Lawrence Berkeley National Laboratory

Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology

Wei Wang, City University if Hong Kong, LBNL

Jaiyu Chen, City University if Hong Kong

Tianzhen Hong, LBNL

Na Zhu, City University if Hong Kong

Energy Technologies Area

June 2018

Please cite DOI: **10.1016/j.buildenv.2018.04.034**

Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology

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

[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

Accurate occupancy prediction can improve building control and energy efficiency. In recent years, WiFi signals inside buildings have been widely adopted in occupancy and building energy studies. However, WiFi signals are easily disturbed by building components and the connections between users and WiFi signals are unstable. Meanwhile, occupancy information is often characterized stochastically and varies with time. To overcome such limitations, this study utilizes WiFi probe technology to actively scan the WiFi connection request and response between WiFi signal and smart devices in existing network infrastructures. The Markov based feedback recurrent neural network (M-FRNN) algorithm is proposed in modeling and predicting the occupancy profiles. One on-site experiment was conducted to collect ground truth data using camera-based occupancy sensors, which were used to validate the M-FRNN occupancy prediction model over a 9-day measurement period. From the results, the M-FRNN based occupancy model using WiFi probes shows best accuracy with a tolerance of 2, 3, and 4 occupants can reach 80.9%, 89.6%, and 93.9%, respectively. This study demonstrated WiFi data coupled with machine learning methods can provide valuable people count information to building control systems and thus improve building energy efficiency.

Keywords: building energy efficiency, occupancy prediction, WiFi probe, data analytics, M-FRNN algorithm, machine learning

[*]Corresponding author. tel.: +852 3442 4696; fax: +852 3442 0427.

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

1. INTRODUCTION

Buildings have received increasing attentions in recent years for their energy-saving potential. Building energy conservation requires effective facility system management and a good understanding of occupants' energy demand and buildings' capacity [1,2]. The energy consumption during the operation stage of a building is expected to be highly dependent on its estimated performance during the design stage. However, many studies suggested that actual energy performances of buildings severely deviates from their original design conditions due to incorrect assumption or estimation of occupancy behavior [3–6]. Significant discrepancies have been observed [7,8] due to complicated interrelationship of energy consumption in building facilities and occupancy behavior [9–11]. Masoso and Grobler studied the electricity consumption in an office building and found 56% was consumed during non-working hours due to leaving on lightings or other devices when offices were unoccupied [12]. Dong and Andrews [13] implemented sensor-based occupancy behavior model and prediction method in building energy and comfort management systems and simulations suggested potential energy saving of 30% when compared with other basic energy savings by HVAC control strategies. Wood and Newborough [14] found an energy reduction of 10% to 20% when studying the effect on real energy consumption by occupant feedback method. The International Energy Agency (IEA) Energy in Building and Community (EBC) Programme Annex 66 has highlighted and concluded the significant roles of occupant behavior in building performance study [15]. It emphasizes that occupant behavior is a key factor in evaluation of energy-saving technology by observing how occupants understand and interact with the technology when building is in use.

Therefore, researchers have developed various occupancy approaches to model and predict occupancy patterns of building. As the most widely implemented approach, CO_2 concentration based occupancy estimations have been used in some studies [16–18]. However, those approaches yield to some limitations, such as low sensitivity to large areas, latency in prediction and potentially high initial investment for installation. Developed in recent years, an alternative approach utilizes WiFi technology to automatically sense occupancy information based on existing network infrastructures [19,20]. However, WiFi signal is unstable and various building components, such as metal separations or concrete walls, can interfere with the signal. In addition, the occupancy information is usually stochastic and varies with time, which also cause the variation of WiFi signal utilization from the occupant side. For example, different occupants can carry different number of WiFi devices, which can also change between

time of the day or day of the week. The connection and disconnection between occupants and WiFi devices is also changing due to sleep mode of devices. Therefore, there is a challenge to use WiFi signal to directly infer occupancy information. To enhance the application of WiFi technology in low-cost and high-resolution occupancy estimation, this study proposes an occupancy prediction model using the Markov based feedforward artificial neural network (M-FRNN) algorithm to derive occupancy profiles in office buildings. To evaluate the performance of the proposed approach, an on-site experiment was conducted for nine days. During the experiment, a CO_2 concentration based occupancy approach was also implemented for comparison. Ground truth occupancy data was acquired from camera-based occupancy sensors.

2. BACKGROUND

In recent years, researchers recognized that building occupancy information played a critical role in energy consumption as well as the discrepancies between the designed/simulated energy performance and actual building energy consumption [8]. The occupant behavior in a building is closely related to its actual energy consumption during its operation stage [20–22]. Occupants can influence a building's energy consumption in three ways [21]: (1) the participating heat balance of the building through occupants' body heat release; (2) occupancy-based demand including thermal comfort and indoor air quality [19,23–25]; (3) occupant interactions with building systems and building controls [26–28]. Based on the occupancy assessment, Kim et al. developed a method to improve building energy simulation and significantly reduced the deviated plug-load estimation [29]. Chen et al. utilized occupancy information to visualize the impact of occupants' behavior on office buildings in the EnergyPlus simulation model [30]. Occupancy information can also be embedded into model predictive control (MPC) to save energy [10]. Occupancy information can be extended to more complicated energy-using behaviors of occupants. For example, Chen et al. [31] emulates occupants' energy consuming behaviors under peer pressure through peer network simulations. Lu et al. investigated the occupants' relapse behavior to assist decision-making in building retrofit projects [32]. Anna et al. proposed new human-based energy retrofits in residential buildings to integrate the post-occupancy information in the simulation [33]. Anna et al. also investigated the occupants' attitudes, education background, and perception in some thermal and energy need studies [34,35].

Therefore, an accurate prediction of occupancy in a building is the premise to improving building energy efficiency in the future [9]. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) recommends fixed

occupancy diversity factors in Standard 90.1-2007 [36] when actual occupancy information is unavailable. However, significant discrepancies were found given uncertainties in occupants' behaviors and occupancy patterns, which can result in an estimation error of up to 40% [37]. The book, Exploring Occupant Behaviors in Buildings, has concluded occupancy studies in terms of sensing and data acquisition, survey and laboratory approaches, and validations [38]. Passive infrared (PIR) sensors, movement sensors, and lighting sensors can respond to occupants' presence/absence within their field-of-view. PIR occupancy sensors is one of the most popular application in lighting controls, but these sensors are not able to predict stationary occupant [39]. Using movement sensors, the study [40] considered occupant presence as an inhomogeneous Markov chain and generated a time series of presence (absent or present) of each of a zone's occupants insides buildings. Studies also employ cameras to record building or room presence information and predict people count information [41–43]. Ahn and Park built an occupancy estimator using a camera to predict occupancy presence and simulated building energy performance under ASHRAE, Markov chain, and random walk models [44].

Another one widely applied medium is carbon dioxide (CO2) concentration in indoor spaces. For example, Wang et al. developed several dynamic CO_2-based models to estimate and predict occupancy in commercial buildings [16,45,46]. Researchers also proposed the inclusion of sensory data from other environmental sensors, such as temperature, humidity, lighting, and acoustic sensors, to improve occupancy prediction accuracy [47,48]. Yang and Becerik-Gerber formulated stochastic processes based on regression, time-series modeling, and pattern recognition modeling approaches to improve accuracy in occupancy prediction from a data analytic perspective [49]. Jiang et al. proposed a feature scaled extreme learning machine (FS-ELM) approach on CO_2 concentration to predict occupancy [17]. However, CO_2-based approaches have several constraints, such as low sensitivity to occupant mobility and slow response to drastic occupancy changes [50]. Diaz and Jimenez conducted an experiment on the power consumption of computers under occupancy variation estimated by CO_2 and the results suggested that CO_2 concentration is informative and expected to serve as a good indicator of occupancy [11]. Another popular stream of occupancy-sensing methods focuses on terminal-based wireless positioning sensors, such as radio-frequency identification (RFID). For example, Li et al. [51] reported an average positioning accuracy of 88% for stationary occupants and 62% for mobile occupants when using active RFID systems. However, RFID systems require occupants carry a dedicated receiver tag which needs additional investment in implementation in supporting

facilities and infrastructures. There is also concern of privacy with such wearable devices.

Alternatively, WiFi signal is a popular medium to predict occupancy information since WiFi signal is more efficient, affordable, and convenient inside buildings. Many researchers have proposed effective WiFi based occupancy approaches to adjust HVAC operation [52–55]. Because multiple WiFi networks are usually installed in most of modern buildings, the setup cost for the positioning network is low. Additionally, occupants' mobile phones can serve as signal tags by measuring the signal strength indicator (RSSI) and MAC address. For example, Balaji utilized existing WiFi infrastructure and smartphones to adjust HVAC operation, achieving 17.8% electricity savings [56]. Some researchers used WiFi data for occupancy sensing by directly counting connections between users and WiFi signals [57]. However, WiFi based occupancy prediction is subject to unstable signal due to building geometry and various interferences. The connection from users is also unstable as some users might not connect to WiFi signals since sleep mode of their devices, as well as users carry different number of mobile devices during a day.

To apply WiFi signal in improving occupancy prediction, this study conducted one on-site experiment in an office testbed to collect the WiFi signal connections through WiFi probe and proposed one occupancy prediction approach using the WiFi signal. One novel Markov-based feedback recurrent neural network (M-FRNN) algorithm is proposed to predict the occupancy information. The time-series, statistic, and stochastic characteristics, and the two-status of occupancy information are fully considered in the prediction model. Cameras were used to obtain ground truth of occupancy to assess the results of the M-FRNN approach. In addition, the CO_2 concentration based occupancy prediction was used to benchmark the performance of the proposed M-FRNN algorithm.

Main contributions of this study can be illustrated as:

(1) It proposed and validated a high-resolution and -accuracy occupancy prediction model, based on a machine learning algorithm M-FRNN, using the low-cost and widely available WiFi utilization data in an office field.

(2) It provides a reference of how we can utilize more occupancy characteristics in occupancy prediction model and improve accuracy.

(3)It provides insights for occupancy study through comparison of the CO_2 concentration based approach and the WiFi data approach.

3. METHODOLOGY

3.1 WiFi network infrastructure

Inside buildings, WiFi access points (APs) are installed to provide Internet services for building occupants by broadcasting WiFi signal around. When an occupant arrives in a building, smart devices carried by the occupant will scan the WiFi APs to look for the signal. The scanning has two types, active scanning (including direct scanning and broadcast scanning) and passive scanning [58]. During direct scanning, clients send a probe request intending to connect to a designated AP with a certain service set identifier (SSID). Under such settings, only the requested SSID AP can response to the probe request. During broadcast scanning, clients broadcast a probe request with a null SSID, which allows all APs in the zone to receive and respond to the probe request. During passive scanning, available APs act like a beacon and broadcast signals, and clients will decide whether to send a connection request. Under all scanning approaches, the request and response will be captured by the WiFi probe with a timestamp and the MAC address of each client. If the request log shows a client's device sent a connection request to one AP, the system infers that the client is within the range of WiFi probe sensing. Each MAC address is assumed to represent one unique client. Figure 1 shows a typical workflow of cooperative WiFi probe devices in the process of occupancy sensing.

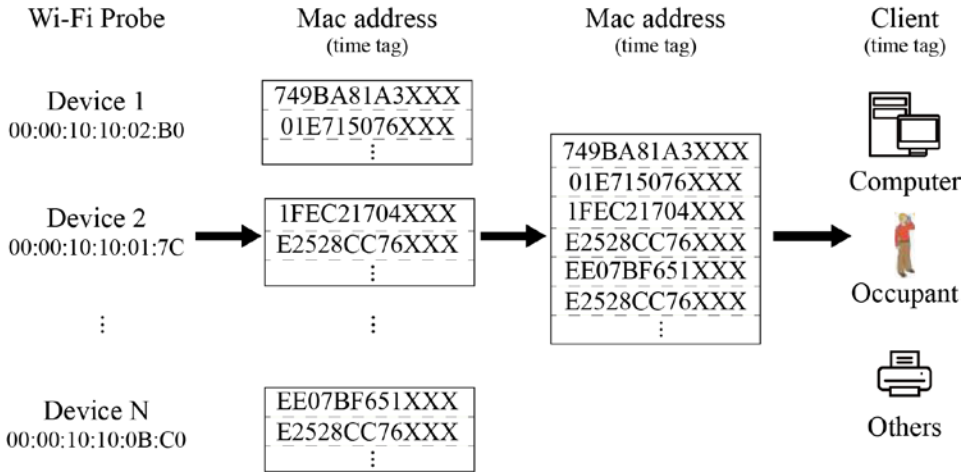

Figure 1. Workflow of occupancy sensing through cooperative WiFi probe devices.

As occupants may use multiple WiFi enabled devices, such as smart phones, laptops, wireless printers and other wearable devices, one MAC address does not imply one valid occupant. Therefore, to filter out invalid MAC addresses based on the signal patterns of the connection request, previous study adopted duration time functions to

pre-process the raw connection data [19] to figure out the raw occupant data from computers and other devices, which can be adopted in this study. However, the details about filtration will not be discussed in this study for brevity.

3.2 Characteristics of occupancy data

As a presentation of the utilization of building functions, occupancy contains unique features. Researchers have discussed and summarized these features into five characteristics [17,59–61]. (1) Time series characteristic. Occupancy is associated with time of a day and shows a periodic pattern. (2) Statistic characteristic. Historical occupancy data can infer the current occupancy pattern. (3) Stochastic characteristic. Current occupancy status is determined probabilistically according to the previous status. (4) Limited occupancy statuses. The occupancy status is simple and commonly contains two types, occupied or unoccupied (in or out). (5) Chronological distribution. The probability of occupant presence varies at different times of day. In addition, when occupants have interrelationships, such as being co-workers, the probability of their presence is a joint probability or conditional probability. Figure 2 illustrates the typical data format of the occupancy transfer probability matrix for the proposed model. In Figure 2, t_m and d_n stand for the *m* th and *n* th day in the dataset. The $V_{m,n}$ means the occupancy information at time interval *m* and day *n* and $P_{2,n}$ stands for the transfer probability from time interval *t2* to *t3* in *n* th day.

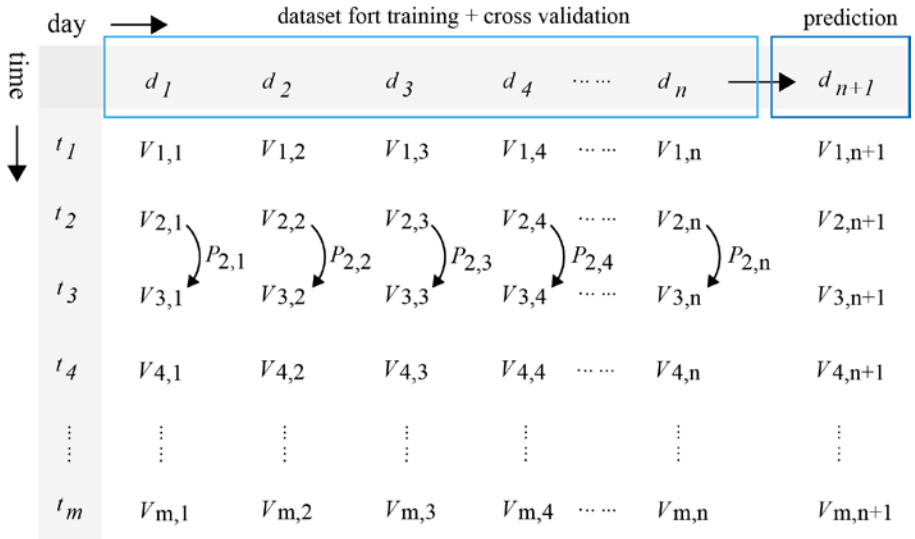

Figure 2. A typical probability correlation matrix of the proposed occupancy model

3.3 Markov-based Feedback Recurrent Neural Network (M-FRNN)

Artificial Neural Network (ANN) machine learning algorithms are a set of computational methods that mimic the human brain's problem-solving process to predict system outcomes through training and pattern learning. A typical ANN model has three network layers: The Input layer, the Hidden layer, and the Output layer. For occupancy prediction settings, the Input layer includes the captured MAC address of the current time interval. An input vector for MAC addresses at time t can be formatted as

$$X(t) = \{x_1, x_2, \dots, x_i, \dots, x_k\} \quad (1)$$

where, $X(t)$ is the collection of all MAC addresses at current time t.

k is the total number of all MAC addresses at current time t.

x_i means the *i* th detected MAC address.

The Hidden layer neutrons are calculated through input and network weights.

$$H(t) = W_1^T * X(t) + b \quad (2)$$

The Output layer converts the outcomes of Hidden layer and aggregates the results.

$$Y(t) = g(W_2^T * H(t)) = g(W_2^T * (W_1^T * X(t) + b)) \quad (3)$$

where, $H(t)$ and $Y(t)$ stand for the output of the Hidden layer and Output layer, respectively. W_1^T and W_2^T stand for the weights from the Input layer to the Hidden layer and from the Hidden layer to the Output layer, respectively. $g(\cdot)$ is the activation function of the Output layer and b is the random bias.

Classic ANN algorithms can effectively represent the time series and statistic characteristics of occupancy information, but are constrained regarding its stochastic, chronologically interdependent characteristics. Due to the unique electronic features of WiFi receivers, the outcome of occupancy prediction is subject to large and random fluctuations. For example, during WiFi probe sensing, discontinuous connection randomly appears as most mobile phones will activate sleep mode once they are without internet services. Short-term or one-time, unexpected visitors could also distort the prediction outcome with random bump ups. Such noises lead to latency in occupancy recognitions and deteriorate the prediction accuracy.

To solve such problems, this study developed a multi-variable Markov based Recurrent Neural Networks algorithm to build the occupancy behaviors prediction model. Figure 3 shows the structure of the proposed M-FRNN algorithm. In the proposed algorithm, an additional Context layer is added to capture the feedback and learn occupancy information from network iterations at a certain time interval. M-FRNN improves the classic ANN occupancy prediction model in three major areas.

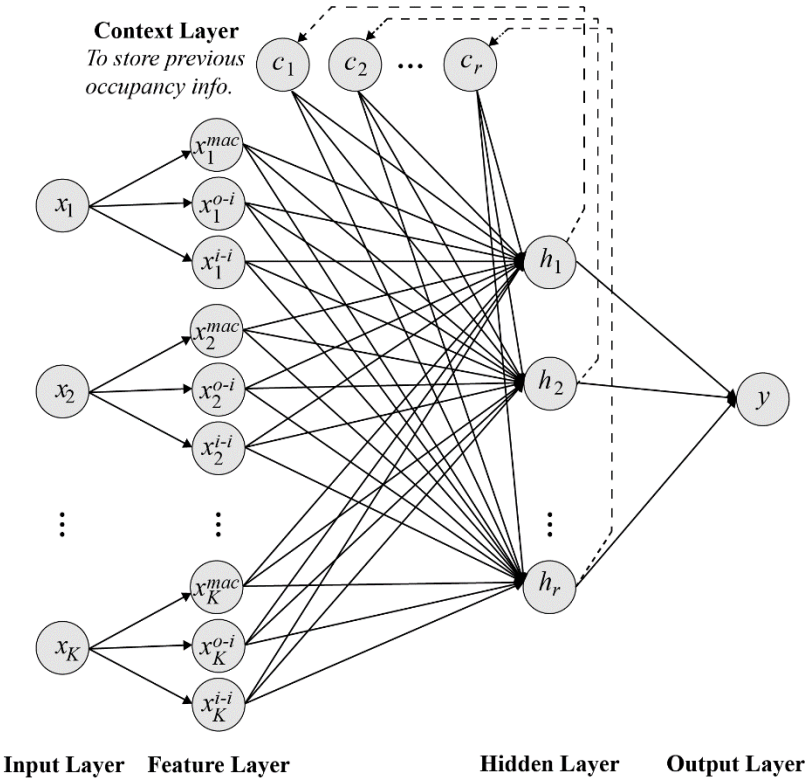

Figure 3. The structure of M-FRNN algorithm

First, the Feature layer is added between the Input layer and the Hidden layer to calculate the transfer probabilities of one MAC address in the Markov process. Since current occupancy status depends on previous occupancy status, the transfer probability and transfer probability matrix can be used to quantify such processes. Assuming there are only two statuses of an occupant in a space, which is "in" or "out", the transfer matrix can be defined as:

$$TPM|_{x_k} = \begin{bmatrix} x_k^{i-o} & x_k^{i-i} \\ x_k^{o-o} & x_k^{o-i} \end{bmatrix} \quad (4)$$

where $TPM|_{x_k}$ represents the transition probability matrix of one occupant x_k. In the transfer matrix, x_k^{i-o} and x_k^{i-i} denote the observed probability that one occupant whose status is "in" at the current time would still be "out" and "in" at the next time, respectively, at the next time. x_k^{o-o} and x_k^{o-i} denote the observed probability that

one occupant whose status is “out” at the current time interval would be “out” and “in” in the next time interval. The probability could be calculated by the observed conditional probability based on Bayesian models. For example,

$$x_k^{i-i} = P(observed\ state = \mathrm{i} | observed\ state = i) \tag{5}$$

Therefore, the occupied probability of one MAC address is

$$x_k^{i-i} = \frac{\sum N_{1-1}}{\sum N_{1-1} + \sum N_{1-0}} \qquad x_k^{o-o} = \frac{\sum N_{0-0}}{\sum N_{0-0} + \sum N_{0-1}} \tag{6}$$

where N_{i-i} is the frequency in which the occupancy status transitioned from “in” to “in” and N_{i-o} is the frequencies in which the occupancy status transitioned from “in” to “out” respectively. Similarly, N_{o-o} and N_{o-i} represent the frequencies in which the occupancy status transitioned from “out” to “out” and from “out” to “in” respectively.

Secondly, the training dataset in occupancy will be automatically updated, making the calculated frequency dynamic. When the training dataset is updated, the transfer probabilities are also updated in the following assessment. With an assigned probability for each MAC address in the room, each MAC address can be formatted as

$$x_k = \left\{x_k^{Mac}, x_k^{o-i}, x_k^{i-i}\right\} \tag{7}$$

Then, the input vector will be updated as

$$X(t) = \left\{x_1^{Mac}, x_1^{o-i}, x_1^{i-i}, x_2^{Mac}, x_2^{o-i}, x_2^{i-i}, \dots, x_k^{Mac}, x_k^{o-i}, x_k^{i-i}\right\} \tag{8}$$

Third, time windows are applied for dynamic modelling, as illustrated in Figure 4. The feature layer can then be formatted as

$$F(t) = \{X(t), X(t-1), X(t-2), \dots., X(t-\Delta t)\} \tag{9}$$

where, Δt is the length of time window ($\Delta \mathrm{t} = 5$ in this study) and $F(t)$ is the vector of the Feature layer at time t. Supposing the number of MAC addresses in the time window is K, then

$$F(t) = \left\{x_1^{Mac}, x_1^{o-i}, x_1^{i-i}, x_2^{Mac}, x_2^{o-i}, x_2^{i-i}, \dots, x_K^{Mac}, x_K^{o-i}, x_K^{i-i}\right\} \tag{10}$$

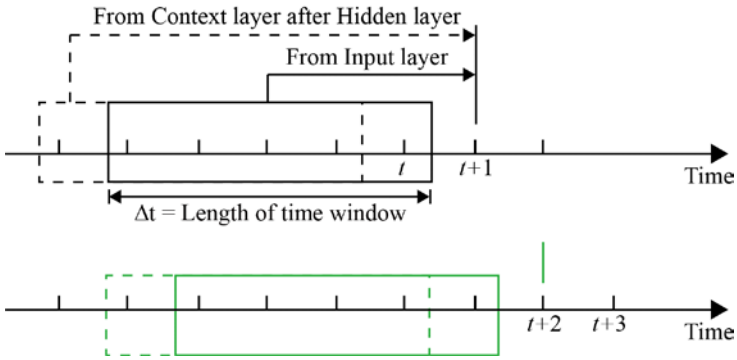

Figure 4. The illustration of time window method in occupancy model.

As the Context layer stores the feedback signals for the Hidden layer in the next interval, it serves as a short-term memory to highlight occupancy interdependencies. Then, the output of the Hidden layer can be formatted as

$$H(t) = f(\omega^1 C(t-1) + \omega^2(F(t))) \tag{11}$$

The output of the Context layer is

$$C(t-1) = \alpha C(t-2) + H(t-1) \tag{12}$$

Where $H(t)$ is the output vector of the Hidden layer at time interval t, and C is the output vector of Context layer. ω^1 is the connection weight from the Context layer to the Hidden layer, and ω^2 is the connection weight from the Feature layer to the Hidden layer. α is the self-connected feedback gain factor ($\alpha = 0$ in this study). $f(\cdot)$ is the activation function of the Hidden layer. In this study, the function is selected as

$$f(x) = \frac{1}{1+e^{-x}} \tag{13}$$

Signal transition from the Hidden layer to the Output layer could be formatted as:

$$y(t) = \omega^3 H(t) = \omega^3 * f(\omega^1 C(t-1) + \omega^2(F(t))) \tag{14}$$

Where, the y(t) is the output variable at time t, which is the predicted occupancy in this study. ω^3 is the connection weight from the Hidden layer to the Output layer.

The cost function to update and learn the connection weights is formulated as

$$E = \sum_{t=1}^{N}[y(t) - d(t)]^2 \tag{15}$$

Where N is the size of training time samples and $d(t)$ is the actual occupancy output.

4. ON-SITE EXPERIMENT

4.1 Experiment testbed

To validate the proposed occupancy prediction method, the research team conducted a two-week experiment using a graduate student office room located in City University of Hong Kong. The office has an area of about 200 m^2 and 25 long-term residents during the experiment period. Fig. 5 shows the space layout and equipment setup of the testbed. The office has two entrances but no window. Inside the room, the dedicated outdoor air system (DOAS) is equipped to ventilate outdoor air to indoor area without air handling process through 24 hours. Indoor air is conditioned by the fan coil unit (FCU) with variable refrigerant flow and therefore, indoor air circulation is driven by positive pressure. On the other hand, this office is covered by several university WiFi signals installed by City University of Hong Kong, which contributes to WiFi utilization data collection in this office room.

4.2 Sensors installation and data collection

During the experiment, TA465-X (environmental sensors produced by TSI Company) were utilized to monitor and record the indoor air temperature, relative humidity, and CO_2 concentration, as shown in Fig. 5. Since DOAS and FCU systems constitute the indoor air circulation and conditioning, the CO_2 concentration of return air of FCU can be approximately represented by CO_2 concentration of the indoor air after air mixing. To eliminate the uneven air mixing, we installed three environmental sensors, which are evenly distributed insides this office room. Air flow meters were installed near outdoor inlets to monitor the air flow rate of the ventilation system. Ground truth of occupancy data is acquired by two overhead cameras installed to record the entrance and exit events of occupants. Table 1 shows the specifications of the installed sensors, which include cost for purchase, measurement variables, data storage types, sensing intervals, range, accuracy, and resolution of each sensor during experiment. The measurement duration is from 09 Sep 2017 to 23 Sep 2017. The occupancy prediction model is scheduled from 09:00 to 18:30.

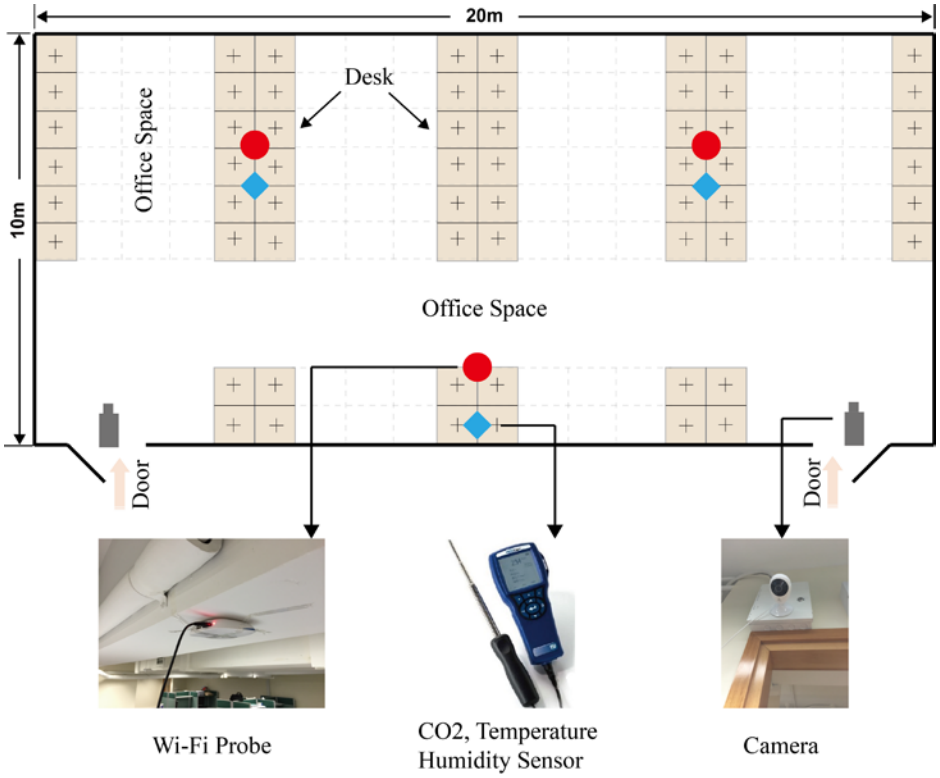

Figure 5. Space layout and equipment setup

Table 1. Sensors used in the experiment.

Sensors	Camera	WiFi Probe	Environmental 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, CO2, Air flow rate, Air pressure, CO			
Data Storage	Online	Online	Local			
Sensing interval		30s	1min	1min	1min	
Range			0 – 5k ppm	14 - 140 °F -10 – 60 °C	0 to 95%	
Accuracy			±3% or ±50 ppm	±0.5°F (±0.3°C)	< 3%	
Resolution			1 ppm	0.1°F (0.1°C)	0.10%	

4.4 Data processing

4.3.1 Ground truth

In this study, two cameras were installed. Since sensing time intervals of sensors were different, two steps were applied in this study to keep sensing data and ground truth consistent. Firstly, we need to obtain the entrance and exit of two doors from videos at any time and generate the occupancy ground truth for a day. Secondly, the WiFi and environmental sensors can be synchronized in one-minute raw data and then we calculate the number of occupants from ground truth at same sample time as the WiFi probe and CO2 concentration.

4.3.2 CO_2-based occupancy model for comparison

A CO_2-based occupant accounting approach was adopted as a comparison. Based on the ASHRAE standard's recommendation, this study assumes that (1) CO_2 is only generated by occupants' metabolism and outdoor air ventilation, (2) the occupant generated CO2 (*S*) at a constant speed, and (3) the air supplied to the space is assumed to be well-mixed. The time variation of CO_2 concentration levels in one zone can be calculated with a mass balance equation

$$V_{room}\frac{\partial C_z(t)}{\partial t} = V_{sa}C_{sa} + P_z * S - V_{ra}C_{ra} - V_{oa}C_{ra} \tag{16}$$

While in the air-handling unit (AHU), mass balance of CO_2 yields to

$$V_{oa}C_o + V_{ra}C_{ra} = V_{sa}C_{sa} \tag{17}$$

Where, V_{room} is the volume of room. C_z is the indoor CO_2 concentration, C_{sa} is the CO_2 concentration of supply air, and C_{ra} is the CO_2 concentration at the return duct level. V_{sa} is the supply air volume, V_{ra} is the return air volume, and V_{oa} is the outdoor air volume. P_z is the number of occupants and S is the CO_2 generation rate of per occupant. When the CO_2 concentration at the return air duct is assumed to be the same as the CO_2 concentration of the indoor air at breathing level, then

$$V_{room}\frac{\partial C_z(t)}{\partial t} = V_{oa}C_0 + P_z * S - V_{oa} * C_z \tag{18}$$

Therefore, we could calculate:

$$P_z = \frac{(V_{oa}^k + V_{oa}^{k-1}) * (C_{ra}^k - C_o^k)}{2S} + \mathrm{V} * \frac{(C_{ra}^k - C_{ra}^{k-1})}{S * \Delta t} \tag{19}$$

The superscript k denotes the index of the time interval and Δt is the data resolution of time. Since the air in the room is well-mixed and occupants generate CO_2 at a constant speed, CO_2 concentration is an obvious indicator of human presence and the number of occupants in a space.

After measurement, interpolation was used to make up default value of outdoor air flow rate and indoor air CO_2 concentration by averaging the around two sample data. For Eq.19, the generation rate (S) of an occupant can refer to ASHARE Standard 62 [62]. The CO_2 concentration of return air (C_{ra}) was replaced by indoor air CO_2 concentration measurement and was inputted by averaging measurement data from three CO_2 sensors. For brevity, the CO_2 concentration sources from filtration were ignored in this study since filtration is difficult to measure. The office room does not have windows so infiltration is expected to be low.

4.3.3 Model configuration

Fig. 6 shows an overview of this study. In the proposed occupancy prediction model, the raw data of Mac address of occupants' smart devices can be illustrated using the duration time filter method [19]. Since the core of this study focused on occupancy prediction method with WiFi data, the details about how to filter Mac addresses were not investigated in this study. In M-FRNN, for the data in input layer, Eq. 4 to 7 were the key step to excavate the features of raw data of Mac addresses and calculate the occupancy frequencies from "out" to "in" and from "in" to "in". The 30-min moving time window method adapted to time series characteristic of occupancy and generated feature layer dataset. The final occupancy prediction profiles were illustrated by 5-min resolution. Therefore, each time window includes 6 occupancy data. In results, Markov chain model based on Eq. 4, 5, and 6 was used to compare M-FRNN model by summing the transfer probabilities from "out" and "in" to "in". On the other hand, results from CO2-based occupancy profiles were also applied to compare the results. However, during the experiment, outdoor air system operates during the 24 hours before the first arriving of the office. To eliminate such error, some researchers use CO_2 concentration increase, such as 50ppm, as an effective indicator of human presence [13]. Therefore, this study assumes the presence of occupants if the CO_2 concentration of indoor air is 50ppm higher than that of the outdoor air. Finally, two types of CO_2 concentration based occupancy profiles were used as comparisons. First type was to use measured CO_2 concentration to calculate occupancy profile and second type considered the first arriving time of occupancy information by calculating occupancy when CO_2 concentration is over 450 ppm (CO_2 concentration of outdoor air was set as 400 ppm).

For each occupancy profile, performance criteria were whether the predicted occupancy profile successfully matches the occupancy profile of ground truth evaluated by four assessment indices.

4.4 Assessment of occupancy prediction

During the experiment, the actual occupancy of the room was acquired through manual video analysis of the camera recordings. To assess the occupancy prediction, four indices were used to compare the results with actual occupancy.

(1) Mean Absolute Error (MAE) compares the mean error between the occupant counts in a zone and can be defined as

$$MAE(O^p) = \frac{1}{N}\sum_{i=1}^{N}\left|O_i^{ob} - O_i^p\right| \tag{20}$$

(2) Mean Absolute Percentage Error (MAPE) shows the mean percentage error between the predicted occupant count and the actual number of occupants.

$$MAPE(O^p) = \frac{1}{N}\sum_{i=1}^{N}\left|(O_i^{ob} - O_i^p)/O_i^{ob}\right| \tag{21}$$

(3) Root Mean Squared Error (RMSE) shows the magnitude of the estimation error.

$$CVRMSE(O^p) := \frac{\sqrt{\sum_{n=1}^{N}(O^{ob}-O^p)^2/N}}{\sum_{n=1}^{N}O^{ob}/N} \tag{22}$$

(4) X-accuracy presents the accuracy when a tolerance x is allowed between the predicted occupancy and actual occupancy. The tolerance x allows for wrong estimation of the number of occupants. For example, the 2-accuracy (x=2) tolerance means that when the number of wrongly estimated occupants is less than 2, the estimation is regarded as correct.

$$\tau(O^p, x) = \frac{\sum_{i=1}^{N} X\left(\left|O_i^{ob}-O_i^p\right|,x\right)}{N} \tag{23}$$

Where,

$$X(|O_i^{ob} - O_i^p|, x) = \begin{cases} 1, & if\,|O_i^{ob} - O_i^p| < x \\ 0, & otherwise. \end{cases} \quad (24)$$

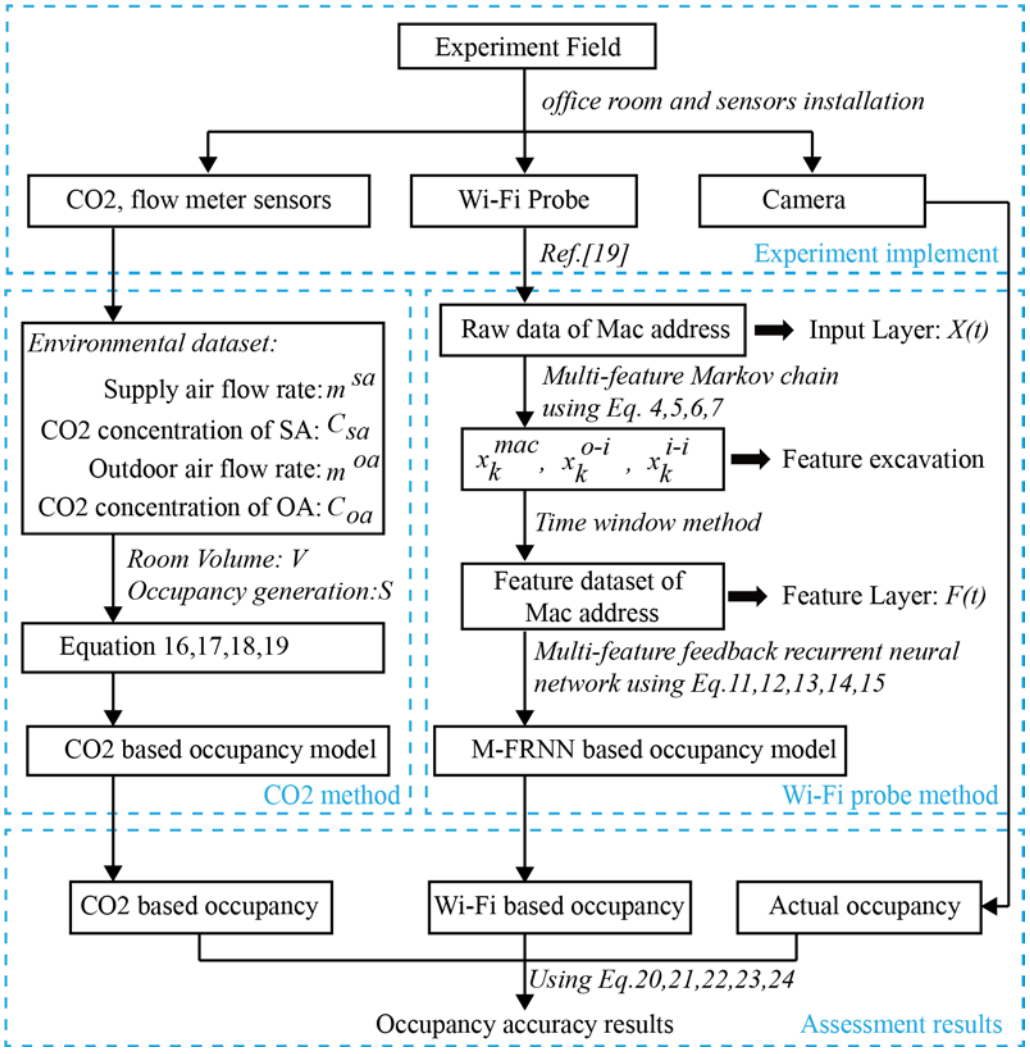

Figure. 6. The flowchart of this study.

5. RESULTS AND VALIDATION

5.1 Data analysis

Figures 7 to 9 show the raw results of the experiment. There are 4 outdoor air inlets and 8 supply air outlets in the experiment. It could be found from Figure 7 that the outdoor air supply flow rate is 180 cfm for each outdoor air inlet consistently, while the supply air flow rate for each supply air inlet is over 300 but less than 400 cfm most of the time. During the experiment, outdoor air was supplied uninterrupted during the night even if the cooling services from supply air terminals were closed by users.

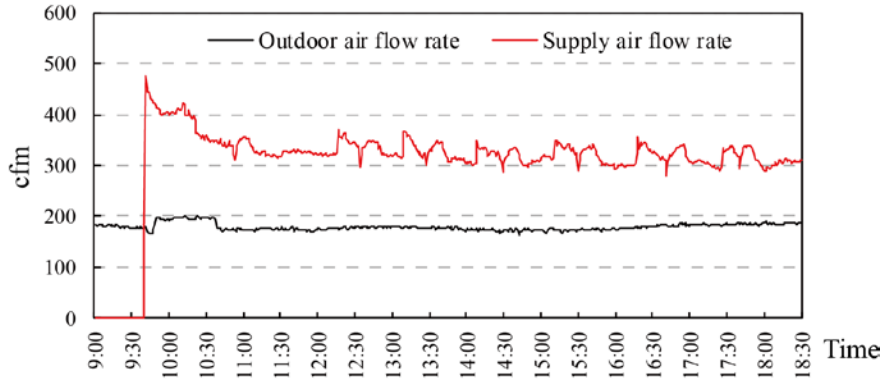

Figure 7. The flow rate of outdoor air and supply air over one day.

Figure 8 shows the CO_2 concentration variation of outdoor air and indoor air. It shows the CO2 concentration of outdoor air is constant around 400 ppm, which is used in the CO2 concentration based occupancy model. The indoor air's CO2 concentration is averaged from data measured by all CO2 sensors. Figure 8 shows that the pattern of CO2 concentration varies similar to the occupancy pattern recommended in ASHRAE standard 90.1-2007 [36]. At noon, the value drops dramatically due to occupants' lunch break. Since MAC addresses were taken as the identities of occupants, Figure 9 shows the total number of MAC addresses during experiment period from 09/10/2017 to 09/23/2017 based on the WiFi probe sensing results. In the experiment, the MAC addresses mainly belong to phones and computers. To filter the results, only MAC address with a duration time over 30 minutes were taken into consideration [19]. The figure shows the number of MAC addresses appearing in the experiment filed in weekdays was more than on weekend days.

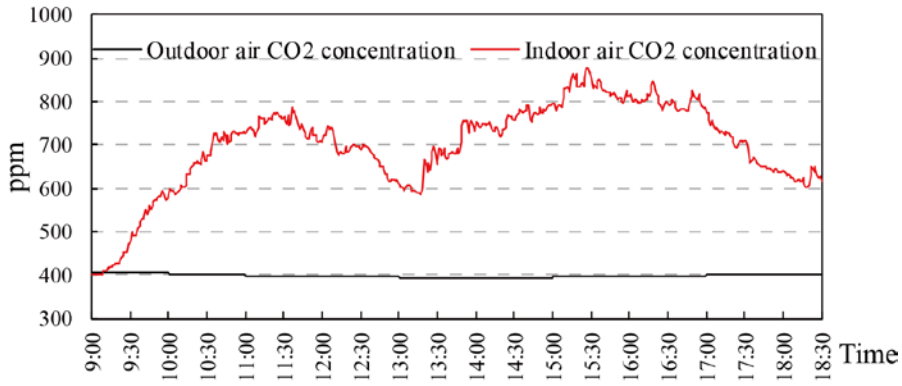

Figure 8. The CO_2 concentration of outdoor air and supply air in one day.

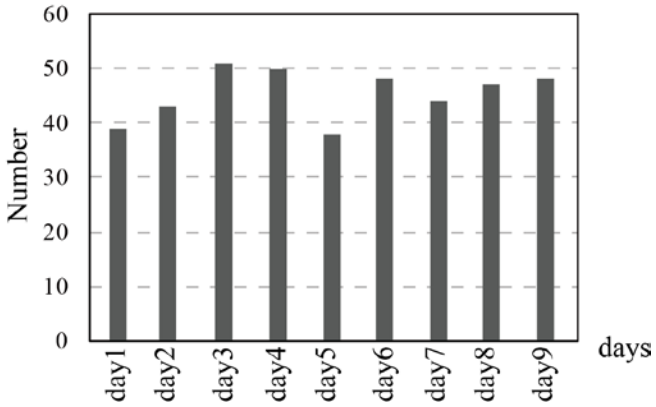

Figure 9. The total number of MAC addresses for each day.

5.2 Results for occupancy prediction and comparison

This section summarizes the occupancy profile results using both the WiFi probe and CO2 concentration approaches. During the experiment, the sensors, including the WiFi probes and cameras, required internet connectivity to upload the data in real time, while the data from the environmental sensors needed to be downloaded manually. Since then, the processed results excluded the experiment days with discontinuous monitoring and unqualified results. Moreover, WiFi probes, CO2 sensors and cameras have different measurement timesteps. Finally, to compare results, the valid experiment period includes 9 days and the 5-min resolution occupancy profiles during office hours (from 09:00 am to 18:30 pm) were estimated. The results beyond the decimal point have been dealt and shows using integer points as count of occupants should be integer. Figures 10-12 show the occupancy profiles based on camera footage, CO2 concentrations, and the WiFi probe.

In the CO2-based occupancy prediction, the outcomes were significantly influenced by the CO2 concentration of indoor air and the flow rate of outdoor air. As the outdoor air was constantly supplied through the whole day, the CO2 concentration of indoor air was close to that of outdoor air at the beginning of a day. Therefore, the occupancy counts based on CO2 concentration method were unreasonably high at the beginning of the day. It can be seen from Figures 10-12 that the occupancy measured with CO2 concentration methods (Occupancy_CO2, the orange line) is very high at the start of all nine days, even when the room was unoccupied around 9:00 am. The results with the improved CO2 concentration based occupancy model (Occupancy_CO2_50ppm, the blue line) are presented in Figures 10-12. Based on the results, it could be found that the accuracy of the occupancy model at the beginning of each day was greatly improved.

In WiFi probe based occupancy results, two occupancy models were compared: the classic Markov model and the Markov-based FRNN model. The classic Markov model introduced stochastic characteristics of occupancy and recognized the occupancy information using transfer probabilities. The Markov-based FRNN model also considered the time series characteristics and interdependency. Figures 10-12 present the WiFi probe based occupancy prediction results, including results from the classic Markov model (Occupancy_WiFi_Markov, the grey line) and the Markov-based FRNN model (Occupancy_WiFi_M-FRNN, the red line). Since the results of Occupancy_WiFi_Markov is more undulatory than that of Occupancy_WiFi_M-FRNN, the proposed is more preferable as the input for building facility systems, such as HVAC systems, lighting systems, etc.

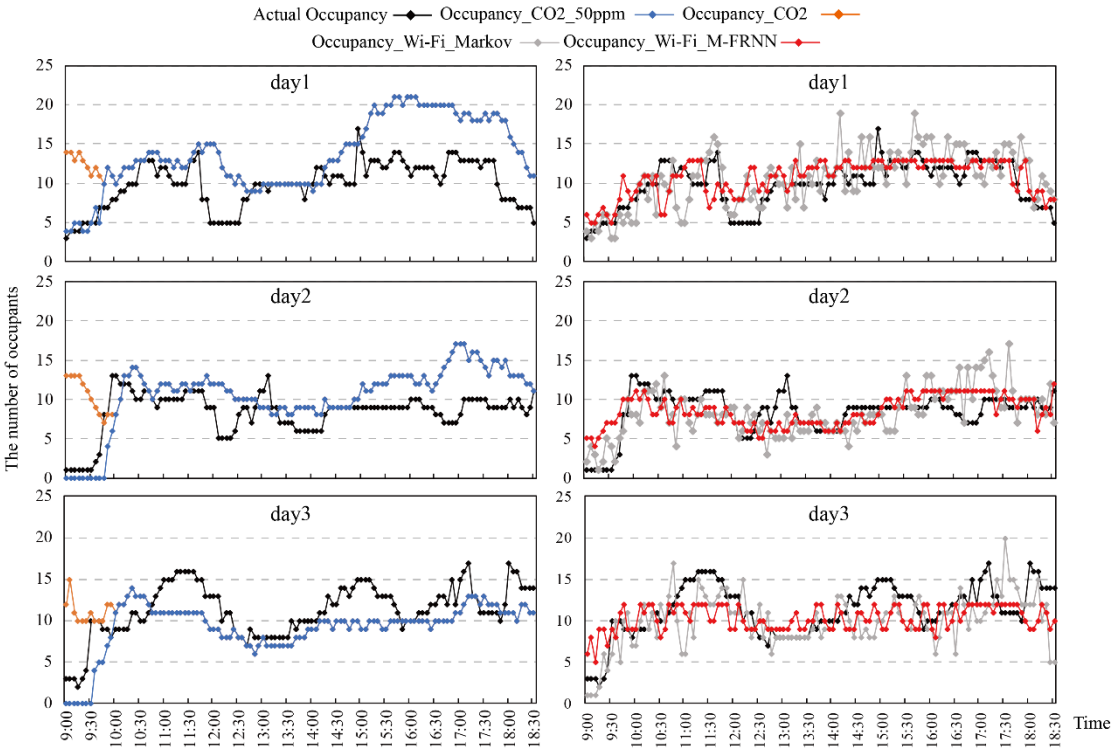

Figure 10. The occupancy results found using each approach from day 1 to day 3.

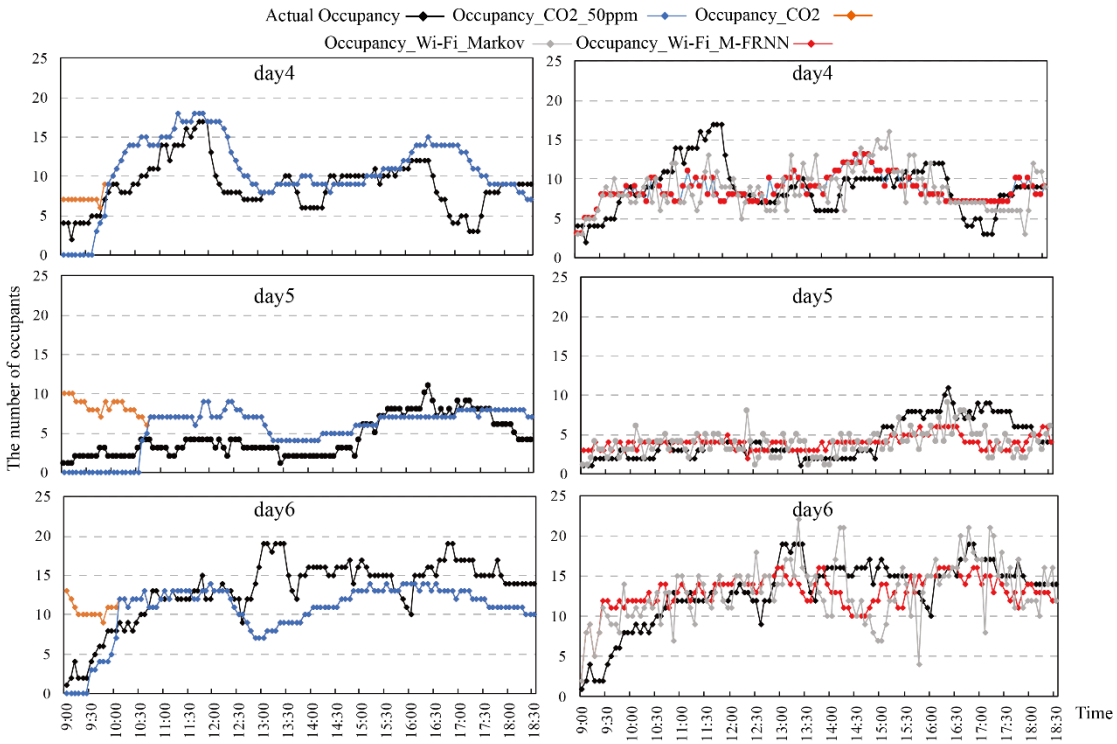

Figure 11. The occupancy results found using each approach from day 4 to day 6.

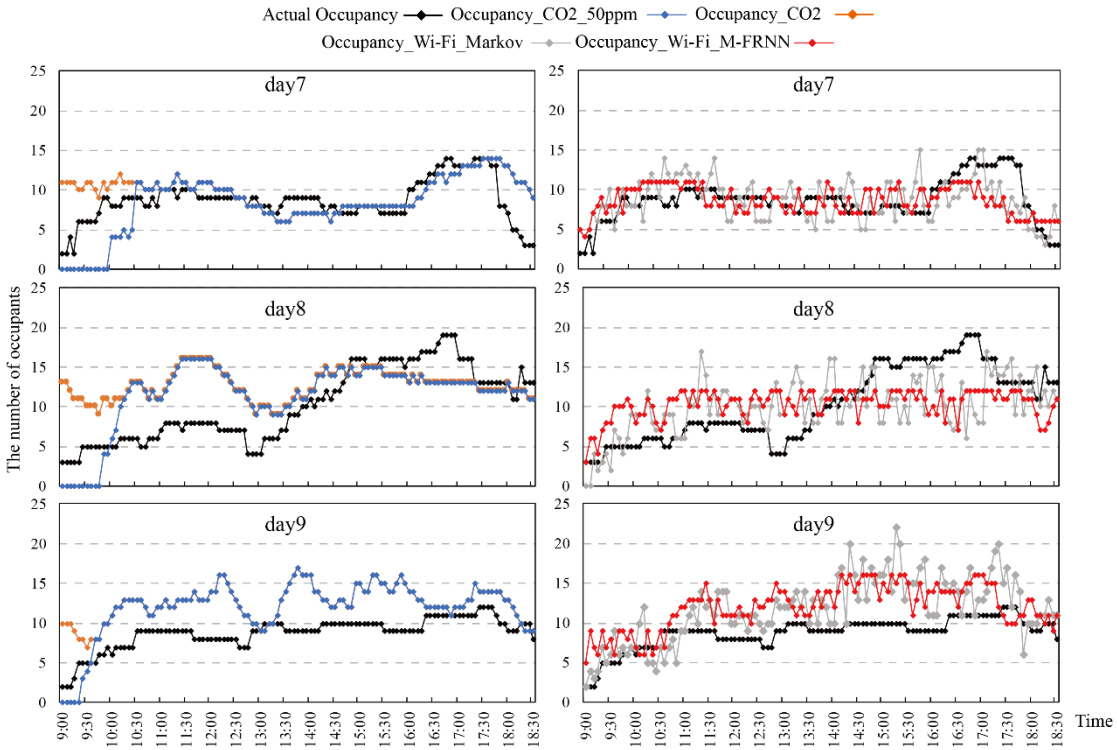

Figure 12. The occupancy results found using each approach from day 7 to day 9.

Figure 13 shows the x-accuracy of all three occupancy models. The x-accuracy of improved CO2-based occupancy model shows over 80% accuracy on 5 days (x=4), 2 days (x=5), 1 day (x=6), and 1 day (x=7). The best performance appears on day 5 and

the maximum number of occupants is 11, while the accuracy can still be 93.9% if 4 error count is allowed.

The WiFi probe based Markov models is over 80% of accuracy for x=3 (1 day), x=4 (5 days), x=5 (1 day), and x=6 (2 days). Similar to the CO2 model, the accuracy on day 5 can reach 94.8% if 4 error count is tolerant. The results also reveal that the M-FRNN model can greatly improve the prediction accuracy, 80% of accuracy on x=2 (1 day), x=3 (4 days), x=4 (2 days) and x=5 (2 days).

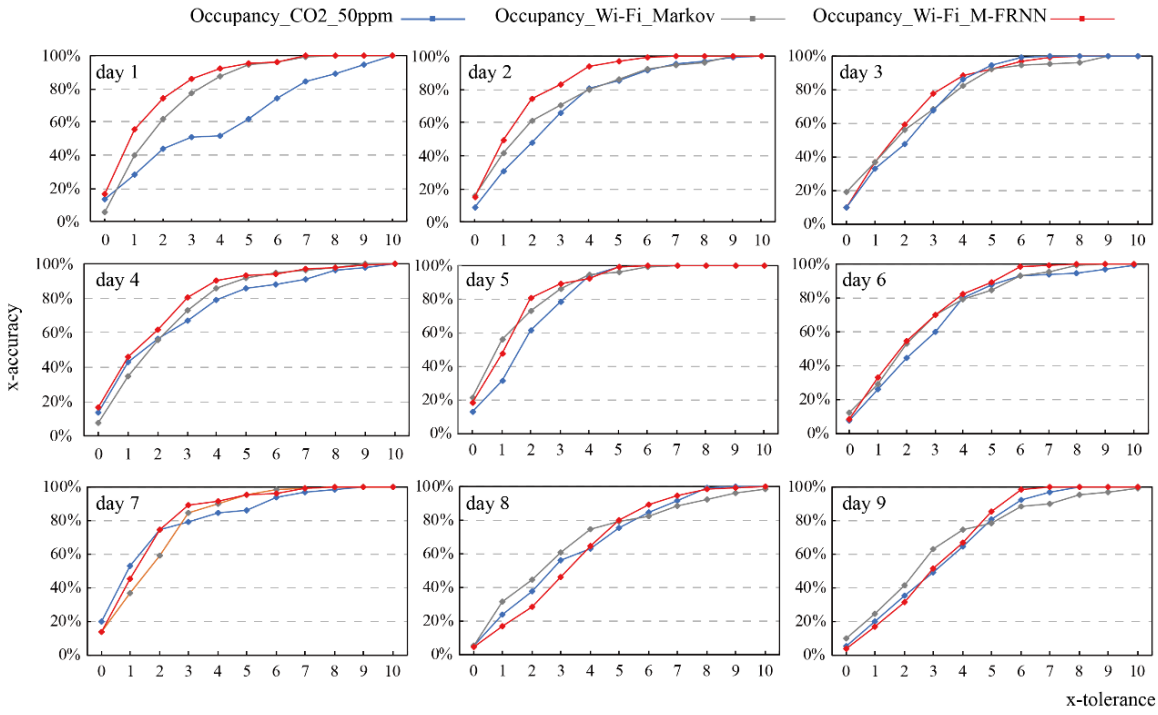

Figure 13. The results of x-accuracy for each approach over 9 days

Table 2 summaries the assessment indices of each occupancy approach during the experiment period, including the max number of occupants, MAE, MAPE, and CVRMSE. It shows that days 5 has the worst prediction performance for all three occupancy models. The proposed WiFi based M-FRNN occupancy model shows higher accuracy in terms of MAE, MAPE, and CVRMSE when compared with CO2-based occupancy model. When compared with the classic Markov model using WiFi probe sensing, the proposed M-FRNN also has a better performance in terms of the MAE, MAPE and CVRMSE. The mean error average of proposed occupancy prediction with WiFi data is lowest among the occupancy models. For example, in day1, the MAE of M-FRNN model is only 1.82 (the number of maximum occupants is 17) while MAE results of CO2 based method and Markov chain based method are 3.99 and 2.3, respectively. Considering MAPE and CVRMSE, the M-FRNN model has the lowest

results implying the proposed predicted model can achieve more stable occupancy profiles results when compared to the ground truth.

Table 2. Results of the MAE, MAPE, and CVRMSE of each approach over 9 days.

		Occupancy_CO2_50ppm			Occupancy_WiFi_Markov			Occupancy_WiFi_M-FRNN		
Index	Max	MAE	MAPE	CVRMSE	MAE	MAPE	CVRMSE	MAE	MAPE	CVRMSE
day1	17	3.99	44%	51.17%	2.3	25.20%	29.22%	1.82	22.14%	24.70%
day2	14	2.97	43.1%	44.2%	2.61	34%	41.9%	1.88	48.3%	29.3%
day3	17	2.64	25.3%	28.1%	2.57	22.8%	29.8%	2.4	25.9%	29.6%
day4	17	2.8	45.7%	42.8%	2.61	32.8%	36.9%	2.23	28.4%	34.6%
day5	11	2.23	78.4%	61.7%	1.72	45.4%	53.6%	1.72	48.2%	57.5%
day6	19	3.22	27.3%	30.8%	2.87	33.4%	27.9	2.66	33.5%	25.2%
day7	14	2.13	34.1%	35.3%	2.23	30%	31.8%	1.94	26.3%	29%
day8	19	3.63	50.5%	42.3%	3.48	40%	43.3%	3.78	46.5%	42.2%
day9	14	3.55	44.2%	46.5%	3.37	38.2%	47.9%	3.47	44.8%	33%

6. DISCUSSIONS

The conventional CO_2-based occupancy prediction methods have been thoroughly investigated over years regarding the occupant counting and demand controlled ventilation systems. As CO2 concentration is a direct indicator of indoor air quality, it effectively improves the wide application of CO2 concentration methods in current HVAC systems. However, high data-resolution CO2 sensors are far more expensive than WiFi probe sensors. The readings of CO2 sensors are usually undulatory, and the CO2 concentration of indoor air will change depending on the status of doors and windows. The ambiguous relationship between CO2 concentration and occupant number potentially results in energy wastes. WiFi probe approach is a good supplement to the existing CO2 method, given the ease and affordability of system deployment. WiFi probe can provide high-resolution occupancy data to assist the CO2-based indoor air quality control. In addition, WiFi networks can greatly improve the prediction accuracy in large multi-zone spaces, where the slow CO2 dilution can potential cause significant error.

In term of prediction accuracy and efficiency, this study compared several popular prediction approaches. The popular occupancy prediction indicator, x-accuracy, shows that the proposed M-FRNN method can significantly improve the prediction accuracy with a tolerance of 2, 3, and 4 occupants to reach 80.9%, 89.6%, and 93.9%, while the CO2-based methods have accuracy of 74.8%, 79.1%, and 93.9%. WiFi networks are installed in most modern buildings, with advances in technologies, WiFi can interplay

a more significant role in building control. For example, the Internet-of-Things technologies allows small appliances and facilities connected via WiFi network and can be predicted and operated. More complicated prediction method can be used to not only prdict the occupants' presence but also actives. In addition, the predicted occupancy also can be extended to intelligent building management. For example, the building energy management can estimation cooling, heating, and ventilation load based on occupancy information (such as occupant number and energy consumption schedule and profiles) to automatically adjust the building service systems, which is one key feature of the grid-interactive efficient buildings [ref].

The WiFi probe-based occupancy prediction method in this study also yields some limitations. The first limitation is privacy concern, since the device MAC address is often associated with occupant's identification. In IEA EBC Annex 66, it recommends that ethics conduct should be concerned by assuring scientific validity and minimum potential harm to participants during occupancy study. Secondly, this study didn't validate the occupancy model in different experiment fields and different time resolutions. For example, the proposed occupancy model can be an acceptable solution in smaller spaces or at finer temporal (e.g., 1-minute) resolution. As building energy control system might vary with types of spaces and different measurement timestep of occupancy, occupancy validation in multi-type buildings and multi resolutions should be prioritized in future. Thirdly, the proposed method is only applicable in WiFi covered areas. Its application may be constrained in very small room with few occupants and the areas with insufficient or no WiFi services. The results also show a mobile phone might turn to sleep mode when the phone is not in use for a long time [56]. Once the mobile phone turns to sleep mode, communication with the WiFi signal is low-frequency and may mislead the prediction algorithm. In the simulation, occupancy data are generated from experiment and model results while other variables were kept identically. Building performance simulation results under different occupancy results may be optimistic.

7. CONCLUSIONS

This study employed the WiFi probe technology to capture the connection requests and responses to dynamically diagnose and assess a building's occupancy information. For comparison, CO2 concentrations based occupancy sensing method was applied while camera based occupancy count was taken as ground truth. Considering the unique characteristics of occupancy data, the Markov-based feedback recurrent neural network

M-FRNN approach was developed. The on-site experiment during nine days demonstrated that M-FRNN based occupancy model using WiFi probes shows the best accuracy with a tolerance of 2, 3, and 4 occupants can reach 80.9%, 89.6%, and 93.9%, respectively. As WiFi signal is popularly used in buildings, the occupancy sensing with WiFi signal enables building control systems to adjust services based on people count, which will lead to energy savings and improved occupant comfort. This study provide insights for WiFi based occupancy studies and occupant-related studies to improve building energy efficiency.

ACKNOWLEDGEMENTS

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] Siano P. Demand response and smart grids—A survey. Renew Sustain Energy Rev 2014;30:461–78. doi:10.1016/j.rser.2013.10.022.

[2] Strbac G. Demand side management: Benefits and challenges. Energy Policy 2008;36:4419–26. doi:10.1016/j.enpol.2008.09.030.

[3] Li N, Yang Z, Becerik-Gerber B, Tang C, Chen N. Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures? Appl Energy 2015;159:196–205. doi:10.1016/j.apenergy.2015.09.001.

[4] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization methods applied to building performance analysis. Appl Energy 2014;113:1043–58. doi:10.1016/j.apenergy.2013.08.061.

[5] Tronchin L, Fabbri K. Energy performance building evaluation in Mediterranean countries: Comparison between software simulations and operating rating simulation. Energy Build 2008;40:1176–87. doi:10.1016/j.enbuild.2007.10.012.

[6] Galvin R. Making the “rebound effect” more useful for performance evaluation

of thermal retrofits of existing homes: Defining the “energy savings deficit” and the “energy performance gap.” Energy Build 2014;69:515–24. doi:10.1016/j.enbuild.2013.11.004.

[7] Pieter de Wilde. The gap between predicted and measured energy performance of buildings: A framework for investigation. Autom Constr 2014;41:40–9. doi:10.1016/J.AUTCON.2014.02.009.

[8] Menezes AC, Cripps A, Bouchlaghem D, Buswell R. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. Appl Energy 2012;97:355–64. doi:10.1016/j.apenergy.2011.11.075.

[9] Hong T, Yan D, D ’oca S, Chen C-F. Ten questions concerning occupant behavior in buildings: The big picture. Build Environ 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.

[10] Oldewurtel F, Sturzenegger D, Morari M. Importance of occupancy information for building climate control. Appl Energy 2013;101:521–32. doi:10.1016/j.apenergy.2012.06.014.

[11] Díaz JA, Jiménez MJ. 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 2017;199:121–41. doi:10.1016/j.apenergy.2017.04.082.

[12] Masoso OT, Grobler LJ. The dark side of occupants’ behaviour on building energy use. Energy Build 2010;42:173–7. doi:10.1016/J.ENBUILD.2009.08.009.

[13] Dong B, Andrews B. sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. Build. Simul. Elev. Int. IBPSA Conf., 2009.

[14] Wood G, Newborough M. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. Energy Build 2003;35:821–41. doi:10.1016/S0378-7788(02)00241-4.

[15] Yan D, Hong T, Dong B, Mahdavi A, D’Oca S, Gaetani I, et al. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy Build 2017;156:258–70. doi:10.1016/J.ENBUILD.2017.09.084.

[16] Wang S, Jin X. CO2-Based Occupancy Detection for On-Line Outdoor Air Flow Control. Indoor Built Environ 1998;7:165–81. doi:10.1159/000024577.

[17] Jiang C, Masood MK, Soh YC, Li H. Indoor occupancy estimation from carbon dioxide concentration. Energy Build 2016;131:132–41. doi:10.1016/j.enbuild.2016.09.002.

[18] Wang S, Burnett J, Chong H. Experimental Validation of CO2-Based Occupancy Detection for Demand-Controlled Ventilation. Indoor Built Environ 1999;8:377–91. doi:10.1177/1420326X9900800605.

[19] Wang W, Chen J, Song X. Modeling and predicting occupancy profile in office space with a WiFi probe-based Dynamic Markov Time-Window Inference approach. Build Environ 2017;124:130–42. doi:10.1016/J.BUILDENV.2017.08.003.

[20] Wang Y, Shao L. Understanding occupancy pattern and improving building energy efficiency through WiFi based indoor positioning. Build Environ 2017;114:106–17. doi:10.1016/j.buildenv.2016.12.015.

[21] Liang X, Hong T, Shen GQ. Improving the accuracy of energy baseline models for commercial buildings with occupancy data. Appl Energy 2016;179:247–60. doi:10.1016/j.apenergy.2016.06.141.

[22] Wang Q, Augenbroe G, Kim J-H, Gu L. Meta-modeling of occupancy variables and analysis of their impact on energy outcomes of office buildings. Appl Energy 2016;174:166–80. doi:10.1016/j.apenergy.2016.04.062.

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

[24] Goyal S, Ingley HA, Barooah P. Occupancy-based zone-climate control for energy-efficient buildings: Complexity vs. performance. Appl Energy 2013;106:209–21. doi:10.1016/j.apenergy.2013.01.039.

[25] Korkas CD, Baldi S, Michailidis I, Kosmatopoulos EB. Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage. Appl Energy 2016;163:93–104. doi:10.1016/j.apenergy.2015.10.140.

[26] Hong T, Taylor-Lange SC, D'Oca S, Yan D, Corgnati SP. Advances in research and applications of energy-related occupant behavior in buildings. Energy Build 2016;116:694–702. doi:10.1016/j.enbuild.2015.11.052.

[27] Yan D, O'Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: Current state and future challenges. Energy Build 2015;107:264–78. doi:10.1016/j.enbuild.2015.08.032.

[28] Wang W, Chen J, Huang G, Lu Y. 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.

[29] Kim Y-S, Heidarinejad M, Dahlhausen M, Srebric J. Building energy model calibration with schedules derived from electricity use data. Appl Energy 2017;190:997–1007. doi:10.1016/j.apenergy.2016.12.167.

[30] Chen Y, Liang X, Hong T, Luo X. Simulation and visualization of energy-related occupant behavior in office buildings. Build Simul 2017;10:785–98. doi:10.1007/s12273-017-0355-2.

[31] Chen J, Jain RK, Taylor JE. Block Configuration Modeling: A novel simulation model to emulate building occupant peer networks and their impact on building energy consumption. Appl Energy 2013;105:358–68. doi:10.1016/j.apenergy.2012.12.036.

[32] Lu Y, Zhang N, Chen J. A behavior-based decision-making model for energy performance contracting in building retrofit. Energy Build 2017;156:315–26. doi:https://doi.org/10.1016/j.enbuild.2017.09.088.

[33] Pisello AL, Asdrubali F. Human-based energy retrofits in residential buildings: A cost-effective alternative to traditional physical strategies. Appl Energy 2014;133:224–35. doi:10.1016/j.apenergy.2014.07.049.

[34] Pisello AL, Castaldo VL, Piselli C, Fabiani C, Cotana F. How peers' personal attitudes affect indoor microclimate and energy need in an institutional building: Results from a continuous monitoring campaign in summer and winter conditions. Energy Build 2016;126:485–97. doi:10.1016/J.ENBUILD.2016.05.053.

[35] Pisello AL, Rosso F, Castaldo V., Piselli C, Fabiani C, Cotana F. The role of building occupants' education in their resilience to climate-change related events. Energy Build 2017;154:217–31. doi:10.1016/J.ENBUILD.2017.08.024.

[36] ASHRAE. ASHRAE Standard 90.1-2007: Energy Standard for Buildings Except Low-Rise Residential Buildings. n.d.

[37] Gunay HB, O'Brien W, Beausoleil-Morrison I. A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices. Build Environ 2013;70:31–47. doi:10.1016/j.buildenv.2013.07.020.

[38] Wagner A, O'Brien W, Dong B. Exploring occupant behavior in buildings : methods and challenges. Springer International Publishing; 2017. doi:10.1007/978-3-319-61464-9.

[39] Benezeth Y, Laurent H, Emile B, Rosenberger C. Towards a sensor for detecting human presence and characterizing activity. Energy Build 2011;43:305–14. doi:10.1016/j.enbuild.2010.09.014.

[40] Page J, Robinson D, Morel N, Scartezzini J-L. A generalised stochastic model for the simulation of occupant presence. Energy Build 2008;40:83–98.

doi:10.1016/j.enbuild.2007.01.018.

[41] Davis JA, Nutter DW. Occupancy diversity factors for common university building types. Energy Build 2010;42:1543–51. doi:10.1016/j.enbuild.2010.03.025.

[42] Li Z, Dong B. Short term predictions of occupancy in commercial buildings—Performance analysis for stochastic models and machine learning approaches. Energy Build 2018;158:268–81. doi:10.1016/J.ENBUILD.2017.09.052.

[43] Zou J, Zhao Q, Yang W, Wang F. Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation. Energy Build 2017;152:385–98. doi:10.1016/J.ENBUILD.2017.07.064.

[44] Ahn K-U, Kim D-W, Park C-S, de Wilde P. Predictability of occupant presence and performance gap in building energy simulation. Appl Energy 2017;208:1639–52. doi:10.1016/J.APENERGY.2017.04.083.

[45] Shan K, Sun Y, Wang S, Yan C. Development and In-situ validation of a multi-zone demand-controlled ventilation strategy using a limited number of sensors. Build Environ 2012;57:28–37. doi:10.1016/j.buildenv.2012.03.015.

[46] Wang S, Burnett J, Chong H. Experimental Validation of CO2-Based Occupancy Detection for Demand-Controlled Ventilation. http://dx.doi.org/101177/1420326X9900800605 2016. doi:10.1177/1420326X9900800605.

[47] Ekwevugbe T, Brown N, Pakka V, Fan D. Real-time building occupancy sensing using neural-network based sensor network. 2013 7th IEEE Int. Conf. Digit. Ecosyst. Technol., IEEE; 2013, p. 114–9. doi:10.1109/DEST.2013.6611339.

[48] Ekwevigbe T, Brown N, Pakka VH, Fan D. Real-time Building Occupancy Sensing for Supporting Demand Driven HVAC Operations. Int. Conf. Enhanc. Build. Oper. , 2013.

[49] Yang Z, Becerik-Gerber B. Modeling personalized occupancy profiles for representing long term patterns by using ambient context. Build Environ 2014;78:23–35. doi:10.1016/j.buildenv.2014.04.003.

[50] Yang J, Santamouris M, Lee SE. Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. Energy Build 2016;121:344–9. doi:10.1016/j.enbuild.2015.12.019.

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

[52] Bisio I, Lavagetto F, Marchese M, Sciarrone A. Smart probabilistic fingerprinting for WiFi-based indoor positioning with mobile devices. Pervasive Mob Comput 2016;31:107–23. doi:10.1016/j.pmcj.2016.02.001.

[53] Wang Y, Shao L. Understanding occupancy pattern and improving building energy efficiency through WiFi based indoor positioning. Build Environ 2017;114:106–17. doi:10.1016/j.buildenv.2016.12.015.

[54] Campos RS, Lovisolo L, de Campos MLR. WiFi multi-floor indoor positioning considering architectural aspects and controlled computational complexity. Expert Syst Appl 2014;41:6211–23. doi:10.1016/j.eswa.2014.04.011.

[55] Chen J, Ahn C. Assessing occupants' energy load variation through existing wireless network infrastructure in commercial and educational buildings. Energy Build 2014;82:540–9. doi:10.1016/j.enbuild.2014.07.053.

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

[57] Pritoni; M, Nordmand; B, Piette MA. Accessing WiFi Data for Occupancy Sensing. 2017.

[58] Sidiropoulos N, Mioduszewski M, Oljasz P, Schaap EdwinSchaap E. Open Wifi SSID Broadcast vulnerability SSN Project Assessment 2012 2012.

[59] Zhao Y, Zeiler W, Boxem G, Labeodan T. Virtual occupancy sensors for real-time occupancy information in buildings. Build Environ 2015;93:9–20. doi:10.1016/j.buildenv.2015.06.019.

[60] Dodier RH, Henze GP, Tiller DK, Guo X. Building occupancy detection through sensor belief networks. Energy Build 2006;38:1033–43. doi:10.1016/j.enbuild.2005.12.001.

[61] Chen Z, Xu J, Soh YC. Modeling regular occupancy in commercial buildings using stochastic models. Energy Build 2015;103:216–23. doi:10.1016/j.enbuild.2015.06.009.

[62] ASHRAE Standard 62.1. 2007:Ventilation for Acceptable Indoor Air Quality. 2007.

LIST OF FIGURES

LISTS OF TABLES

Sensor	Camera	WiFi Probe	Environment Sensors			
			CO2 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, CO2, Air flow rate, Air pressure, CO			
Data Storage	Online	Online	Local			
Frequency		30s	1min	1min	1min	
Range			0 – 5k ppm	14 - 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%	

		Occupancy_CO2_50ppm			Occupancy_WiFi_Markov			Occupancy_WiFi_M-FRNN		
Index	Max	MAE	MAPE	CVRMSE	MAE	MAPE	CVRMSE	MAE	MAPE	CVRMSE
day1	17	3.99	44%	51.17%	2.3	25.20%	29.22%	1.82	22.14%	24.70%
day2	14	2.97	43.1%	44.2%	2.61	34%	41.9%	1.88	48.3%	29.3
day3	17	2.64	25.3%	28.1%	2.57	22.8%	29.8%	2.4	25.9%	29.6
day4	17	2.8	45.7%	42.8%	2.61	32.8%	36.9%	2.23	28.4%	34.6%
day5	11	2.23	78.4%	61.7%	1.72	45.4%	53.6%	1.72	48.2%	57.5%
day6	19	3.22	27.3%	30.8%	2.87	33.4%	27.9	2.66	33.5%	25.2%
day7	14	2.13	34.1%	35.3%	2.23	30%	31.8%	1.94	26.3%	29%
day8	19	3.63	50.5%	42.3%	3.48	40%	43.3%	3.78	46.5%	42.2%
day9	14	3.55	44.2%	46.5%	3.37	38.2%	47.9%	3.47	44.8%	33%