A software development kit (SDK) was provided if further development with the Wi-Fi data was needed, where the SDK consists of typically a set of software development tools that enrich users to apply advanced functionalities. Data from the environmental sensing considered in this study were indoor air temperature, RH, and CO_2 concentration. The synchronization between two types of occupancy sensing is the time label, which can be used as an important index of data fusion. For the environmental sensors, a sensor hub can help upload data if the sensor is wireless, or it can directly store data locally.

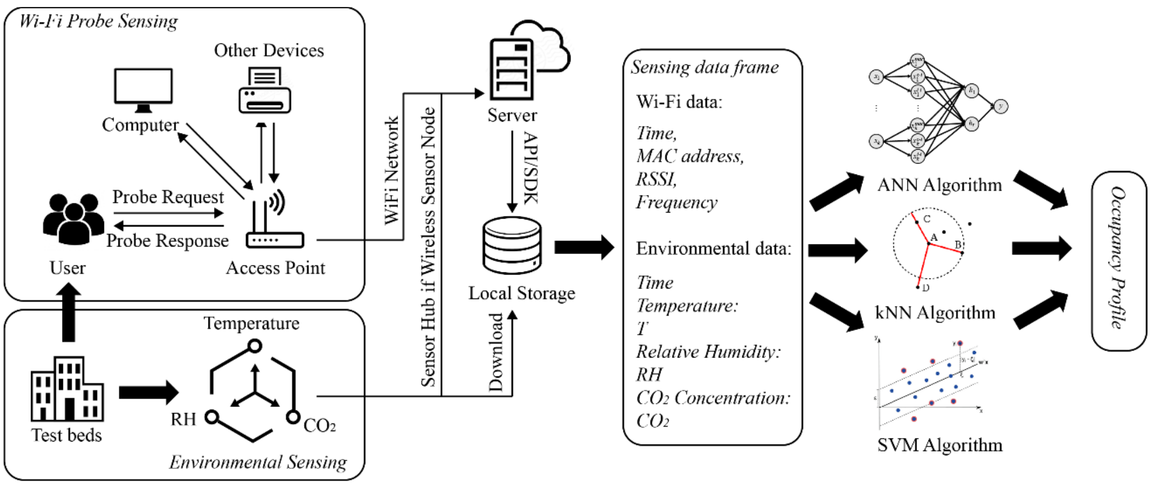


Fig. 1. Study overview.

Three machine learning techniques—kNN, ANN, and SVM—were used to predict the occupancy profile. The test was divided into three groups, to model occupancy profiles with (1) the environmental parameters dataset, (2) the Wi-Fi dataset, and (3) both. The mean average error (MAE), mean average percentage error (MAPE), and root mean squared error (RMSE) were used to compare results and determine good machine techniques for different datasets.

3.2 Feature-based machine learning techniques

3.2.1 Backpropagation (BP)-based ANN model

An artificial neural network (ANN) algorithm is used to solve problems in the same way that a human brain would; a network of simple neuron elements connect the output to input with a directed and weighted graph. The capabilities of the ANN algorithm can be applied in regression analysis, including time series prediction and modeling, classification (including pattern recognition and sequential decision making), and so on. A backpropagation is a method used in ANN algorithms to calculate the error contribution of each neuron. In the context of learning, backpropagation is commonly used by the gradient descent optimization to adjust the weight of neurons by calculating the gradient of the loss function.

The BP-based ANN algorithm usually has three layers—an input layer, a hidden layer, and an output layer—and the weights between the layers are usually random. Given training samples $(\boldsymbol{x_k}, \boldsymbol{y_k})$, $\boldsymbol{x_k}$ is the collection of data of selected features and $\boldsymbol{y_k}$ is the corresponding occupancy profile feature.

ANN models usually have three layers: an input layer, a hidden layer, and an output layer. Once an occupancy feature selection is determined, the input layer can be defined as in Eq. 1. The output of the hidden layer, output of the output layer, weights from the hidden layer, and weights from the hidden layer to the output layer are defined as in Eq. 2, Eq. 3, Eq. 4, and Eq. 5, respectively.

$$\mathrm{X_{input}} = (\mathrm{x_1}, \mathrm{x_2}, \dots, \mathrm{x_i}, \dots, \mathrm{x_n})^{\mathrm{T}} \tag{1}$$

$$\mathrm{H_{hidden}} = (\mathrm{h_1}, \mathrm{h}, \dots, \mathrm{h_j}, \dots, \mathrm{h_m})^{\mathrm{T}} \tag{2}$$

$$Y_{output} = (y_1, y_2, \dots, y_k, \dots, y_l)^T \tag{3}$$

$$V = (v_1, v_2, \dots, v_j, \dots, v_m) \tag{4}$$

$$W = (w_1, w_2, \dots, w_k, \dots, w_l) \tag{5}$$

Where x_n comes from the combination of selected occupancy feature data and n is the length of feature data length or the number of neural cells of the input layer. The m and l are the length of the hidden layer and output layer, respectively. The v_j donates the weight vector of the jth neural cell of the hidden layer, and w_k donates the weight vector of the kth neural cell of the output layer. The length of the input layer is determined by the number of elements of input data, while the length of the hidden layer (m) is usually randomly selected. The length of the output layer (l) usually equals the number of expected output elements.

The mathematic information transfer between each layer can be expressed as in Eq. 6 and Eq. 7:

$$h_j = f(\sum_{i=0}^{n} v_{ij} x_i)\,, \quad j = 1,2, \dots, m \tag{6}$$

$$y_k = f(\sum_{j=0}^{m} w_{jk} h_{j)}\,, \quad k = 1,2,3 \tag{7}$$

Where the transfer function f(x) can usually be the sigmoid function, which is formatted as in Eq. 8:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

If we define the ground truth of occupancy information as Eq. 9, assuming one output neuron, the squared error function is shown in Eq. 10:

$$d = (d_1, d_2, \dots, d_k, \dots, d_l)^T \tag{9}$$

$$E = \frac{1}{2}\sum_{k=1}^{l}(d_k - y_k)^2 \quad (10)$$

The gradient descent method involves calculating the derivative of the squared error function with respect to the weights of the network. Eq. 11 and Eq. 12 can be used to correct the weights between layers.

$$\Delta v_{jk} = \eta(\sum_{k=1}^{l}(d_k - y_k)y_k(1 - y_k)w_{jk})(1 - h_i)h_i \quad (11)$$

$$\Delta w_{ij} = \eta(d_k - y_k)y_k(1 - y_k)h_i \quad (12)$$

Implementation of the ANN algorithm can be generated in three steps:

(1) Randomly assign the length of the hidden layer (m) and weights of layers (V, W).

(2) Calculate the occupancy feature in the output layer.

(3) Calculate the error and reassign the weights.

3.2.2 K-nearest neighbors (kNN) model

The k-nearest neighbors (kNN) algorithm is quite popular for classification and prediction, as it is a nonparametric, nonlinear, distance-based method. Given a training dataset(x_k, y_k), where k = 1, 2, …, N, and a test sample t, the distance, d_k, between t and x_k can be calculated as in Eq. 13:

$$d_k = ||t - x_k|| \quad (13)$$

Where ||·|| is the distance. One of the most widely applied distance calculations is Euclidean distance. After obtaining the distance d_k, k =1, 2, …., N, the labels of k training samples with the smallest distance can be used. Then, a majority voting will be performed to determine the label of the testing sample.

3.3.3 Support Vector Machine (SVM) model

Support vector machine (SVM) was also chosen as a means of estimating the expectation of variables, and this method has been already utilized extensively in the previous research for regression and prediction [16, 42]. Analogously, the model produced by SVM depends only on a subset of the training data, because the cost function for building the model ignores any training data close to the model prediction. Several SVM model techniques have been developed to analyze and forecast occupancy models [42, 43]. Training the original SVM means solving Eq. 14 and Eq. 15:

$$\text{Minimize } \frac{1}{2}\|\omega\|^2 \tag{14}$$

$$\text{Subject to } \begin{cases} y_k - \langle \omega, x_i \rangle - b \leq \varepsilon \\ \langle \omega, x_i \rangle + b - y_k \leq \varepsilon \end{cases} \tag{15}$$

Where x_i is a training sample with target value y_i. The inner product plus the intercept $\langle \omega, x_i \rangle + b$ is the prediction for that sample, and ε is a free parameter that serves as a threshold: all predictions must be within an ε range of the true predictions. Slack variables are usually added to the above to allow for errors and to allow approximation in the case the above problem is infeasible.

4. Experiment setting and validation procedure

4.1 Description of the testbed

The testbed is a graduate student office located inside an institutional building in City University of Hong Kong. The office has an area of about 200 square meters (m^2) and had 25 long-term residents during the experiment period. Figure 2 shows the testbed's space layout and equipment setup. The office has two entrances. Ground truth is acquired by two overhead cameras installed to record the entrance and exit events of occupants. The number of occupants were manually counted based on the recorded video for each minute. Wi-Fi probes recorded the connection requests and responses of all wireless devices within the space. The sensing timestep Wi-Fi probe was set as 30 seconds and collected data were automatically uploaded to the server and downloaded from the SDK. The TA465-X (environmental sensors by TSI Company) were utilized to monitor and record the air temperature, RH, and CO_2 concentration. The sensing interval default was 1 minute. Table 1 shows the specifications of the installed sensors.

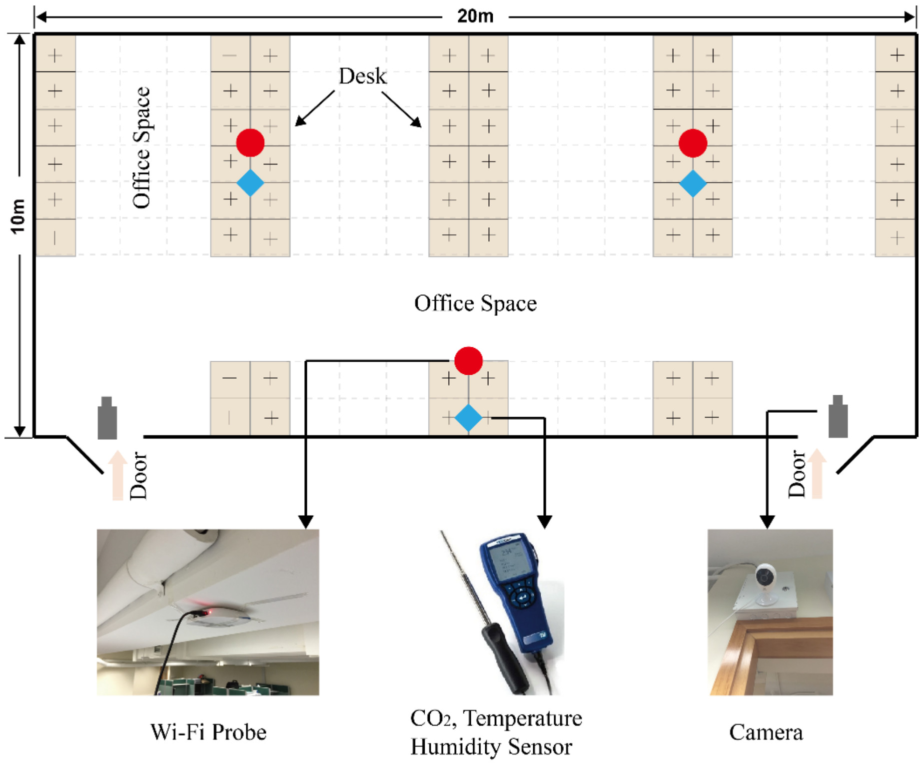


Fig. 2. Space layout and equipment setup

Table 1. Sensors used in the experiment

Sensor	Camera	Wi-Fi Probe	Environment Sensors			
			CO_2 Sensors	Temperature Sensors	Humidity Sensors	Other Sensors
Cost (USD)	45	30	400			
Recorded Variables	time, actual occupancy	time, MAC address, RSSIs	time, temperature, relative humidity, CO_2, air flow rate, air pressure			
Data Storage	online	online	local			
Measurement timestep		30s	1min	1min	1min	
Range			0–5k ppm	14°F-140°F -10℃–60℃	0 to 95%	
Accuracy			±3% or ±50 ppm	±0.5°F (±0.3℃)	< 3%	
Resolution			1 ppm	0.1°F (0.1℃)	0.10%	

4.2 Model configuration

The five-minute time interval of occupancy profiles was modeled, and the occupancy model was run from 08:30 to 19:00, total 127 data points for one day; however, the overtime occupancy profile model was not considered in this study. Machine learning techniques like ANN, kNN, and SVM are not sophisticated enough to accurately predict the real-time occupancy level directly from raw data, such as measured environmental parameters and Wi-Fi data, especially when different parameters have a different scale. After measurement, the feature-driven machine learning techniques require normalization, to eliminate the impact of the data scale on the results. With the dataset, this study built three occupancy models using three data groups, with one including only the environmental parameters (T, RH, CO_2), one including only the Wi-Fi data, and one including both datasets. The results were compared among the three different machine learning techniques on those three data groups. This study generated three datasets with "feature-target" type, where the feature were the inputs of three occupancy model, including sensed environment parameters and Wi-Fi data and target were the outputs, which are the actual occupancy data. To eliminate the scale of features, this study normalized three datasets before implementing three occupancy models. In the parameter selection for ANN model setting, the size of input layers were determined by the number of feature parameter for each dataset, while the hidden layer was selected as 3 and output layer was 1. In kNN model, the k value was selected as 15, which was to compare the predicted occupancy value with nearest 15 targets using Euclidean distance. For SVM model, the ε was set as 0.2 and Gaussian kernel function was used. 9-day on-site experiment were conducted in this study. For the model training, the training dataset composed of former 6-day experiment data, including 6 x 127 x n data points, where n is the feature size of different model. The validation data used in this study was later 3-day experiment data. To apply the three machine learning techniques, the Python programming language has been used to fit the models and predict occupancy offline and numpy and sklearn were the two of mainly used packages.

4.3 Assessment indices

During the experiment, the actual occupancy of the room (ground truth) was acquired through manual video analysis of the camera recordings. In this study, assessment of the prediction were conducted between two continuous variables, which were the predicted occupancy value and actual occupancy value. Thus, to assess the prediction accuracy, three indices were used to compare the predicted results with the actual occupancy. Mean Absolute Error (MAE) is a measure of difference of predicted versus observed, which compares the mean error between the occupant counts in a zone, and can be defined as in Eq. 16. Mean Absolute Percentage Error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics and shows the mean error as a percentage between the predicted occupant count and the actual count, as in Eq. 17. The MAE and MAPE provide the evaluation indices to assess the distance between predicted and observed values, finally giving the error levels for HVAC studies, such as the cooling load estimation with occupant count. Root Mean Squared Error (RMSE) is a frequently used measure of the differences between the values predicted by the model and the values actually observed and shows the magnitude of the estimation error, as in Eq. 18. The three issues are widely applied to assess the prediction accuracy issues for occupancy studies as a statistical method.

$$\mathrm{MAE}(O^{p}) = \frac{1}{N}\sum_{i=1}^{N}\left|O_i^{ob} - O_i^{p}\right| \tag{16}$$

$$\mathrm{MAPE}(O^{p}) = \frac{1}{N}\sum_{i=1}^{N}\left|(O_i^{ob} - O_i^{p})/O_i^{ob}\right| \tag{17}$$

$$\mathrm{RMSE}(O^{p}) := \sqrt{\sum_{n=1}^{N}(O^{ob} - O^{p})^2/N} \tag{18}$$

Where, O^{ob} is the observed occupant count as the ground truth and O^{p} is the predicted occupant count. N is the size of occupancy data.

5. Results

5.1 Model performance with the environmental parameter sensing data

In this study, we first modeled occupancy profiles with only the environmental parameters, including the indoor air temperature, RH, and CO_2 concentration. Figures 3, 4, and 5 show the results of the occupancy profile under the ground truth, kNN, SVM, and ANN models during three days using the environmental sensing data. Table 2 shows the assessment results of different methods with MAE, MAPE, and RMSE using the environmental parameters. The occupancy profile and accuracy results demonstrated that environmental sensing is a good approach and that it can learn to predict occupancy. The overall trend of the predicted occupancy matches the actual occupancy. From the results in Table 2, the best performance in terms of MAE, MAPE, and RMSE appeared on Day 1, with a minimum MAE of 1.8 and RMSE of 2.2. The MAE of the three algorithms is around

1.8, while the maximum occupant number was 14, on Day 1. However, the MAPE results show that the occupancy results fluctuate. On Day 2, the three occupancy models showed similar results in MAE, MAPE, and RMSE; while on Day 3, the ANN had better results. Overall, based on results for the three days, the ANN-based occupancy model showed the best performance. Therefore, we could conclude that the ANN model is more effective and accurate to use when environmental parameters are utilized to learn occupancy information.

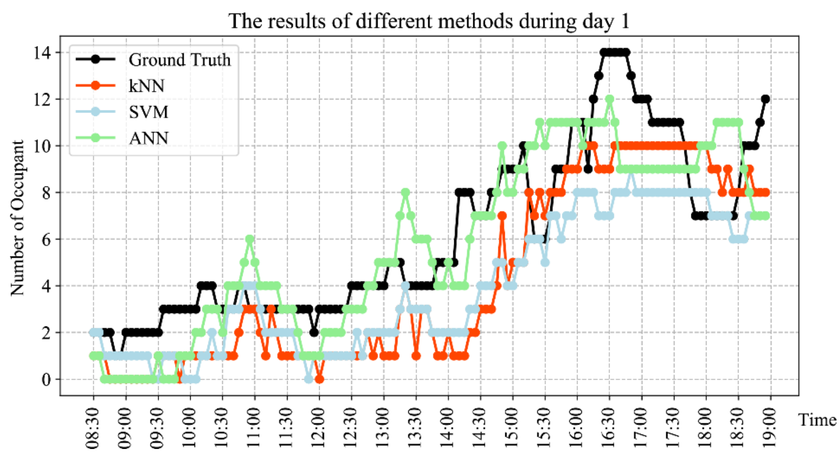


Fig. 3. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 1 using only the environmental parameters data

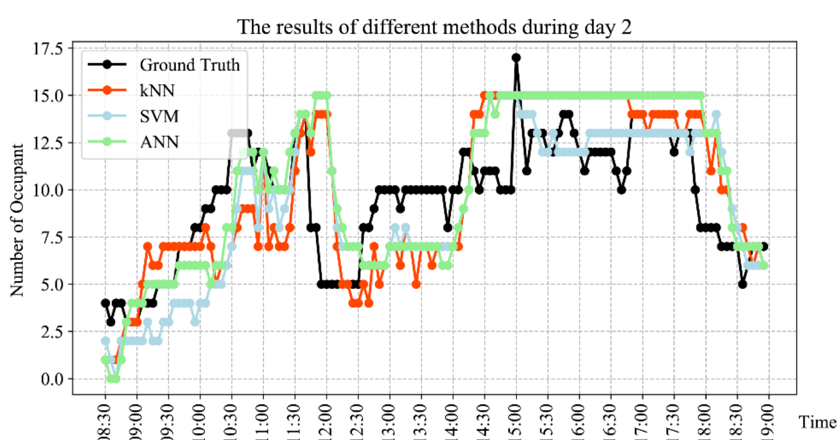


Fig. 4. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 2 using only the environmental parameters data

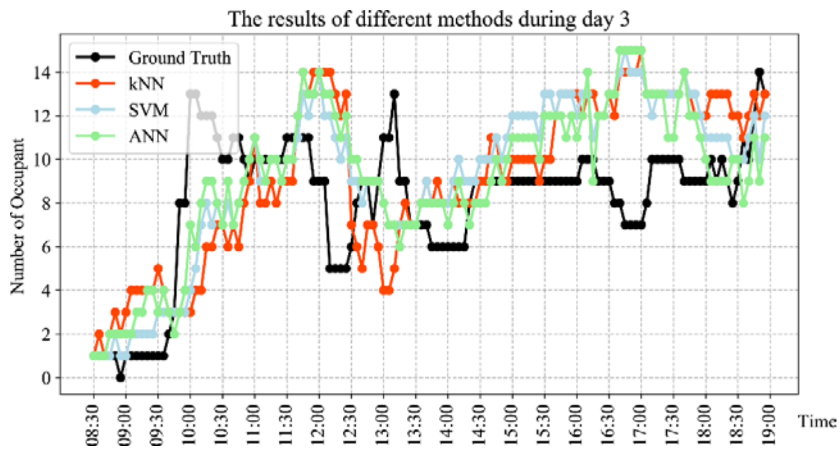


Fig. 5. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 3 using only the environmental parameters data

Table 2. The results of three machine learning occupancy models with MAE, MAPE, and RMSE using only the environmental parameters data.

	kNN			SVM			ANN		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Day 1	2.3	47.6%	2.7	2.3	41.0%	2.9	1.8	37.5%	2.2
Day 2	2.6	30.6%	3.2	2.4	30.8%	3.0	2.5	30.9%	3.1
Day 3	3.0	55.2%	3.8	2.6	38.0%	3.3	2.5	43.5%	3.2
All 3 days	2.6	44.4%	3.2	2.4	36.6%	3.1	2.3	37.3%	2.9

5.2 Model performance with the Wi-Fi data

This study also modeled occupancy profiles with Wi-Fi data, and Figures 6, 7, and 8 show the results of occupancy profiles under the ground truth, kNN, SVM, and ANN models during three days. Table 3 shows the assessment results of different machine learning methods with MAE, MAPE, and RMSE using only the Wi-Fi data. The occupancy results using Wi-Fi data on Day 1 are less accurate than occupancy profiles that used only environmental parameters. The mismatch increased between the predicted occupancy and the actual occupancy during the afternoon. The MAE results in Day 1 are about 2.7, 2.7, and 3.7, using kNN, SVM, ANN, respectively, and the variation results in Day 1 are 3.6, 3.7, and 3.9. Those results show that the Wi-Fi sensing method in Day 1 is more unstable. However, accuracy on Day 2 was reasonable, especially when the SVM model was applied. MAE results on Day 2 showed the best performance on Day 2. The ANN-based occupancy model had higher MAE than kNN and SVM; while considering MAPE and RMSE, kNN and SVM occupancy models had close performance. For Day 3, SVM occupancy models had the best result of MAE when compared with kNN and ANN. Therefore, in a Wi-Fi based occupancy model, the kNN and SVM models can be two good choices, whereas the simplified backpropagation-based ANN model could not directly predict occupancy well, so more adaptions of the ANN model should be applied.

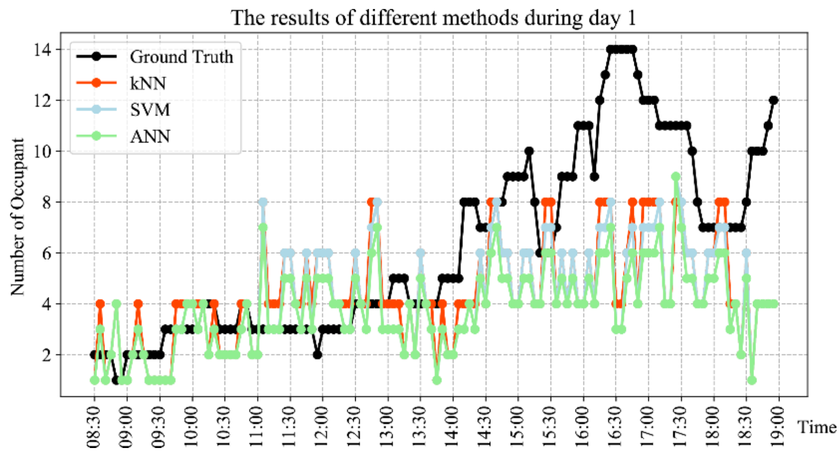


Fig. 6. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 1 using the Wi-Fi data

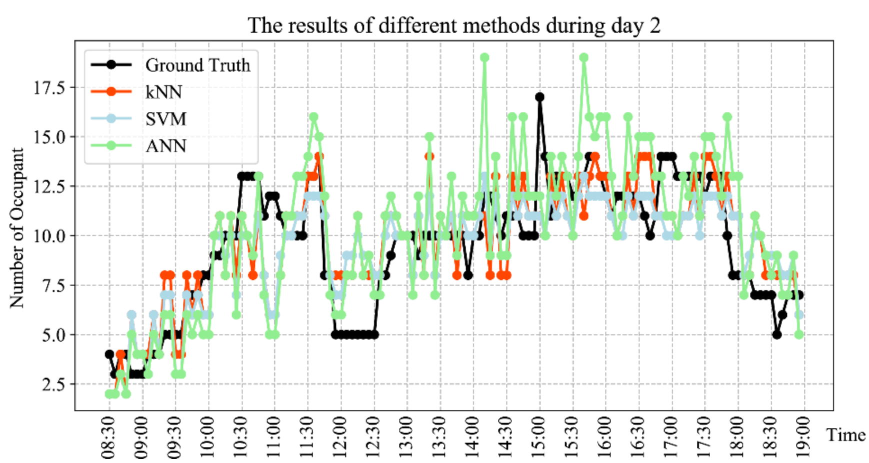


Fig. 7. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 2 using the Wi-Fi data

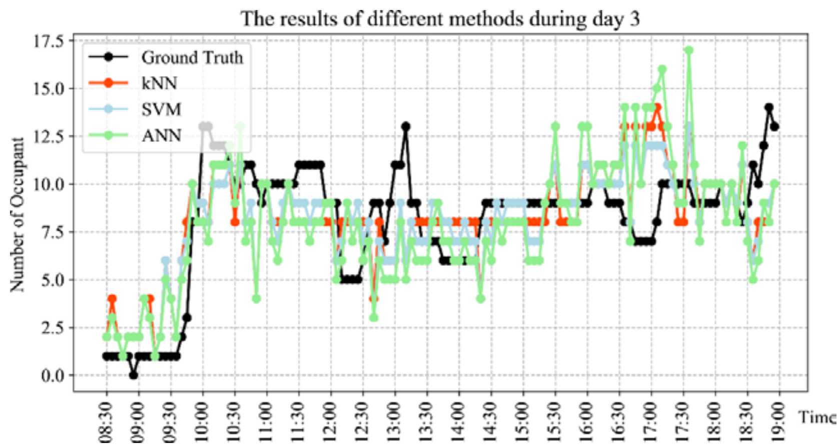


Fig. 8. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 3 using the Wi-Fi data

Table 3. The results of three occupancy models with MAE, MAPE, and RMSE using the Wi-Fi data

	kNN			SVM			ANN		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Day 1	2.7	45.5%	3.6	2.7	43.8%	3.7	2.8	39.5%	3.9
Day 2	1.8	22.7%	2.3	1.7	21.8%	2.2	2.3	24.9%	2.8
Day 3	2.1	42.0%	2.6	1.8	37.2%	2.4	2.3	37.0%	3.0
All 3 days	2.2	36.7%	2.9	2.1	34.3%	2.8	2.5	37.0%	3.3

5.3 Model performance with the combined environmental parameters and Wi-Fi data

Finally, this study tested three occupancy models by combining the environmental parameters and the Wi-Fi data to determine if the occupancy prediction accuracy could be improved. Figures 9, 10, and 11 show the results of occupancy profiles under the ground truth, kNN, SVM, and ANN models during three days using the combined environmental parameters and Wi-Fi data. Table 4 shows the assessment results of three occupancy models with MAE, MAPE, and RMSE.

Compared to the occupancy models using only the environmental parameters or only the Wi-Fi data, the ANN occupancy model provided the best performance, with lowest MAE, MAPE, and RMSE compared with kNN and SVM models on Day 1, 2, and 3. Especially on Day 1, the MAE is about 1.9, MAPE is about 27.2%, and RMSE is about 2.7, which are rather lower than those of kNN and SVM. Therefore, the ANN-based occupancy model performs the best in analyzing the combined environmental parameters and the Wi-Fi data.

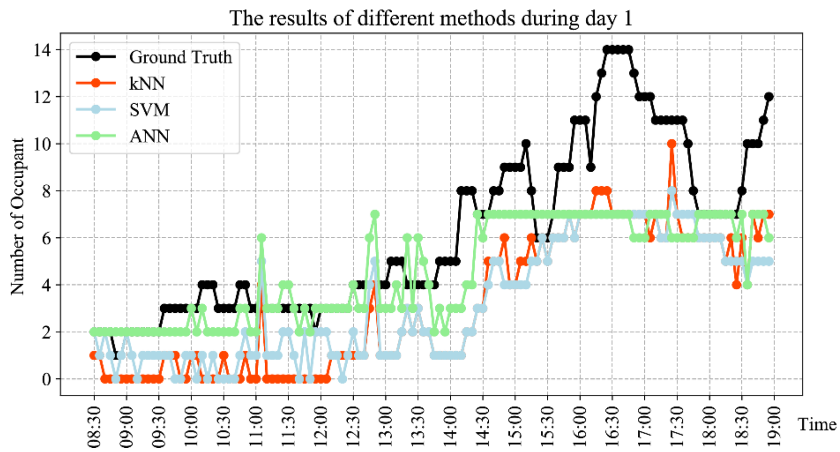


Fig. 9. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 1 using the combined environmental parameters and the Wi-Fi data

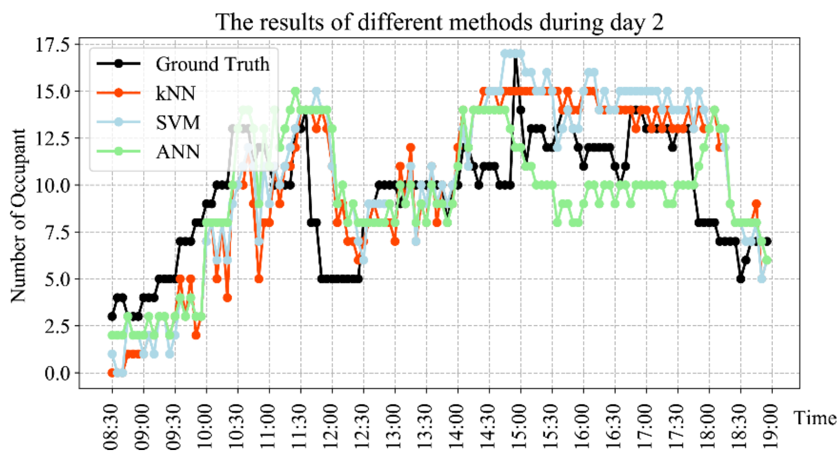


Fig. 10. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 2 using the combined environmental parameters and the Wi-Fi data

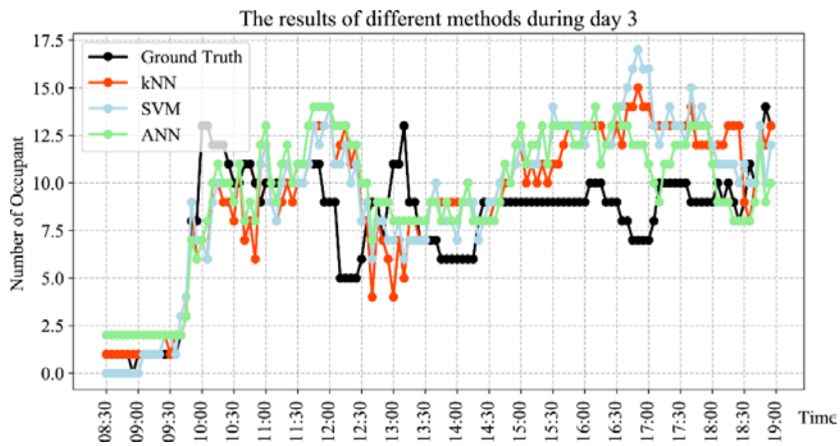


Fig. 11. Results of the occupancy profile under the kNN, SVM, and ANN models on Day 3 using the environmental parameters and the Wi-Fi data

Table 4. The results of three occupancy models with MAE, MAPE, and RMSE using the combined environmental parameters and the Wi-Fi data

	kNN			SVM			ANN		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Day 1	3.1	60.7%	3.5	2.9	51.8%	3.5	1.9	27.2%	2.7
Day 2	2.5	33.4%	3.1	2.6	33.4%	3.2	2.55	31.4%	3.1
Day 3	2.5	31.3%	3.2	2.6	36.7%	3.3	2.4	38.0%	3.0
All 3 days	2.7	41.7%	3.3	2.7	40.6%	3.3	2.3	32.2%	3.0

5.4 Performance evaluation of using three data groups

When evaluating data fusion of environmental parameters and Wi-Fi sensing, two questions must be answered. First, are occupancy results more accurate when using both datasets rather than using the environment dataset alone? Table 2 shows the occupancy results from using only the environmental parameters, and Table 4 shows results from using combined environmental parameters and Wi-Fi data. Both tables show approximately similar results for MAE and RMSE. However, occupancy results under both have a lower MAPE, the range of which varies from 30.9% to 43.5% observed from ANN model-based occupancy results under the environmental parameters. Therefore, the combination of datasets cannot improve accuracy significantly during three days.

Second, do accuracies increase when comparing occupancy results under the Wi-Fi dataset to those under both datasets? The overall three days' results show that using the combination of datasets under the ANN model decreased the accuracy, and the same results can be found on Day 2 and 3. However, on Day 1, the MAE results of occupancy are 2.7 and 1.9, respectively, under the Wi-Fi dataset and both datasets; MAPE results are 43.8% and 27.2%, respectively, under the Wi-Fi dataset and both datasets; and RMSE results are 43.8% and 27.2%, respectively under the Wi-Fi dataset and both datasets. Those results show that the combination of both datasets improve accuracy when the Wi-Fi sensing method did not work well.

On the other hand, three-day accuracy results using the Wi-Fi data in the occupancy model vary a lot; for example, using SVM, MAPE results vary from 21.8% to 43.8%, and RMSE results vary from 2.2 to 3.7, but the three-day accuracy results using both datasets are close to one another. Since the connections between Wi-Fi signals and building occupants are not very steady, adding environmental parameters into the Wi-Fi data can improve robustness of occupancy models, for example, on Day 1.

6. Discussion

Using environmental parameters to sense occupancy information is a popular way to acquire indirect occupancy data, and some studies have used environmental parameters to calculate occupancy information; for example, with a mass conservation equation [44, 45] or model occupancy information [20, 31], or with machine learning techniques, since the temperature, humidity, and CO_2 concentration are highly related to occupancy presence and count. However, the time delay of those parameters is an important drawback. On the other hand, as Wi-Fi signals are popularly used in buildings, Wi-Fi utilization varies with occupants as they carry a variety of mobile phones, computers, tablet, and wearables. A Wi-Fi sensing method is a direct occupancy acquisition because it learns occupancy information from the connections between Wi-Fi signals and users. Therefore, both indoor environmental parameters and Wi-Fi data are available and suitable for acquiring occupancy information.

This study used three feature-based occupancy models to learn occupancy by using both the indoor air parameters and Wi-Fi data. Two issues were addressed in this study: (1) which feature-based occupancy models perform better in occupancy prediction, and (2) whether or not combining the indoor environmental parameters and the Wi-Fi data improve the accuracy of occupancy prediction in feature-based occupancy models.

Combined with the MAE, MAPE, and RMSE results in Sections 5.1, 5.2, and 5.3, the first issue has been resolved. The ANN-based occupancy model can predict occupancy better than the kNN and SVM models, not only when using environmental parameters only, but also when using a combination of environmental parameters and Wi-Fi data. For the Wi-Fi data, the SVM model had the best performance.

This study can be regarded as one baseline case study for future feature-based occupancy studies in selecting machine learning models and inputs—especially in investigating the data fusion of environmental sensing and Wi-Fi sensing technologies, and how combining both technologies can improve the robustness of results. However, accuracy is not necessarily improved, and that issue should be further investigated. Also, more adaptions of other machine learning techniques should be studied. It should be noted that this study yielded some unsolved problems. The first important one is the occupancy sensing cost, which is a key issue in data collection and application. Although in feature-based occupancy models, Wi-Fi technology does not show an overwhelming advantage, and results on Day 1 show a lower accuracy of occupancy prediction, the cost of Wi-Fi technology is far lower than that of environmental sensors, since Wi-Fi technology has already been installed in most buildings. Second, for brevity, this study did not resolve how to reduce the sensing cost by using the minimal set of environmental sensors. For example, combining one or two of the environmental parameters with Wi-Fi data (such as CO_2 + Wi-Fi, T + Wi-Fi, RH + Wi-Fi, or CO_2 + T + Wi-Fi, and so on) could reduce costs. Neither issue was included in this study, but they are important, and should be addressed in future occupancy studies. Third, in previous studies, researchers have widely used environmental parameters to predict occupancy by considering that, when one space is occupied, environmental parameters of its indoor air will increase and vary with occupancy number (for example, the variation of CO_2 concentration level is highly related to occupancy). However, building operation, especially HVAC system operation, has a critical impact on indoor air parameters; especially indoor air temperature and relative humidity. Such two indoor air parameters usually are maintained at the thermal comfort range through the conditioned air supply. Therefore, the operational conditions of supply air should be considered; for example, how much conditioned air is supplied (and at what temperature) to remove the indoor thermal load and maintain an occupant's thermal comfort. In addition, if using environmental sensing, it is important to determine which key parameters to collect to better represent occupancy and show the best accuracy. Further, to apply data fusion for occupancy study in multi-size and multi-type rooms and generate the wider applicability of results should also be prioritized in the future. On the other hand, to validate the robustness of our method, the occupancy experiment can be run for a long period, e.g. one month or a whole year, in working days, weekend days, and holidays.

7. Conclusions

Occupancy prediction using machine techniques has captured growing attention in many studies by addressing the applications of environmental parameters (e.g., CO_2, temperature, and relative humidity) or the Wi-Fi technology. This study has investigated the performance of data fusion by combining the environmental parameters data and the Wi-Fi data. One on-site experiment was conducted in a large office room in City University of Hong Kong to collect indoor air CO_2 concentration level, temperature, relative humidity, and also Wi-Fi utilization data for three days. To learn and predict the occupancy information, this study selected three feature-based occupancy models using machine learning algorithms of kNN, SVM, and ANN. To assess the occupancy models, the MAE, MAPE, and RMSE indices were utilized.

Results showed that when only the environmental parameters data were applied to learn occupancy, the ANN-based occupancy model is more suitable and accurate, and so will be with the combination of environmental parameters and the Wi-Fi data. The SVM model is more suitable and accurate in learning occupancy information with Wi-Fi data. On the other hand, the combination of datasets cannot improve accuracy significantly during three days when compared with occupancy results under environmental parameters. The combination of the datasets decreases the accuracy under the Wi-Fi dataset. The three-day accuracy results using Wi-Fi data in the occupancy model vary a lot. For example, using SVM and MAPE results vary from 21.8% to 43.8%, and RMSE results vary from 2.2 to 3.7. Therefore, this study showed that the combination of both datasets can improve robustness of occupancy prediction. However, how to improve accuracy needs further investigation. Two potential methods are suggested that, (1) more features of environmental and Wi-Fi sensing datasets can be excavated, e.g. the change of sensed status and value or the change points of related parameters, to enhance and interfere occupancy prediction; (2) more occupancy prediction algorithms or improved kNN, SVM, and ANN algorithm can be built. Further, as in discussion, the occupancy sensing cost can also compromise with sensed parameters and accuracy, or how to reduce the cost with a higher accuracy should be an important and inspiring work pursued to achieve necessary improvement.

Acknowledgement

The work described in this paper was sponsored by the project JCYJ20150518163139952 of the Shenzhen Science and Technology Funding Programs and the National Natural Science Foundation of China (NSFC #51508487). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Science Technology and Innovation Committee of Shenzhen and NSFC. This work was also supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] T. Hong, D. Yan, S. D ’oca, C.-F. Chen, Ten questions concerning occupant behavior in buildings: The big picture, Build. Environ. 114 (2017) 518–530. doi:10.1016/j.buildenv.2016.12.006.

[2] T. Hong, Y. Chen, Z. Belafi, S. D’Oca, Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs, Build. Simul. 11 (2018) 1–14. doi:10.1007/s12273-017-0396-6.

[3] T. Hong, S. D’Oca, S.C. Taylor-Lange, W.J.N. Turner, Y. Chen, S.P. Corgnati, An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema, Build. Environ. 94 (2015) 196–205. doi:10.1016/J.BUILDENV.2015.08.006.

[4] T. Hong, H.-W. Lin, Occupant Behavior: Impact on Energy Use of Private Offices, in: Asim 2012, 1st Asia Conf. Int. Build. Perform. Simul. Assoc., 2013.

[5] I. Gaetani, P.-J. Hoes, J.L.M. Hensen, Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy, Energy Build. 121 (2016) 188–204. doi:10.1016/j.enbuild.2016.03.038.

[6] Xin Zhou, Da Yan, Guangwei Deng, Influence of occupant behavior on the efficiency of a district cooling system, in: Proc. BS2013 13th Conf. Int. Build. Perform. Simul. Assoc. Chamberym Fr. August 26-28, Chambery, 2013: pp. 1739–1745.

[7] W. Wang, J. Chen, Y. Lu, H.-H. Wei, Energy conservation through flexible HVAC management in large spaces: An IPS-based demand-driven control (IDC) system, Autom. Constr. 83 (2017) 91–107. doi:10.1016/J.AUTCON.2017.08.021.

[8] W. Wang, J. Chen, G. Huang, Y. Lu, Energy efficient HVAC control for an IPS-enabled large space in commercial buildings through dynamic spatial occupancy distribution, Appl. Energy. (2017). doi:10.1016/J.APENERGY.2017.06.060.

[9] W. Wang, J. Wang, J. Chen, G. Huang, Multi-zone Outdoor Air Coordination through Wi-Fi Probe-based Occupancy Sensing, Energy Build. (2017). doi:10.1016/J.ENBUILD.2017.11.041.

[10] S. Goyal, H.A. Ingley, P. Barooah, Occupancy-based zone-climate control for energy-efficient buildings: Complexity vs. performance, Appl. Energy. 106 (2013) 209–221. doi:10.1016/j.apenergy.2013.01.039.

[11] D. Bourgeois, I. Macdonald, Adding advanced behavioural models in whole building energy simulation: A study on the total energy impact of manual and automated lighting control, Energy Build. 38 (2006) 814–823. doi:10.1016/j.enbuild.2006.03.002.

[12] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, Autom. Constr. 24 (2012) 89–99. doi:10.1016/j.autcon.2012.02.013.

[13] M.S. Gul, S. Patidar, Understanding the energy consumption and occupancy of a

multi-purpose academic building, Energy Build. 87 (2015) 155–165. doi:10.1016/j.enbuild.2014.11.027.

[14] H. Rafsanjani, C. Ahn, M. Alahmad, A Review of Approaches for Sensing, Understanding, and Improving Occupancy-Related Energy-Use Behaviors in Commercial Buildings, Energies. 8 (2015) 10996–11029. doi:10.3390/en81010996.

[15] L.M. Candanedo, V. Feldheim, Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models, Energy Build. 112 (2016) 28–39. doi:10.1016/j.enbuild.2015.11.071.

[16] B. Dong, B. Andrews, K.P. Lam, M. Höynck, R. Zhang, Y.-S. Chiou, D. Benitez, An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy Build. 42 (2010) 1038–1046. doi:10.1016/j.enbuild.2010.01.016.

[17] Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz, A systematic approach to occupancy modeling in ambient sensor-rich buildings, Simulation. 90 (2014) 960–977. doi:10.1177/0037549713489918.

[18] Q. Zhu, Z. Chen, M.K. Masood, Y.C. Soh, Occupancy estimation with environmental sensing via non-iterative LRF feature learning in time and frequency domains, Energy Build. 141 (2017) 125–133. doi:10.1016/J.ENBUILD.2017.01.057.

[19] M.O. Z. Yang, N. Li, B. Becerik-Gerber, A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations, Proc. 2012 Symp. Simul. Archit. andUrban Des. Soc. Comput. Simul. Int. San Diego, CA,USA. (2012) 49–56.

[20] A. Szczurek, M. Maciejewska, T. Pietrucha, Occupancy determination based on time series of CO2 concentration, temperature and relative humidity, Energy Build. 147 (2017) 142–154. doi:10.1016/J.ENBUILD.2017.04.080.

[21] Z. Yang, B. Becerik-Gerber, Modeling personalized occupancy profiles for representing long term patterns by using ambient context, Build. Environ. 78

(2014) 23–35. doi:10.1016/j.buildenv.2014.04.003.

[22] M.M. Ouf, M.H. Issa, A. Azzouz, A.-M. Sadick, Effectiveness of using WiFi technologies to detect and predict building occupancy, Sustain. Build. 2 (2017) 7. doi:10.1051/sbuild/2017005.

[23] J. Chen, C. Ahn, Assessing occupants' energy load variation through existing wireless network infrastructure in commercial and educational buildings, Energy Build. 82 (2014) 540–549. doi:10.1016/j.enbuild.2014.07.053.

[24] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: Occupancy Based HVAC Actuation using Existing WiFi Infrastructure within Commercial Buildings, in: Conf. Proc. 11th ACM Conf. Embed. Networked Sens. Syst., 2013. doi:10.1145/2517351.2517370.

[25] W. Wang, J. Chen, X. Song, Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach, Build. Environ. 124 (2017) 130–142. doi:10.1016/J.BUILDENV.2017.08.003.

[26] W. Chung, Y.V. Hui, A study of energy efficiency of private office buildings in Hong Kong, Energy Build. 41 (2009) 696–701. doi:10.1016/j.enbuild.2009.02.001.

[27] D. Yan, W. O'Brien, T. Hong, X. Feng, H. Burak Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: Current state and future challenges, Energy Build. 107 (2015) 264–278. doi:10.1016/j.enbuild.2015.08.032.

[28] K. Sun, T. Hong, A simulation approach to estimate energy savings potential of occupant behavior measures, Energy Build. 136 (2017) 43–62. doi:10.1016/J.ENBUILD.2016.12.010.

[29] D. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, X. Feng, IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings, Energy Build. 156 (2017) 258–270. doi:10.1016/J.ENBUILD.2017.09.084.

[30] S. Wang, Dynamic simulation of building VAV air-conditioning system and evaluation of EMCS on-line control strategies, Build. Environ. 34 (1999) 681–705.

doi:10.1016/S0360-1323(98)00052-3.

[31] S. Wang, J. Burnett, H. Chong, Experimental Validation of CO2-Based Occupancy Detection for Demand-Controlled Ventilation, Indoor Built Environ. 8 (1999) 377–391. doi:10.1177/1420326X9900800605.

[32] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, Energy Build. 131 (2016) 132–141. doi:10.1016/j.enbuild.2016.09.002.

[33] R.H. Dodier, G.P. Henze, D.K. Tiller, X. Guo, Building occupancy detection through sensor belief networks, Energy Build. 38 (2006) 1033–1043. doi:10.1016/j.enbuild.2005.12.001.

[34] M. Liu, K. Qiu, F. Che, S. Li, B. Hussain, L. Wu, C. Patrick Yue, Towards indoor localization using Visible Light Communication for consumer electronic devices, in: 2014 IEEE/RSJ Int. Conf. Intell. Robot. Syst., IEEE, 2014: pp. 143–148. doi:10.1109/IROS.2014.6942553.

[35] K. Qiu, F. Zhang, M. Liu, Visible light communication-based indoor environment modeling and metric-free path planning, in: 2015 IEEE Int. Conf. Autom. Sci. Eng., IEEE, 2015: pp. 200–205. doi:10.1109/CoASE.2015.7294062.

[36] F. Manzoor, D. Linton, M. Loughlin, Occupancy Monitoring Using Passive RFID Technology for Efficient Building Lighting Control, in: 2012 Fourth Int. EURASIP Work. RFID Technol., IEEE, 2012: pp. 83–88. doi:10.1109/RFID.2012.10.

[37] J.A. Díaz, M.J. Jiménez, Experimental assessment of room occupancy patterns in an office building. Comparison of different approaches based on CO 2 concentrations and computer power consumption, Appl. Energy. 199 (2017) 121–141. doi:10.1016/j.apenergy.2017.04.082.

[38] T.H. Pedersen, K.U. Nielsen, S. Petersen, Method for room occupancy detection based on trajectory of indoor climate sensor data, Build. Environ. 115 (2017) 147–156. doi:10.1016/J.BUILDENV.2017.01.023.

[39] Q. Zhu, Z. Chen, M.K. Masood, Y.C. Soh, Occupancy estimation with

environmental sensing via non-iterative LRF feature learning in time and frequency domains, Energy Build. 141 (2017) 125–133. doi:10.1016/J.ENBUILD.2017.01.057.

[40] Z. Chen, M.K. Masood, Y.C. Soh, A fusion framework for occupancy estimation in office buildings based on environmental sensor data, Energy Build. 133 (2016) 790–798. doi:10.1016/J.ENBUILD.2016.10.030.

[41] J. Page, D. Robinson, N. Morel, J.-L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2008) 83–98. doi:10.1016/j.enbuild.2007.01.018.

[42] Y. Peng, A. Rysanek, Z. Nagy, A. Schlüter, Using machine learning techniques for occupancy-prediction-based cooling control in office buildings, Appl. Energy. 211 (2018) 1343–1358. doi:10.1016/J.APENERGY.2017.12.002.

[43] S.F. Lin, J.Y. Chen, H.X. Chao, Estimation of number of people in crowded scenes using perspective transformation, IEEE Trans. Syst. Man, Cybern. Part ASystems Humans. 31 (2001) 645–654. doi:10.1109/3468.983420.

[44] A. Szczurek, M. Maciejewska, T. Pietrucha, Occupancy determination based on time series of CO2 concentration, temperature and relative humidity, Energy Build. 147 (2017) 142–154. doi:10.1016/J.ENBUILD.2017.04.080.

[45] S.H. Ryu, H.J. Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Build. Environ. 107 (2016) 1–9. doi:10.1016/J.BUILDENV.2016.06.039.

[46] C. Martani, D. Lee, P. Robinson, R. Britter, C. Ratti, ENERNET: Studying the dynamic relationship between building occupancy and energy consumption, Energy Build. 47 (2012) 584–591. doi:10.1016/j.enbuild.2011.12.037.

[47] I. Bisio, F. Lavagetto, M. Marchese, A. Sciarrone, Smart probabilistic fingerprinting for WiFi-based indoor positioning with mobile devices, Pervasive Mob. Comput. 31 (2016) 107–123. doi:10.1016/j.pmcj.2016.02.001.

[48] Y. Wang, L. Shao, Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning, Build. Environ. 114

(2017) 106–117. doi:10.1016/j.buildenv.2016.12.015.

[49] R.S. Campos, L. Lovisolo, M.L.R. de Campos, Wi-Fi multi-floor indoor positioning considering architectural aspects and controlled computational complexity, Expert Syst. Appl. 41 (2014) 6211–6223. doi:10.1016/j.eswa.2014.04.011.

[50] Y. Wang, L. Shao, Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning, Build. Environ. 114 (2017) 106–117. doi:10.1016/j.buildenv.2016.12.015.

[51] H. Zou, Y. Zhou, H. Jiang, S.-C. Chien, L. Xie, C.J. Spanos, WinLight: A WiFi-based occupancy-driven lighting control system for smart building, Energy Build. 158 (2018) 924–938. doi:10.1016/J.ENBUILD.2017.09.001.

[52] F. Adib, D. Katabi, Wi-Vi: See Through Walls with Wi-Fi Signals, (2013). http://people.csail.mit.edu/fadel/wivi/index.html.

[53] R. Rana, B. Kusy, J. Wall, W. Hu, Novel activity classification and occupancy estimation methods for intelligent HVAC (heating, ventilation and air conditioning) systems, Energy. 93 (2015) 245–255. doi:10.1016/j.energy.2015.09.002.

[54] R. Rana, B. Kusy, R. Jurdak, J. Wall, W. Hu, Feasibility analysis of using humidex as an indoor thermal comfort predictor, Energy Build. 64 (2013) 17–25. doi:10.1016/j.enbuild.2013.04.019.

[55] S. Wang, X. Jin, CO2-Based Occupancy Detection for On-Line Outdoor Air Flow Control, Indoor Built Environ. 7 (1998) 165–181. doi:10.1159/000024577.

[56] S. Wang, J. Burnett, H. Chong, Experimental Validation of CO2-Based Occupancy Detection for Demand-Controlled Ventilation, http://dx.doi.org/10.1177/1420326X9900800605. (2016). doi:10.1177/1420326X9900800605.