Building and Environment 124 (2017) 130–142

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach

Wei Wang [a], Jiayu Chen [a, *], Xinyi Song [b]

[a] Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621, AC1, Tat Chee Ave, Kowloon, Hong Kong
[b] School of Building Construction, Georgia Institute of Technology, Atlanta, GA 30332, United States

A R T I C L E I N F O

Article history:
Received 7 June 2017
Received in revised form
15 July 2017
Accepted 1 August 2017
Available online 3 August 2017

Keywords:
Occupancy prediction
Wi-Fi probe
Time window approach
Markov inference

A B S T R A C T

Demand-based HVAC control methods in buildings show great energy saving potential when accurate occupancy information is available. Appropriate service based on actual occupant demand could prevent unnecessary energy waste caused by system overcooling or overheating. Therefore, various occupancy detection approaches had attracted increasing attentions in recent years. Among them, Wi-Fi based detection approaches have been thoroughly discussed since Wi-Fi access points (APs) and wireless devices are ubiquitously used in modern buildings. Compared with traditional request and response based occupancy assessment, the newly developed Wi-Fi probe technology can actively scan Wi-Fi enabled devices even if they are not connected to the network. However, Wi-Fi probe detection still subjects to significant errors due to unstable signal and unpredictable occupant behavior. This study stresses the time-series and stochastic characteristics of detected signals and proposes a novel Dynamic Markov Time-Window Inference (DMTWI) model to predict reliable occupancy. The conventional Auto-Regressive Moving Average (ARMA) model and Support Vector Regression (SVR) model are also examined and compared with the proposed approach. Also, an on-site experiment was conducted to validate the proposed model, and the results reveal that the prediction accuracy is over 80% when x-accuracy tolerance is less than 4 for weekdays, 3 for holidays, and 2 for weekend days.

1. Introduction

A building consumes energy to maintain its indoor environment and provide services such as lighting, heating, cooling, and ventilation [1]. Among all power consuming sectors in commercial buildings, HVAC systems dominate more than 40% of energy consumption [2]. Improving HVAC control efficiency could dramatically reduce a building's energy consumption and waste [3,4]. In the past, HVAC control models focused on adjusting preset indoor parameters, such as temperature, humidity, and CO_2 concentration [5–9], without considering occupancy information. Even when a building is unoccupied, it is common that the system keeps working to maintain the room's indoor environment requirements as if it were occupied [10]. Therefore, energy wastes occur when the supply of HVAC systems is more than the demand [6,11]. In recent years, the role of occupancy has risen in profiling the actual demand for building service systems [12–14]. Research indicates a building could save about 30–42% of its energy with accurate occupancy detection [15–17]. Many studies have worked to incorporate occupancy in HVAC systems and explored HVAC research based on occupancy patterns [4,11,18–20]. However, occupancy detection is a challenging work that requires high data resolution and automatic data collection. This research extends the exiting Wi-Fi based technology and utilizes a Wi-Fi probe-enabled sensing approach to detect the presents of occupants. Media Access Control (MAC) is regarded as the identity of the occupants and a dynamic Markov inference method is proposed to model and predict occupancy for HVAC system operation in commercial buildings. This method utilizes a time-window enabled filtering algorithm to acquire detailed occupancy profiles of long-term end-users. The proposed Dynamic Markov Time-Window Inference (DMTWI) model aims to predict occupancy based on stochastic properties of the raw Wi-Fi presence data. Also, an experiment in two offices had been conducted to assess the prediction accuracy in both working days and weekend days.

* Corresponding author.
E-mail address: jiaychen@cityu.edu.hk (J. Chen).

http://dx.doi.org/10.1016/j.buildenv.2017.08.003
0360-1323/

2. Background

2.1. Occupancy data resolution and detection

The data resolution can significantly impact the detection accuracy and the compatibility with HVAC systems. Christensen et al. [21] defined occupancy resolution as a three-dimensional data system (i.e. temporal resolution, spatial resolution, and occupant resolution). Teixeira and Dublon [22] categorised occupancy resolution into several levels, such as the presence, count, location, track, identity, and behavior. The most compatible data resolution for HVAC operation is count level. Count level is defined as the duration time of people and the number of people in one zone or room in certain time intervals. It could provide the visibility of occupancy profiles for occupant loads and occupant-related loads in evaluating demand-based control strategies [6,7,23]. Demand-based control methods utilized the occupancy information to allow adjustable HVAC system settings instead of fixed preset system parameters. However, occupancy is a dynamic profile because occupants do not regularly arrive or leave one place at the exact same time each day [24] and many occupants were temporary visitors [25]. A noticeable occupancy variation of up to 60% in a 1-h period could be possible [26]. It is possible that the people entering and exiting the building reach the maximum occupancy within 1 h [27]. All these facts have revealed the non-negligible uncertainties of occupancy. Therefore, many approaches have been developed to mitigate uncertainties and predict accuracy occupancy information.

When actual occupancy is unknown or difficult to capture, ASHRAE recommends approximate occupancy diversity factors based on ASHRAE standard 90.1–2007 [28] to standardize building occupancy for different building types and locations. However, researchers found a 40% variation between the actual occupancy and the standardized ASHRAE occupancy schedules [29]. CO_2 sensors have been employed to estimate the occupant count. In Wang's research [30], dynamic CO_2-based occupancy models have estimated and predicted occupancy profiles in a typical office. However, the model is limited by environmental sensitivity and a slow response rate [12]. Additionally, some works tested CO_2 sensors with temperature, humidity, lighting, and sound sensors and reported the accuracy of occupancy detection between 75 and 84.5% [19,31]. Jiang et al. [32] proposed an indoor occupancy estimation model based on CO_2 concentration with a feature scaled extreme learning machine (FS-ELM) algorithm which has a 50% of probability that the predicted occupancy is exactly same as actual occupancy. Radio-frequency identification (RFID) tags are another occupancy detection system, which have been extensively used in recent years for demand-driven HVAC control systems. Li et al. [11] reported the average accuracy of RFID systems was 88% for stationary occupants and 62% for mobile occupants. However, additional received tags and a complete implementation of supporting facilities are required. Six machine-learning methods were evaluated for occupancy estimation in single-occupancy and multi-occupancy offices, including Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN) [33]. Various ambient sensor types were installed in a box and placed in the area of interest while accuracy reports above 90%.

Using statistical sensing data in occupancy models to predict occupancy probability is an option for a more reliable detection. Yang and Becerik-Gerber [1] tested stochastic processes with the regression model, timer-series modeling, and pattern recognition modeling by using ambient sensing data to model personalized occupancy profiles representing long-term patterns. Dong et al. [34] studied the artificial neural network (ANN) with the Hidden Markov approach through environmental sensing networks to detect and model occupancy information. The Markov chain model is one of most applied models in predicting building occupancy, and some studies based on the Markov chain have achieved occupancy profiles with an average of 73% accuracy [24,35–37]. In their work, Wang et al. [36] proposed a first-order homogeneous Markov chain method to model occupant movement in buildings, which generates occupancy by the results of their movement. The same first-order Markov chain technology is adopted in Ref. [38] to build occupancy models for domestic buildings. Page et al. [24] applied the Markov models to simulate occupancy in building energy simulations, and Virote and Neves-Silva [39] used the hidden Markov Chain model to integrate observable motivations of occupant behavior.

2.2. Wi-Fi based occupancy studies and Wi-Fi probe

Wi-Fi, a common and widely implemented network, offers a more efficient, affordable, and convenient option for occupancy detection, and it is especially suitable for commercial and residential buildings. Also, there are many Wi-Fi enabled devices (laptops, smartphones, tablets.) that allow occupants connect to Wi-Fi networks. With Wi-Fi networks, the occupants can send requests and responses via the internet and leave a connection footprint. Many studies adopted Wi-Fi technology to detect occupancy patterns and imporve HVAC system operation efficiency [40–43], demonstrating Wi-Fi connections and disconnections can be utilized as indicators of building occupancy [44]. Wi-Fi connections can determine presence occupancy profiles with an 83% accuracy, and Balaji utilized existing Wi-Fi infrastructure in commercial buildings along with smartphones with Wi-Fi connectivity to actuate HVAC zones and measured 17.8% in HVAC electrical energy savings [45].

The latest Wi-Fi probe technology allows the APs to scan a specific zone and discover devices without connection. Fig. 1 shows the connection scanning types and the scanning process of Wi-Fi probe. Devices have two common types of connection scanning: active scanning (including direct scanning and broadcast scanning) and passive scanning [46]. In direct scanning, the user sends a probe request to a Service Set Identifier (SSID) access point (AP), and only this requested SSID AP can reply with a probe response. In broadcast scanning, the user broadcasts a probe request with a null SSID received by all APs within range. Those APs give a probe response, and the user chooses to connect to the preferred AP. In passive scanning, available APs act as a beacon to broadcast signals, and the user decides if one connection is requested, instead of actively sending a probe request. The request and response are captured by the Wi-Fi probe with a timestamp and the MAC address of each user, and such information is remotely uploaded to the server and could be downloaded by users. This paper proposed an automatic occupancy detection algorithm based on Wi-Fi probe and validated the proposed methodology in an on-site experiment.

2.3. Existing occupancy prediction models

Various quantitative models have been developed to predict occupancy profiles based on sensory data sources. The Support Vector Regression (SVR) model and the Auto-Regressive Moving Average (ARMA) model were most widely used. For example, Yang and Becerik-Gerber [1] modeled occupancy profiles by the SVR and ARMA time series model using ambient contexts, such as CO_2 concentration, temperature, and humidity. With accelerometers and phone microphones, Rana e.t. [47] proposed occupancy activity classification and occupancy estimation methods by applying the SVR model. This study compared both methods with the proposed model in the following sections.

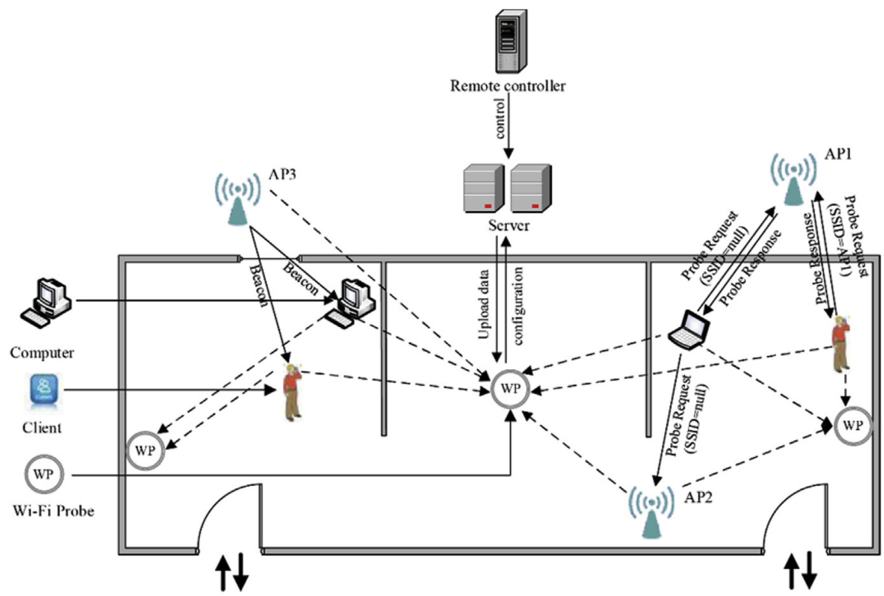

Fig. 1. Connection scanning types of Wi-Fi probe.

2.3.1. Support Vector Regression model

The SVR method has been utilized extensively for occupancy regression and prediction [1,47]. A version of SVM for the regression method is called Support Vector Regression (SVR). Analogously, the model produced by SVR 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. If a training set of sample data is considered as $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, where n is the scope of data, training the original SVR means solving

$$\text{minimize} \quad \frac{1}{2}\|\ \omega\ \|^2 \tag{1}$$

subject to

$$\begin{cases} y_i - \langle \omega, x_i \rangle - b \le \varepsilon \\ \langle \omega, x_i \rangle + b - y_i \le \varepsilon \end{cases} \tag{2}$$

Where

ω and b are the optimization variables to search;
$\langle \cdot, \ \cdot \rangle$ denotes the dot product;
x_i is a training sample with target value y_i;

The inner product plus 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 have to be within an ε range of the true predictions.

2.3.2. Auto-Regressive Moving Average model

Auto-regression techniques have been developed in the study of linear time series, such as ARMA for stationary smoothing time series data models and the Auto-Regressive Integrated Moving Average (ARIMA) for nonstationary time series data [1,48,49]. The selection of the appropriate model depends on the user's preference and stationarity test. Both time series models apply the auto-regression sub-model and moving average sub-model to predict the data in time series. The ARMA model was applied to understand the time series characteristics of the occupancy profile. The model is usually referred to as the ARMA (p,q) model where p is the order of the autoregressive ($AR_{(p)}$) part and q is the order of the moving average ($MA_{(q)}$) part.

$$\begin{aligned} AR_{(p)}:\ x_t &= \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t \\ &= \phi_0 + \varepsilon_t + \textstyle\sum_{i=1}^{p} (\phi_i X_{t-i}) \end{aligned} \tag{3}$$

where the

x_t represents the occupancy profile at time t influenced by previous occupancy series, which shows the occupancy profile at time t is the p-order regression of $(x_{t-1},x_{t-2},\ldots,x_{t-p})$;
ϕ_{i} is the coefficient parameter of the model;
ε_t is a random-variable disturbance;

The moving average can be modeled as

$$\begin{aligned} MA_q:\ x_t &= \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \\ &= \mu + \varepsilon_t - \textstyle\sum_{i=1}^{p} (\theta_i \varepsilon_{t-i}) \end{aligned} \tag{4}$$

where the

x_t is the occupancy profile at time t influenced by the previous occupancy errors, which is the multivariate linear function of q-order regression of random disturbance $(\varepsilon_{t-1},\ \varepsilon_{t-2},\ldots,\varepsilon_{t-q})$;
μ is the expectation of occupancy data;
θ_{i} is the coefficient parameter of the model;
ε_t is a random-variable disturbance;

By integrating the ambient context, the study can get the ARMA (p, 1, q) if a first-order difference operation is chosen

$$\begin{aligned} x_t = \ & \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \\ & - \ldots - \theta_q \varepsilon_{t-q} \end{aligned} \tag{5}$$

The appropriate values of p and q in the model can be facilitated by plotting the partial autocorrelation functions for an estimate of p and using the autocorrelation functions for an estimate of q. The details on how to determine those two values are ignored. In this study, the first order for the AR model and the second order for the MA model was determined for the occupancy series in the ARMA model.

3. Methodology

3.1. Data preprocessing

The devices within the detection area of the Wi-Fi AP coverage can be detected by the Wi-Fi probe and recorded by the AP request logs. In a typical scanning setting, the Wi-Fi probe scanned the network area every 60 s and recorded the timestamp, MAC address, and device type into the data log. Fig. 2 shows the pseudocode for data preprocessing and how the received signals are converted to the duration of presence.

The time t_0 is the detection time when the device first communicates an authentication request to an AP. The code iterates through the whole detection period and associate MAC addresses and connections. This study differentiate short-term occupants and long-term occupants by comparing the length of signal presence. For MAC address have a connection duration less than 20 min in one day is regarded as short-term occupants. Also, this study ignores the corresponding MAC addresses of those clients that have an extremely long connection period, from 14 h up to the entire schedule. These records indicates the MAC address belongs to a facility, such as a computer, a printer, or other networked devices.

3.2. Duration time functions

To filter out computers and short-term occupants, the occupancy resolution t_R was set as 1 min since the Wi-Fi probe scanning frequency is 1 min per read. 20 min were chosen as the length of time window, since there 20 reads are sufficient to smooth the fluctuation and long enough to differentiate temporary visitors. However, proper time window selection for various conditions still need to be studied in future. As shown in Fig. 3, the number of times one user is detected in one window can be calculated as

$$l_{\Delta t} = l_{(n+1)}\Big|_{t=tn+\Delta t} - l_{(n)}\Big|_{t=tn} \quad (6)$$

In one window,

$$l_{\Delta t} \xrightarrow{count} n_{\Delta t} = 1 \quad (7)$$

$$N_{\Delta t} = \sum n_{\Delta t} \quad (8)$$

where

tn is the start time;
Δt is the length of a time window;
l_n and l_{n+1} represent the number of times the user is detected from $T = 0$ to tn and $tn + \Delta t$ respectively;
$l_{\Delta t}$ is the number of times one user is detected in one time window;
$n_{\Delta t} = 1$ means the number of occupants in the time window is one;
$N_{\Delta t}$ counts the total number of occupants in the time window;

The occupancy schedule in the time window can be profiled with $(\Delta t,\ N_{\Delta t})$. The occupancy schedule for the entire day can be profiled at each time window, and this occupancy would be used in the occupancy profile models.

The duration of an occupant stay in the area within a time window is defined as ΔT

$$\Delta T = (l_{\Delta t} - 1)*t_R \quad (9)$$

while in this case, t_R is the resolution (1 min). For example, if an occupant was detected 10 times in a time window means the occupant stayed in the space for 10 min.

The study define the duration time function $f(t)$ in one day for one user as

$$f(t)|_{client} = \sum_{i=1}^{k=\frac{24*60}{\Delta T}} \Delta T = \sum_{i=1}^{k} (l_{n+1}|_{t=tn+\Delta t} - l_n|_{t=tn} - 1)*t_R \quad (10)$$

```
Pseudocode:
Set N(t) = the number of occupants at time t, T_begin = the beginning time of detection, T_end = the
end time of detection, T_tl = the time length from T_begin to T_end, tR = occupancy resolution
Detect:
For t = T_begin:
    If client (probe request from client):
    N(t) =1, t (client)= Tbegin; MAC (client)
Else for t in T_tl:
    for i in T_tl/tR
    if client (probe request at time T_begin+ i*tR ):
   N(t) = N(client), t (client) = T_begin + i*tR , MAC(client)
   return N(t), t(client), MAC(client)
   end
Else for t = T_end:
   if client (probe request from client):
   N(t) = 1, t(client) = T_end; MAC(client)
Return N(t), t(client), MAC(client)
End
```

Fig. 2. The pseudocode for data preprocessing.

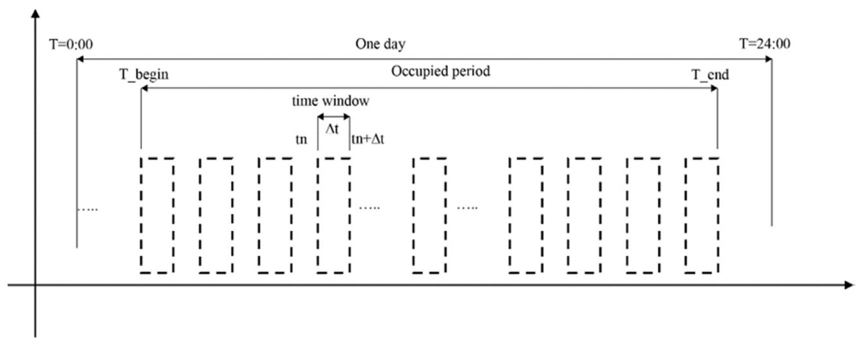

Fig. 3. Illustrated time window approach during occupied period.

The occupancy patterns of computers, short-term occupants, and long-term occupants have different duration times. Some occupants may keep their computers turned on for an entire 24-h period, while others might turn off their computers when they leave and reboot the computers in the morning. However, computers, especially desktops, would not move in or out of the room more frequently than the occupants. These computers would be under the detection of the Wi-Fi probe during occupied hours, but the occupants could be out of the room for various reasons. Therefore, in this case, one day can be divided into two parts: occupied time (from T_begin to T_end) and unoccupied time (from T = 0:00 to T_begin and T_end to T = 24:00). The duration time function for one client can be defined as piecewise functions. When the duration time function $f(t)$ is continuous during unoccupied time, the MAC address is associated as non-occupant-related clients.

3.3. Dynamic Markov Time-Window Inference (DMTWI) model

In existing occupancy stochastic modeling studies [32,37,50,51], researchers suggested that a valid occupancy model should concern unique features of occupancy data format. These features include: (1) Temporal. Occupancy changes over time, and is an ordered sequence of probability at equally spaced time intervals. (2) Stationary. Occupancy evolves regularly and could be learned from previous records. (3) Stochastic. Occupancy is influenced by non-deterministic factors and the status at each time point is determined probabilistically according to the previous status. (4) Dichotomous. Occupancy status often is binary, either occupied/unoccupied or in/out. This study modeled the occupancy as a stochastic process and the current occupancy schedule status analogously depends on the previous time series of occupancy status.

The Markov chain is a method used to model stochastic processes through analyzing the transitions from one state to another [1,37,51]. The transition is quantified by two input parameters: the initial status and transfer probability. In this study, the two occupancy statuses are defined as "in" and "out". If the MAC address of one user is detected by the Wi-Fi probe, the model identifies this user as "in", otherwise, as "out". The Markov chain model also embeds a moving time window, all possible MAC addresses detected in the time window are considered as the "in" state. Fig. 4 shows the illustration of the time-window based Markov chain model.

For the two initial occupancy statuses, the transfer probability indicates the probability of status changes form "in" to "out" during time t to time t+1, and denoted as

$$TPM|_t = \begin{bmatrix} P_t^{i-o} & P_t^{i-i} \\ P_t^{o-o} & P_t^{o-i} \end{bmatrix} \ (t > 1) \tag{11}$$

where

$TPM|_t$ means the transition probability matrix at the current time t of one occupant;
P_t^{i-o} denotes the observed probability that one occupant whose status is "in" at the current time t would be "out" in the next time window;
P_t^{i-i} presents the possibility that the status is still "in";
P_t^{o-i} and P_t^{o-o} denote the probabilities of switching from "out" to "in" and remaining "out";

These probabilities are calculated by the observable conditional probability of historical records. For example, P^{o-i} is calculated as

$$P_t^{o-i} = P(observed\ state = \text{i}|observed\ state = o) \ \ (t > 1) \tag{12}$$

Fig. 5 shows a sample process of the Markov chain model. In the sample model, there are n MAC addresses in the current time window, respectively, and occupancy resolution is t_R. All detected MAC addresses are represented in a Hamming matrix, and the occupancy status of these MAC addresses are assigned as either "in" or "out" for the matrix elements.

Based on Eq. (11), the transfer probability can be calculated as

$$P_t^{i-i} = \frac{\sum N_{i-i}}{\sum N_{i-i} + \sum N_{i-o}} \quad P_t^{o-o} = \frac{\sum N_{o-o}}{\sum N_{o-o} + \sum N_{o-i}} \tag{13}$$

where

N_{i-i} is the frequency that occupancy status transited from "in" to "in";
N_{i-o} is the frequencies that occupancy status transited from "in" to "out";
N_{o-o} and N_{o-i} mean the frequencies that occupancy status transited from "out" to "out" and from "out" to "in" respectively.

However, transition probability could be different in a day. For example, if one occupant is "in" during working time, he or she still occupies with higher probability. If one occupant is "in" during overtime working period, he or she will have a higher probability for "out" state in the next time window. To model the impact of such temporal behaviors, this study adopted the piecewise transfer probability functions for the Markov chain model. In a study by Simona [26], occupancy schedules for a typical office are clustered into the main typical working activities for four mined patterns: (1) going to work (P1) increases in the global occupancy curve; (2) working (P2) stabilizes the global occupancy curve; (3) lunch/breakfast (P3) one valley decreases in the global occupancy curve; (4) overtime (P4) decreases in the global occupancy curve. Table 1 illustrates the divided occupancy pattern in this study based on Simona's research.

Therefore, the transition probability is calculated in a sectional

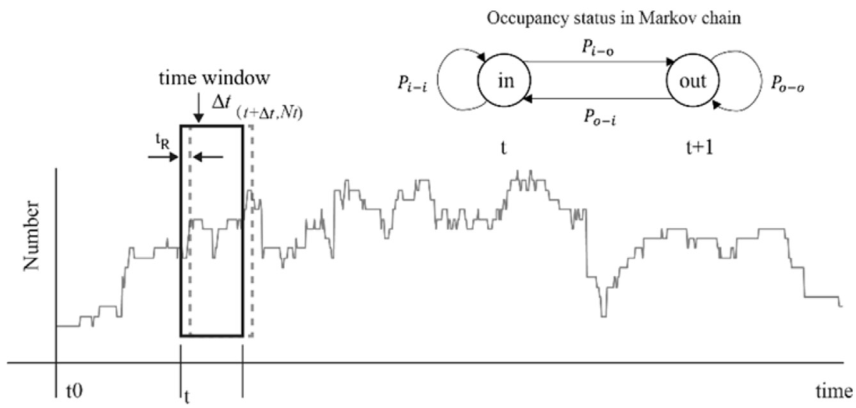

Fig. 4. Illustration of the time-window based Markov chain model.

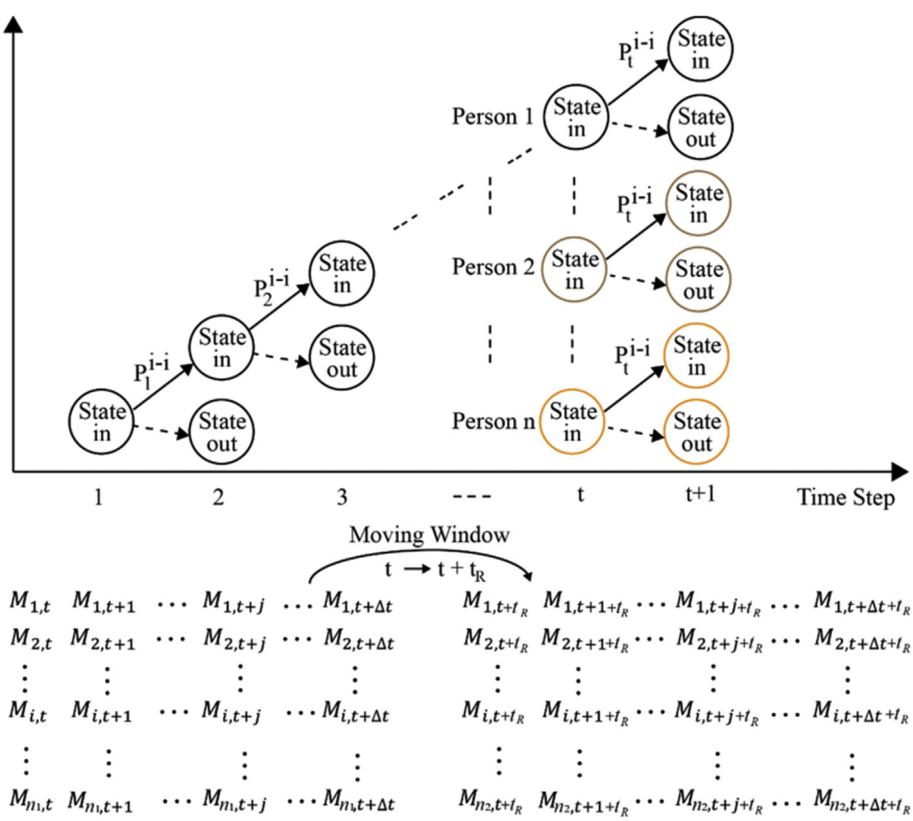

Fig. 5. A sample process of the Markov chain.

pattern, and four transitional matrices should be formulated for the whole day. Finally, in a chosen time window, the total number of MAC address detected in the time window (Δt, $N_{\Delta t}$) and the occupancy status at the time $(t+\Delta t, N_t)$ can be formulated as

$$N_{t+1} = \sum\nolimits_{n=1}^{N(t)} 1 * P_{1-1}^{n} + \sum\nolimits_{n=1}^{N_{\Delta t}-N(t)} 1 * P_{0-1}^{n} \quad t \in [P1, P2, P3, P4] \tag{14}$$

Table 1
The divided occupancy pattern in this study.

Pattern	Going to work	Working	Lunch	Overtime
schedule	8–10:00	10–12:00, 13:30–18:00	12–13:30	18–23:00

The inputs of occupancy model then can be calculated based on the results from Wi-Fi probe devices.

Fig. 6 shows the flow chart of the whole occupancy assessment process. The occupancy profiles are sketched with the data processing procedures, and the MAC addresses of all clients were recorded with the request timestamp. The preliminary experiment data recordings from all Wi-Fi probe devices were conducted in an office room, and the system detected all possible users in target areas containing request signals from occupants and computers. A filtering algorithm is developed to eliminate MAC addresses from computers and short-term occupants to identify actual occupants. Six occupancy profiling results are recorded in Phase I and compared in Phase II. x-accuracy assessment is used to evaluate the performance of different occupancy models.

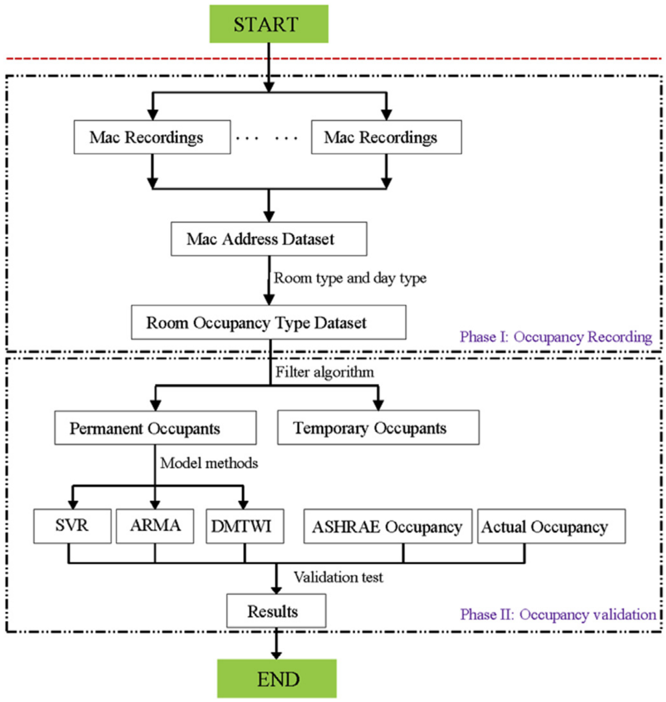

Fig. 6. Flow chart of the occupancy assessment process.

3.4. Validation experiment design

3.4.1. Experiment sites

The validation experiment sites are two Ph.D. research students office rooms on the City University of Hong Kong campus. Both rooms are covered by Wi-Fi networks and the Wi-Fi probe is enabled. The experiment lasted for three days, including a weekday, a holiday, and a weekend day. Seven APs covered the entire rooms and created overlapping and free signal networks in both areas during the experiment. Fig. 7 shows the layouts of both office rooms. Office room A is a large room with a gross area of about 200 square meters and a design maximum of 68 seats. Only about 25 seats were used during the experiment period. Room B is a small room with an area of about 25 square meters and a design maximum of 10 seats, and only 3 seats were used during the experiment period. The research team installed three Wi-Fi probe devices in office room A and two Wi-Fi probe devices in office room B. The data sampling frequency of the Wi-Fi probe was set as 1 min per read to observe Wi-Fi requests and responses from occupants every minute. Occupants in the area were identified by MAC addresses.

3.4.2. The ground truth acquisition

Overhead cameras were installed close to the doors (as shown in Fig. 7) to capture the actual number of occupants in each office room. In addition, two cameras in room A were synchronized with internet time. Because the sampling data of the Wi-Fi probe device was 1 min, the video analysis obtained the number of entrances and exits through each door at the same sampling frequency. The number of occupants were manually counted based on the recorded video for each minute.

3.4.3. X-accuracy

For HVAC control and operation in practice, one or two occupants cause no practical differences. Therefore, to assess the performance of all models, the x-accuracy method was adopt. Proposed by Jiang et al., x-accuracy reports the prediction accuracy when x occupants error is allowed [32]. In the study, the tolerance x represents the prediction error of number of occupants when the prediction was compared to actual occupants. The x-accuracy can be calculated through adjusting the threshold of correct prediction. For example, the 0-accuracy (x = 0) means one prediction is regarded correct prediction only if the prediction is exactly same as actual occupancy.

$$\tau(N_{PO}, x) = \frac{\sum_1^M X(|N_{AO} - N_{PO}|, x)}{M} \tag{15}$$

while

$$X(|N_{AO} - N_{PO}|, x) = \begin{cases} 1, & |N_{AO} - N_{PO}| \le x \\ 0, & |N_{AO} - N_{PO}| > x \end{cases} \tag{16}$$

Where

$\tau(N_{PO}, x)$ is the x-accuracy of a prediction model;
x is the tolerance level;
M is the sampling size;
N_{AO} and N_{PO} are the number of actual occupants and predicted occupants in a zone, respectively;

4. Results and analysis

4.1. Measured occupancy profiles in office room A

The detection period for room A was 30 June to 08 July 2016, including five sequent weekdays, one holiday, and one weekend day. Fig. 8 illustrates the results of one weekday (Monday, 04 July 2016), one holiday (01 July 2016), and one weekend day (Saturday, 02 July 2016). This office was usually occupied from around 8:00 to 23:00. Generally, the occupancy level of weekday is higher than that on the holiday and weekend day. Occupancy reached the maximum level during the morning and afternoon office hours and the minimum level after the office hours. Room A was close to half-occupied around 12:00–13:00 on a weekday, 12:00–14:30 on weekend day. Occupancy profiles during 18:00–19:00 on three days had a similar drop at lunch time, and the room was close to unoccupied during 18:00–19:00 on holiday.

Three different occupancy schedules were compared: (1) the detected occupancy schedule (DOS) using Wi-Fi probe and the proposed model, (2) the actual occupancy schedule (AOS) using cameras, and (3) the ASHRAE recommended occupancy schedule (AROS). The number of occupants for AROS was calculated with the maximum number of the room and occupancy factors that suggested by ASHRAE. The data patterns of the DOS and AOS is highly consistent with each other. The observed results showed that the occupancy profiles for those three days in room A were not steady, which might be caused by random occupants' behaviors, given the occupants are students with a flexible time schedule. For example, during the working hours, occupants might leave the room for various reasons. Although the duration time function can be used to filter out most disturbances caused by computers and other devices, some MAC scanning noise still can not be avoided.

During regular work hours (8:00 to 18:00), AOS and AROS do show some similarity, however, on the weekend, the AOS and AROS are significantly distinctive to each other. The AROS was ignored on the holiday since the ARSHRE did not have a recommended holiday schedule. During overtime work period (18:00 to 23:00), there were discernible differences for both schedule. Many occupants chose to work in office overtime.

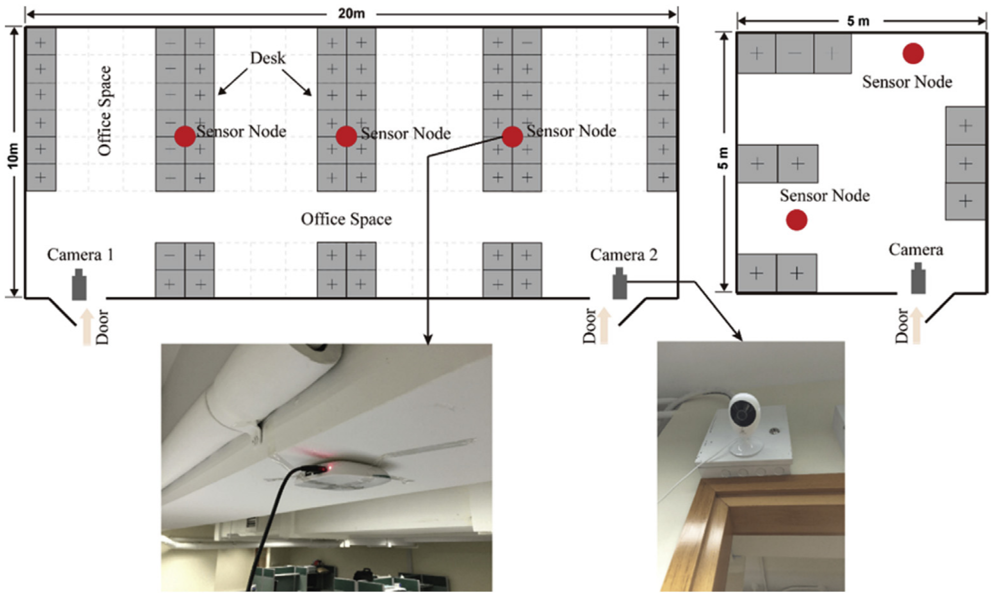

Fig. 7. Layouts of room A (left) and room B(right).

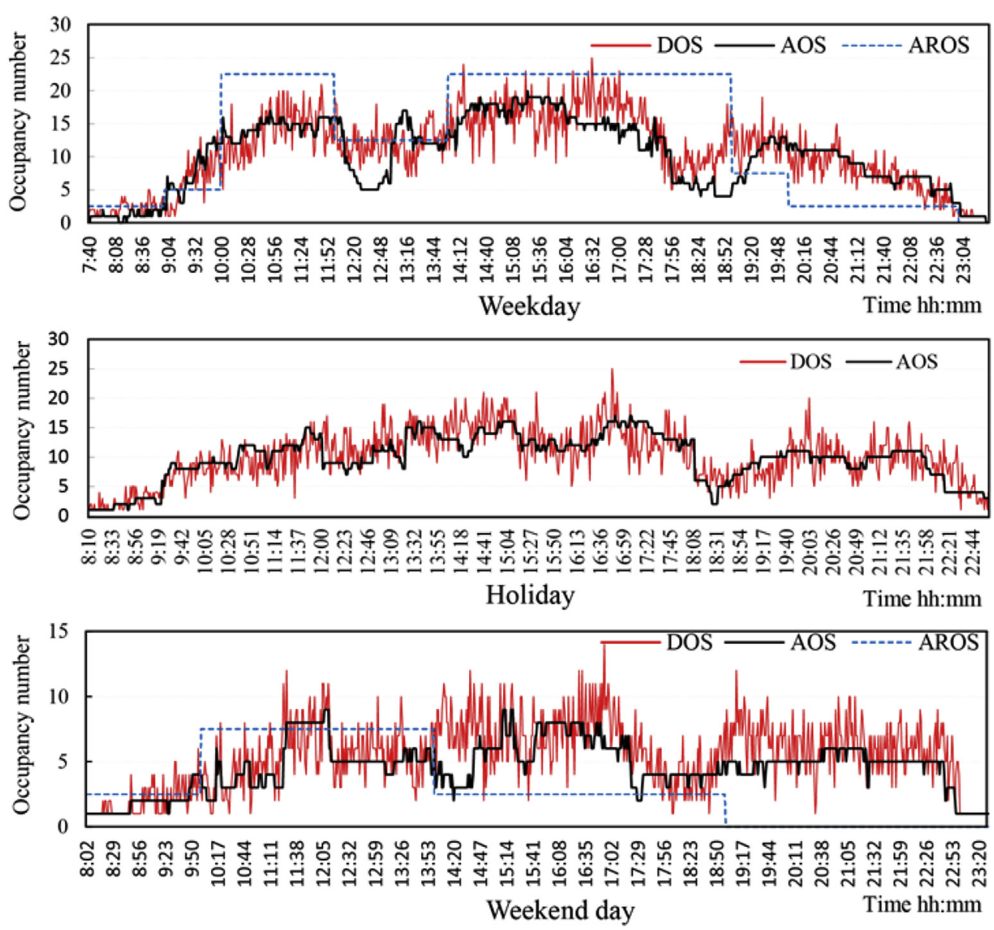

Fig. 8. Occupancy profiles for three types of days in room A.

4.2. Measured occupancy profiles in office room B

Three permanent occupants stayed in office room B during the experiment and the room has one entrance. Actual occupancy was generated by one camera on the door, while the detected occupancy schedule was obtained by two Wi-Fi probe devices. Fig. 9 presents the occupancy schedule of room B on a weekday (11 July 2016), and weekend day (09 July 2016). Similarly, DOS, AOS, and AROS were also compared for room B. The detected occupancy is fluctuated due to the small number of occupants in room B. Comparatively, the occupant profile in the small room was more stable. Also in room B, the transition between unoccupied and occupied was more frequently, which suggests a higher energy saving potential.

4.3. Predicted occupancy for HVAC operation

Although occupancy may change temporarily, habitual occupancy profiles will follow a similar pattern over a long period, and it is possible to develop specific occupancy profiles for different types of days. The occupancy pattern results from the Wi-Fi probe devices were too undulatory to input in occupancy-related HVAC studies. Therefore, the proposed model is able to forecast the occupancy of the next time window based on previous time windows and historical records. SVR and ARMA were also chosen for the comparison as counterpart occupancy prediction models.

The occupancy profile was calculated based on the cleaned tidy raw data detected by Wi-Fi probe. The unoccupied status or predicted unoccupied status were excluded. To avoid to frequency adjustment and maintain the stability of captured data, 5-min data resolution was used. Similarly, the time window was also chosen as 5 min. Four patterns pieces (P1-P4) are constructed in sectionalized calculations for transitional matrices calculation for all valid MAC addresses. The prediction was applied only for large-scale room, such as room A, from 08:00 a.m. to 23:00 p.m. on a weekday, holiday, and weekend day. Figs. 10–12 show the occupancy profile prediction results for a weekday, holiday, and weekend day.

Figs. 10–12 show that all three predicted occupancy based on Wi-Fi probe is less fluctuated compared to DOS. The predicted occupancy have a strong similarity at the most time of a day, expect the period from 13:00 to 15:00. Since the outcomes of all three model vary at each period and difficult to compare, the x-accuracies of the collected prediction results are compared in Figs. 13–15.

The x-accuracy of all three days show that 0-tolerance accuracy (only when the predication and actual results are the same, the prediction regards as correct) of three occupancy models is less than 40%. The 1-tolerance accuracy (when the erorr of a prediction less than one occupant, the prediction regards as correct) is close to 67% and 2-tolerance accuracy (when the erorr of a prediction less than two occupants, the prediction regards as correct) is close to 80%. In a large office space with densely populated residents, small errors do not insignificantly affect the HAVC operation, therefore, the accuracy of the proposed model is acceptable.

When comparing SVR, ARMA, and DMTWI, three models perform similar for x-accuracy that has higher tolerance. For example, for the weekday, the 5-tolerance accuracy for three models are 83.9%(SVR), 85%(ARMA), and 85%(DMTWI). When comparing x-accuracy with low tolerance, DMTWI perform bests for the weekday and weekend day and SVR is more accuracy for holiday.

5. Discussion

In this study, the authors investigated the possibility of applying Wi-Fi probe-enabled occupancy sensing methods to profile occupancy in two office rooms on a weekday, holiday, and weekend day. The Wi-Fi probe devices are able to select the MAC addresses of permanent occupants with regarding to occupancy profiles by a time-window based duration time functions. The results of occupancy prediction of the proposed DMTWI model demonstrated its effectiveness and feasibility in occupancy acquisition. For large commercial offices, when the occupancy is stationary, the proposed model can aggregated the historical occupancy pattern. For smaller space, the model can be adaptive by integrating on-site detection and historical data for occupancy forecasting. Also, given the model is a time-window based approach, it is also flexible in coupling with HVAC adjustment frequency.

Moreover the predicted occupancy data has unique features,

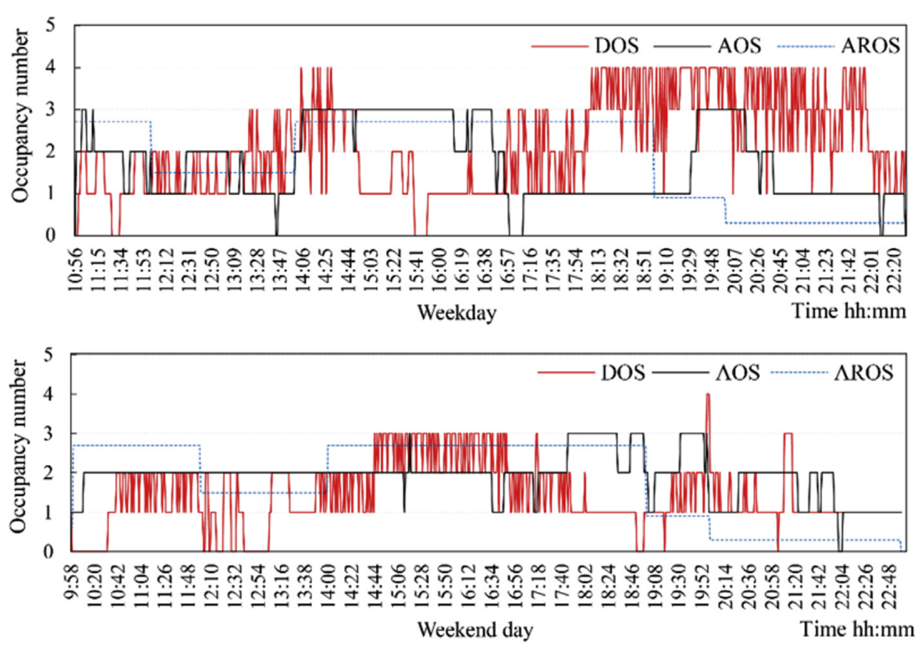

Fig. 9. Occupancy profiles on a weekday and weekend day in room B.

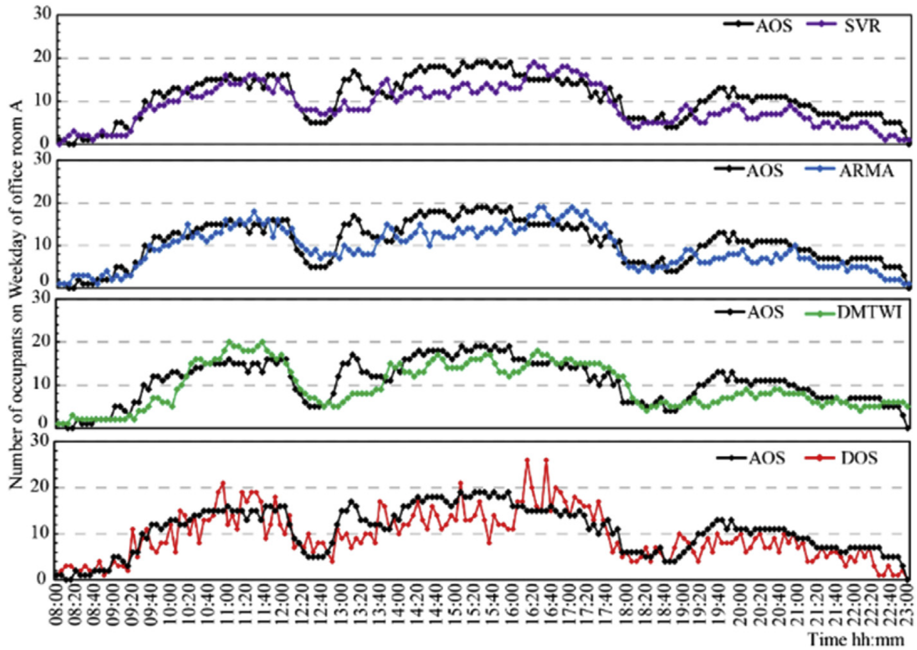

Fig. 10. Occupancy results under different models on a weekday.

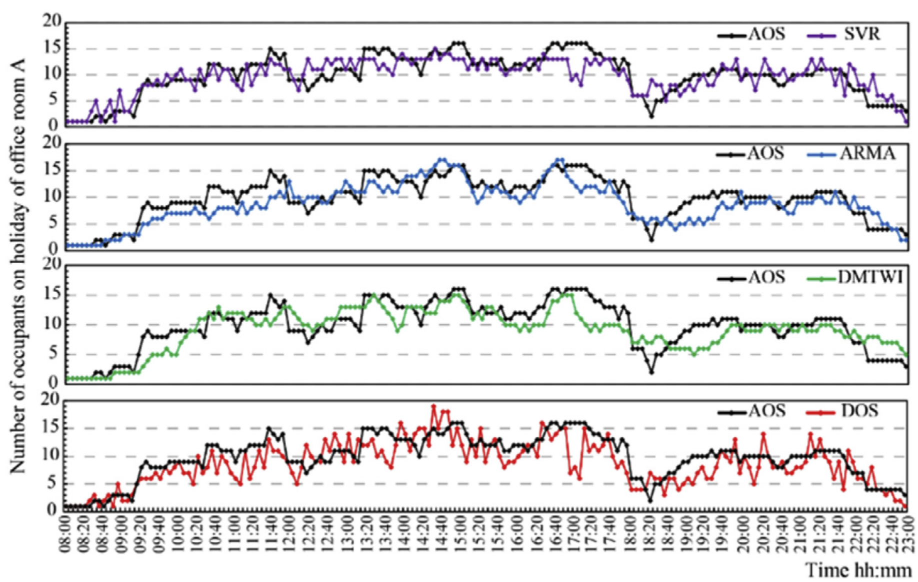

Fig. 11. Occupancy results under different models on a holiday.

including real-time, online, and identifiable tags. The frequency of Wi-Fi probe scanning could customised, and the data is uploaded immediately to online servers. High temperatal data resolution can be used to associate occupancy conditions and building loads. For example, the supply air volume can be controlled according to occupancy level rather than the system suggested setting. Many research suggested reliable occupancy prediction is the premise of accurate energy load and fresh air amounts estimation for building thermal conditions [52] and indoor air quality [9,53]. The cloud based datas base also can be integrated with other on-line sensing technologies, such as temperature sensors and CO_2 concentration sensors. Coupling these systems, not only can bridge the indoor environment and occupancy information but also imporve detection accuracy through data fusion. The identifiable tag, the MAC address of users' devices, can be studied more extensively in the area of occupancy behavior and indoor environment customization.

This study also yields some limitations need to be resolved in future. First, it is assumed that each occupant had at least one Wi-Fi enabled device. Although the vast majority of occupants own one smart phone, the occupants may not turn on their phones' Wi-Fi service or single occupant may bring more than one phones. It is difficult to distinguish the phones from other mobile devices purely based on MAC addresses. One possible solution is developing an event-related approach by detecting the interaction patterns between users and devices to differentiate devices. Also, for the privacy protection purpose, the MAC addresses detection could be a concern. Second, the results showed mobile phones might enter sleep mode if the phones were not in use for a long time [45]. When

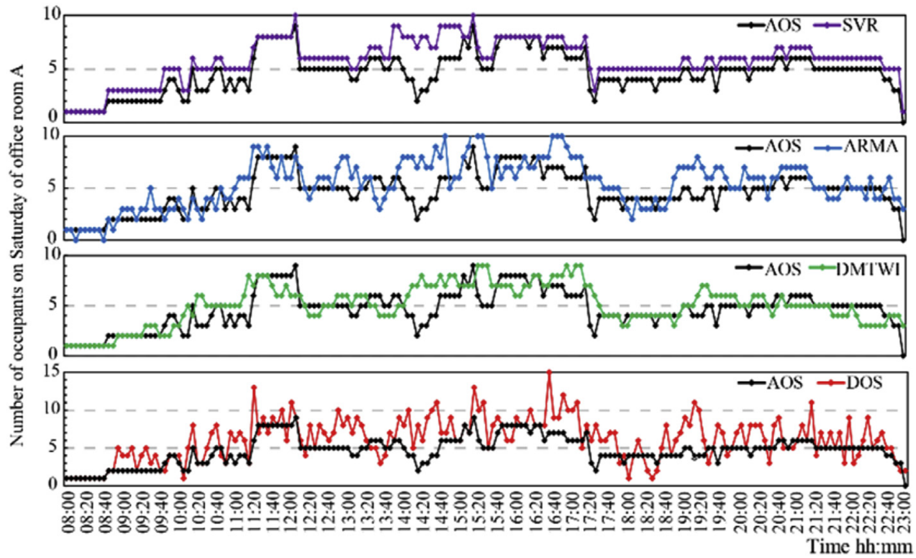

Fig. 12. Occupancy results under different models on a weekend day.

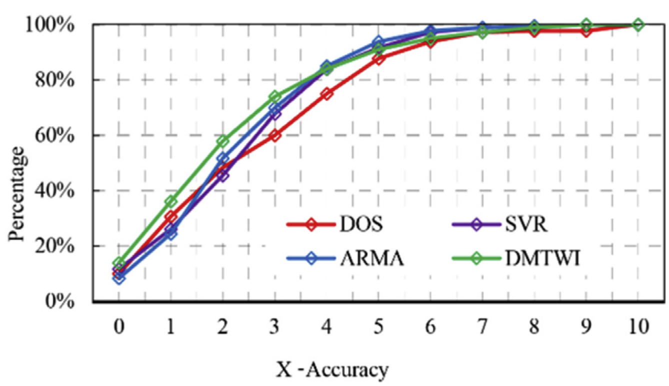

Fig. 13. X-accuracy of different models on a weekday.

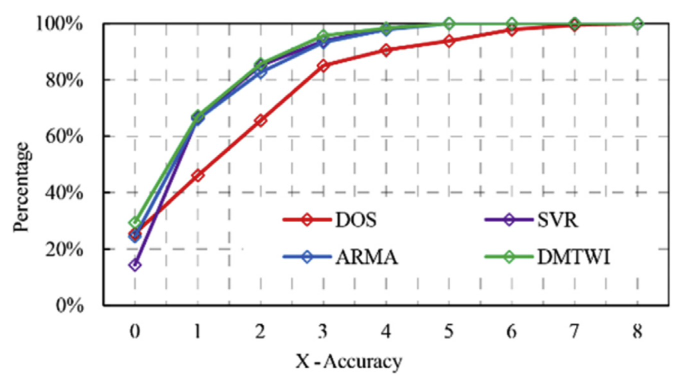

Fig. 15. X-accuracy of different models on a weekend day.

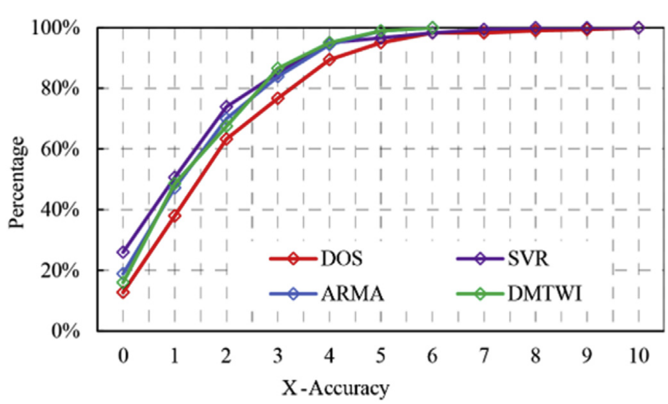

Fig. 14. X-accuracy of different models on a holiday.

a mobile phone enters sleep mode, data communications frequency is reduced and the device may not be discovered by the probe. Third, the data resolution and length of time windows need to be future discussed. The HVAC system should avoid too frequent adjustments for efficiency and stability consideration. In future work, it is essential to integrate the Wi-Fi probe with other conventional occupancy detection methods, such as environment sensors and motion sensors, to increase the scope and efficiency of indoor occupancy determination. Wi-Fi probe based occupancy detection also needs to be expanded and investigated in other building types, such as residential buildings, hospitals, shopping malls, etc. In residential buildings, the number of occupants are relatively small and more home appliances use Wi-Fi network. Such conditions create new challenges for occupancy detection, but it also has a significant research potential in using the human-appliance interactions to automatically detect and record occupants' behaviors.

6. Conclusion

This study utilized the time-series and stochastic characteristics of detected Wi-Fi request to develop a reliable and automatic occupancy detection and prediction mechanism. A DMTWI model to formulate the occupancy dynamics as a Markov process to measure the number of occupants in a given Wi-Fi space. MAC addresses of devices were used as the identification tags to differentiate occupants and facilities. Conventional SVR and ARMA were also used to benchmark the performance of the proposed model. An experiment was conducted to validate the measured occupancy schedule on three days using Wi-Fi probe devices. Cameras were also used to collect actual occupancy information for comparison. The experiment results show that the Wi-Fi probe can effectively reduce the detection fluctuation and DMTWI is more accurate when the error tolerance is high for weekdays and weekend days.

Acknowledgements

The work described in this paper was fully sponsored by the project JCYJ20150518163139952 supported by the Shenzhen Science and Technology Funding Programs (SSTFP) and the Hong Kong Environment and Conservation Fund (ECF) project #9211074 (25/2014),. 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 Hong Kong ECF.

References

[1] Z. Yang, B. Becerik-Gerber, Modeling personalized occupancy profiles for representing long term patterns by using ambient context, Build. Environ. 78 (2014) 23–35, http://dx.doi.org/10.1016/j.buildenv.2014.04.003.
[2] M. Trčka, J.L.M. Hensen, Overview of HVAC system simulation, Autom. Constr. 19 (2010) 93–99, http://dx.doi.org/10.1016/j.autcon.2009.11.019.
[3] F. Oldewurtel, D. Sturzenegger, M. Morari, Importance of occupancy information for building climate control, Appl. Energy 101 (2013) 521–532, http://dx.doi.org/10.1016/j.apenergy.2012.06.014.
[4] S. Goyal, P. Barooah, T. Middelkoop, Experimental study of occupancy-based control of HVAC zones, Appl. Energy 140 (2015) 75–84, http://dx.doi.org/10.1016/j.apenergy.2014.11.064.
[5] G. Lin, D.E. Claridge, A temperature-based approach to detect abnormal building energy consumption, Energy Build. 93 (2015) 110–118, http://dx.doi.org/10.1016/j.enbuild.2015.02.013.
[6] P. Zhou, G. Huang, Z. Li, Demand-based temperature control of large-scale rooms aided by wireless sensor network: energy saving potential analysis, Energy Build. 68 (2014) 532–540, http://dx.doi.org/10.1016/j.enbuild.2013.10.005.
[7] N.A. Nassif, Robust CO2-based demand-controlled ventilation control strategy for multi-zone HVAC systems, Energy Build. 45 (2012) 72–81, http://dx.doi.org/10.1016/j.enbuild.2011.10.018.
[8] V. Congradac, F. Kulic, HVAC system optimization with CO2 concentration control using genetic algorithms, Energy Build. 41 (2009) 571–577, http://dx.doi.org/10.1016/j.enbuild.2008.12.004.
[9] B.F. Warren, N.C. Harper, Demand controlled ventilation by room CO2 concentration: a comparison of simulated energy savings in an auditorium space, Energy Build. 17 (1991) 87–96, http://dx.doi.org/10.1016/0378-7788(91)90001-J.
[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, http://dx.doi.org/10.1016/j.apenergy.2013.01.039.
[11] 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, http://dx.doi.org/10.1016/j.autcon.2012.02.013.
[12] J. Yang, M. Santamouris, S.E. Lee, Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings, Energy Build. 121 (2016) 344–349, http://dx.doi.org/10.1016/j.enbuild.2015.12.019.
[13] W. O'Brien, H.B. Gunay, The contextual factors contributing to occupants' adaptive comfort behaviors in offices – a review and proposed modeling framework, Build. Environ. 77 (2014) 77–87, http://dx.doi.org/10.1016/j.buildenv.2014.03.024.
[14] S. Zikos, A. Tsolakis, D. Meskos, A. Tryferidis, D. Tzovaras, Conditional Random Fields - based approach for real-time building occupancy estimation with multi-sensory networks, Autom. Constr. 68 (2016) 128–145, http://dx.doi.org/10.1016/j.autcon.2016.05.005.
[15] G.Y. Geun Young Yun, H.J. Hyo Joo Kong, J.T. Jeong Tai Kim, A field survey of occupancy and air-conditioner use patterns in open plan offices, Indoor Built Environ. 20 (2011) 137–147, http://dx.doi.org/10.1177/1420326X10388883.
[16] B. Dong, B. Andrews, Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings, in: Eleventh International IBPSA Conference Glasgow, Scotland, 2009, pp. 1444–1451.
[17] A. Mahdavi, Patterns and implications of user control actions in buildings, Indoor Built Environ. 18 (2009) 440–446, http://dx.doi.org/10.1177/1420326X09344277.
[18] J. Zhang, G. Liu, R. Lutes, M. Brambley, Energy Savings for Occupancy- Based Control (OBC) of Variable- Air-volume (VAV) Systems, 2013. Virginia The United States of America.
[19] T. Ekwevigbe, N. Brown, V.H. Pakka, D. Fan, Real-time Building Occupancy Sensing for Supporting Demand Driven HVAC Operations, 2013. Int. Conf. Enhanc. Build. Oper.
[20] S. Wang, X. Jin, CO2-Based occupancy detection for on-line outdoor air flow control, Indoor Built Environ. 7 (1998) 165–181, http://dx.doi.org/10.1159/000024577.
[21] K. Christensen, R. Melfi, B. Nordman, B. Rosenblum, R. Viera, Using existing network infrastructure to estimate building occupancy and control plugged-in devices in user workspaces, Int. J. Commun. Netw. Distrib. Syst. 12 (4) (2014), http://dx.doi.org/10.1504/IJCNDS.2014.057985.
[22] T. Teixeira, G. Dublon, AS. A survey of human-sensing: methods for detecting presence, count, location, track, and identity, ENALAB Tech. Rep. 09-2010 Vol. 1 (1) (2010).
[23] A. Mavrogianni, M. Davies, J. Taylor, Z. Chalabi, P. Biddulph, E. Oikonomou, et al., The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments, Build. Environ. 78 (2014) 183–198, http://dx.doi.org/10.1016/j.buildenv.2014.04.008.
[24] 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, http://dx.doi.org/10.1016/j.enbuild.2007.01.018.
[25] M.S. Gul, S. Patidar, Understanding the energy consumption and occupancy of a multi-purpose academic building, Energy Build. 87 (2015) 155–165, http://dx.doi.org/10.1016/j.enbuild.2014.11.027.
[26] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408, http://dx.doi.org/10.1016/j.enbuild.2014.11.065.
[27] J.A. Davis, D.W. Nutter, Occupancy diversity factors for common university building types, Energy Build. 42 (2010) 1543–1551, http://dx.doi.org/10.1016/j.enbuild.2010.03.025.
[28] ASHRAE Standard 90.1-2007: Energy Standard for Buildings Except Low-Rise Residential Buildings. n.d.
[29] H.B. Gunay, W. O'Brien, I. Beausoleil-Morrison, A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices, Build. Environ. 70 (2013) 31–47, http://dx.doi.org/10.1016/j.buildenv.2013.07.020.
[30] K. Shan, Y. Sun, S. Wang, C. Yan, Development and In-situ validation of a multi-zone demand-controlled ventilation strategy using a limited number of sensors, Build. Environ. 57 (2012) 28–37, http://dx.doi.org/10.1016/j.buildenv.2012.03.015.
[31] T. Ekwevugbe, N. Brown, V. Pakka, D. Fan, Real-time Building Occupancy Sensing Using Neural-network Based Sensor Network, 2013, pp. 114–119, http://dx.doi.org/10.1109/DEST.2013.6611339, 2013 7th IEEE Int. Conf. Digit. Ecosyst. Technol., IEEE.
[32] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, Energy Build. 131 (2016) 132–141, http://dx.doi.org/10.1016/j.enbuild.2016.09.002.
[33] 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, http://dx.doi.org/10.1177/0037549713489918.
[34] B. Dong, B. Andrews, K.P. Lam, M. Hoynck, R. Zhang, Y.-S. Chiou, et al., An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy Build. 42 (2010) 1038–1046, http://dx.doi.org/10.1016/j.enbuild.2010.01.016.
[35] K. Sun, D. Yan, T. Hong, S. Guo, Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration, Build. Environ. 79 (2014) 1–12, http://dx.doi.org/10.1016/j.buildenv.2014.04.030.
[36] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simul. 4 (2011) 149–167, http://dx.doi.org/10.1007/s12273-011-0044-5.
[37] Z. Chen, J. Xu, Y.C. Soh, Modeling regular occupancy in commercial buildings using stochastic models, Energy Build. 103 (2015) 216–223, http://dx.doi.org/10.1016/j.enbuild.2015.06.009.
[38] E. McKenna, M. Krawczynski, M. Thomson, Four-state domestic building occupancy model for energy demand simulations, Energy Build. 96 (2015) 30–39, http://dx.doi.org/10.1016/j.enbuild.2015.03.013.
[39] J. Virote, R. Neves-Silva, Stochastic models for building energy prediction based on occupant behavior assessment, Energy Build. 53 (2012) 183–193, http://dx.doi.org/10.1016/j.enbuild.2012.06.001.
[40] 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, http://dx.doi.org/10.1016/j.pmcj.2016.02.001.
[41] Y. Wang, L. Shao, Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning, Build. Environ. 114 (2017) 106–117, http://dx.doi.org/10.1016/j.buildenv.2016.12.015.
[42] 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, http://dx.doi.org/10.1016/j.eswa.2014.04.011.
[43] 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, http://dx.doi.org/10.1016/j.enbuild.2014.07.053.
[44] 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, http://dx.doi.org/10.1016/j.enbuild.2011.12.037.
[45] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings, in: Proc. 11th ACM Conf. Embed. Networked Sens. Syst. - SenSys 13, ACM Press, 2013, http://dx.doi.org/10.1145/2517351.2517370.

[46] N. Sidiropoulos, M. Mioduszewski, P. Oljasz, E. Schaap EdwinSchaap, Open Wifi SSID Broadcast Vulnerability SSN Project Assessment 2012, 2012.
[47] 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, http://dx.doi.org/10.1016/j.energy.2015.09.002.
[48] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Springer New York, New York, NY, 1991, http://dx.doi.org/10.1007/978-1-4419-0320-4.
[49] C. Chatfield, The Analysis of Time Series: an Introduction, Chapman & Hall/CRC, 2004.
[50] Y. Zhao, W. Zeiler, G. Boxem, T. Labeodan, Virtual occupancy sensors for real-time occupancy information in buildings, Build. Environ. 93 (2015) 9–20, http://dx.doi.org/10.1016/j.buildenv.2015.06.019.
[51] R.H. Dodier, G.P. Henze, D.K. Tiller, X. Guo, Building occupancy detection through sensor belief networks, Energy Build. 38 (2006) 1033–1043, http://dx.doi.org/10.1016/j.enbuild.2005.12.001.
[52] S.S.K. Kwok, E.W.M. Lee, A study of the importance of occupancy to building cooling load in prediction by intelligent approach, Energy Convers. Manag. 52 (2011) 2555–2564, http://dx.doi.org/10.1016/j.enconman.2011.02.002.
[53] S. Wang, J. Burnett, H. Chong, Experimental validation of co2-based occupancy detection for demand-controlled ventilation, Indoor Built Environ. 8 (1999) 377–391, http://dx.doi.org/10.1177/1420326X9900800605.