2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)
Bari, Italy. October 6-9, 2019

Online Self-learning for Smart HVAC Control

Tzu-Yin Chao, Manh-Hung Nguyen, Ching-Chun Huang, Chien-Cheng Liang and Chen-Wu Chung

Abstract –**In this paper, we introduce an online-learning method to model the property of an office building. Unlike conventional control methods where the building property is modeled via a simulator or through offline learning, our building model is adaptively updated according to the dynamic response of a real environment. Upon the building model for environment prediction, the proposed action agent can control the heating, ventilation, and air conditioning (HVAC) system in a smarter way by scheduling the temperature reference point. To online learn the model and improve the agent, two practical and seldom discussed issues are addressed. The first challenge is data bias where the collected initial training dataset can only partially reveal the statistical mapping between the control input and the environment response. Hence, the trained model may lack generalization. To overcome the data bias issue, a data augmentation method is proposed to embed physical logic in order to train a proper initial model. Next, an online learning process is introduced to update the model generality during the system operation phase. The second practical issue is the constraints on agent exploration for discovering unknown data samples. During the business hours, to comfort employees, a control agent is not allowed to explore the possible controlling space randomly. To balance data collection and control stability, we introduce a hybrid control strategy that considers both the human control rule and the agent action. A confidence score of the agent model is also automatically estimated to determine a suitable control strategy finally. Our experiments have realized in an office building. The results outperform conventional methods and show its superior in terms of control stability.**

I. INTRODUCTION

The rise of the Internet of Things (IoT) has opened a door for big data collection and intelligent discovery [1]. Nowadays, sensor networks have been widely deployed in order to collect different kinds of sensing data for environment understanding. However, for each particular application, it is still challenging to analyze the sensing data and to create more add-on values. For instance, the heating, ventilation, and air conditioning (HVAC) system may cost up to 50% of the total energy-consuming in an office building. For energy saving, a smarter way to control the HVAC system thus becomes an emerging topic in an intelligent building. Some works [2], [3] have been introduced for HVAC control in recent years. Among these possible control aspects, the optimal scheduling for HVAC reference temperature setting is the major task of our interest due to its high relevance to energy saving.

A general framework of an HVAC control in a building includes two main parts. The first part is building modeling that aims to model the relationship between environment response and control factors. Conventional methods[4], [5], and [6] proposed different physical models to describe the thermal reaction of a building. The method in [4] described a building by modeling the heat-storage capacity and heat transmissibility. Later, the heat transfer was represented by an electronic circuit. While the equivalent circuit can help to build a physical model for testing, its system response may not characterize precisely the heat transfer. Instead, the authors in [5] model the transfer process by applying extra assumptions and simplify the building model as multiple linearized functions. The model parameters can be estimated by minimizing a mean-square error over a training dataset. Though the method is efficient, the model accuracy may not always be high due to the over-simplified assumptions. Similar to [5] but going further, the authors in [6] separated the HVAC system into four subsystems. Each subsystem was modeled by a parametric function whose parameters can be estimated by a calibration process.

While the aforementioned physical-based modeling is efficient, their frameworks usually require extra assumptions to complete the analysis. However, these assumptions may not be satisfied in a dynamic environment. To improve the model generalization, data-driven methods have been recently introduced. Instead of funding on the physical principles, the data-driven methods leverage the power of big data to understand the environment and build up its approximation model through training. In [7], the author uses a neural network to fit the collected data whereas the authors in [8] used a recurrent neural network to enhance the temporal consistency over sequential data. The data-driven framework requires few physical assumptions and well adapts to many kinds of building environments. However, its success highly relied on the completion and quality of the training dataset. It is expected that the dataset can cover all the possible building statuses under different control settings in a statistic fashion.

The second key module of an air condition system is the control agent that helps to manipulate the system according to

This work was partially supported by Ministry of Science and Technology of Taiwan under Grant No. 108-2634-F-009 -007, 108-2634-F-009-006, 107-2218-E-009-062, 106-2628-E-194-002-MY3 and 108-2811-E-194-501.

Tzu-Yin Chao is with the Dept. of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan. (e-mail: chaoziyin@gmail.com).

Manh-Hung Nguyen is with the Dept. of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan and HCMC University of Technology and Education, Vietnam. (e-mail: nmhung.spkt@gmail.com).

Ching-Chun Huang is with the Dept. of Computer Science, National Chiao Tung University, Hsinchu, Taiwan.
(e-mail: chingchun.huang6@gmail.com, mobile: +886919259377).

Chien-Cheng Liang is with the Public Utilities Service Department, Taiwan Semiconductor Manufacturing Company, Ltd.
(e-mail: cclianga@tsmc.com).

Chen-Wu Chung is with the Public Utilities Service Department, Taiwan Semiconductor Manufacturing Company, Ltd.
(e-mail: chenwu_chung@tsmc.com).

the building model. Jun et al. [9] used a Proportional Integral Derivative (PID) control to keep the indoor temperature stable. The hyper-parameters Kp, Ki, and Kd of the PID controller were estimated automatically by particle swarm optimization (PSO) [10]. The PID control is simple and efficient; the low deployment cost is also its advantage. However, PID control adjusts the control parameters based on the error feedback simply without understanding the property of an environment. When we aim to optimize multiple objective targets, such as creating a comfortable working space and saving more energy at the same time, the basic PID control may not be a suitable way to approach the goal. To control HVAC under multiple objective constraints, rule-based control has been manually designed by experienced experts in [11]. The control rules are well tuned to fit a specific building; for other buildings, extra tuning and human efforts are necessary. An alternative solution is to use a fuzzy [12] method to embed expert-knowledge into the controller. Although the control methods work well, the requirement of experienced tuning makes them less flexible.

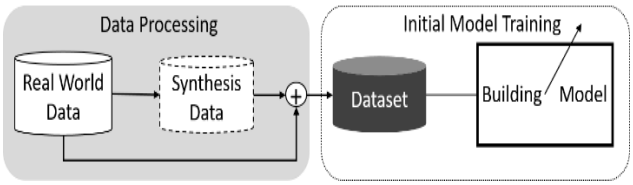
Most of the PID-based control methods discussed before focus on tacking quick actions for the sensor responses. Nowadays, a few model-based control strategies, so-called Model Predictive Control (MPC) [13], have been proposed to target at acting according to prevision. To realize the concept, building modeling and agent control should be co-designed. Particularly, the building model is used for the environment prediction; the agent makes decisions upon the model prevision. In MPC, the impact of future disturbances, such as outdoor temperature, weather conditions, and statuses of devices, could be considered by using model forecasts. The agent then selects an optimal control action at a regular time interval to satisfy the objective goal over time.

Besides MPC, the data-driven HVAC control methods [14], [15], and [16] upon reinforcement learning (RL) are of interest. These RL control methods aim to make the right decision at each moment in order to maximize the long-term future reward. By using reinforcement learning, multiple control objectives can be embedded into a unique reward to guide the learning of the decision agent. In [14] and [15], the classical Q-learning techniques are presented to train the control agent. By interacting with the environment to receive the rewards, the tabular Q value function would be updated successively. The optimal control action is then determined according to the Q table. However, the table-based Q-learning may not be suitable for our HVAC control problem due to the huge state and action space. To solve the problem, deep reinforcement learning (DRL) can be a rescue by using the Deep Neural Network (DNN) to approximate the Q function [16]. Even so, the standard RL framework requires many trial-and-error explorations before learning a converged and stable policy agent. Practically, an environment simulator is therefore needed during the training process. However, for a dynamic environment such as an office, the environment property is changed over time. This time-variant property limits the application of the RL-based methods on HVAC control.

In this paper, to capture the dynamic property of a real building, we proposed an online method to adaptively learn the environment model over time. Meanwhile, based on the environment model, a control agent was designed in order to determine the optimal reference temperature setting for HVAC control. However, both the model and the agent are not ready in the initial stage. To consider the model learning and the HVAC control simultaneously, we also proposed an adaptive strategy to balance control decision and training data exploration. The technical challenges and our method would be detailed in the following section. Below, we briefly summarize the main features of the proposed control method.

- As shown in Fig. 1, our framework includes two learning phases. The first stage is an initial stage that learns a building model based on an offline collected dataset. The second stage is to update the model during the online operation period so that our model can be improved day by day. To the best of our knowledge, still few works try to model the system response of a real building through an online learning procedure. For supplementary, we provide more dataset description and experimental results in [17].
- In the initial training stage, we address the data bias issue where the initially-collected training data only partially reveals the statistical mapping between the control setting and the environmental response. With the help of the proposed data augmentation, we embed physical logics as priors for initial training. Hence, the imperfect building model can be applied to the online learning procedure without unexpected side effects. Although the initial model is imperfect, it plays an important role in the second stage regarding data exploration and model converging.
- In the online learning stage, we introduce a hybrid control strategy that determines the HVAC reference temperature point by considering both the default setting from the experienced experts and the designed agent. Also, the control strategy is able to balance the agent exploration for online training and the control stability for comforting employees. Our system is operated and evaluated in a sensor-deployed office belonged to a semiconductor manufacturing company in Taichung, Taiwan. The results point out that our agent is progressively improved over time. Eventually, our agent shows more efficiency than the conventional controllers in terms of both the comfort level and energy saving.

(a)



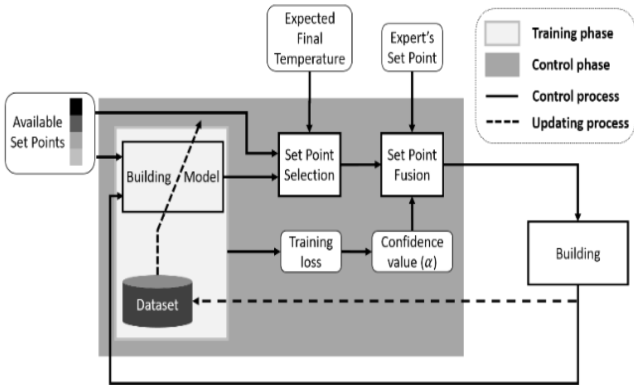
(b) 

Figure 1. The proposed method includes two learning phases. (a) The initial phase and (b) the online learning and control phase.

II. PROBLEM STATEMENT AND CHALLENGES

In the office building, the Honeywell Building Automation System (BAS) is used to manage the HVAC system. As shown in Fig. 2, the cooling system includes a Make-up Air Unit (MAU), an Air Handler Unit (AHU), many Variable Air Volume systems (VAV), and Fan Coil Units (FCU). MAU helps to process the fresh air from outdoors; AHU is used for cooling the return air to the temperature setpoint; VAV controls the air volume and recycles some air from the room; FCU is an independent cooling system for reducing the radiant heat from windows. Note that the temperature measured at each VAV can be considered as a sample of the terminal temperature of the nearby working space. To provide a comfortable working environment, the task of the operator (or agent) is to keep all the VAV temperatures stable and close to the expected temperature level (25.25°C in this work) by controlling the AHU setpoint.

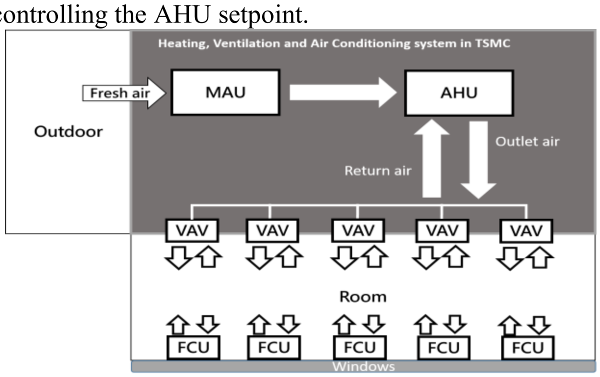


Figure 2. The HVAC system in an office building.

To systematically manage the HVAC system, the end users are not allowed to directly adjust the VAV temperature setting in the office. Instead, a centralized control schedule for the AHU reference temperature setting is designed. Before introducing our learning-based agent for assistance, a safer way for operation is to determine the reference temperature setting according to the experience of HVAC experts and the feedbacks from staffs. A workable AHU setting schedule (also named as the expert mode in this paper) is presented in Fig. 3. By adding the agent into the HVAC system, our first goal is to explore a better schedule for energy saving seamlessly without complaints from staffs. The second goal is to understand the dynamic property of the building and update the AHU setting schedule accordingly.

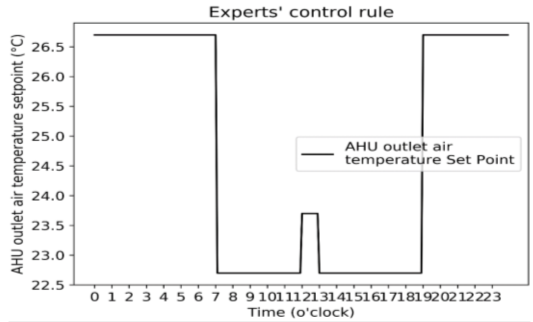


Figure 3. AHU temperature setting schedule designed by experts. 22.7°C for 7:00 to 12:00 and 13:00 to 19:00; 23.7°C during 12:00 to 13:00; 26.7°C from 19:00 to 7:00 to save energy at the off-working time.

To fulfill the goals, a building model is needed that can predict the future VAV temperature given the status of the current environment and the AHU temperature setpoint. Hence, we recorded the sensor responses, weather conditions, and the expert-defined setting schedule over time to form the initial training dataset. However, the initial dataset has a data bias issue. As we can find in Fig. 3, the temperature set points within the expert's schedule are sparse. Thus, the collected training dataset can only partially reveal the statistical mapping between the control and environment response. If the model is learned purely based on the bias data, the future prediction and the agent decision would be wrong.

To reduce the data bias problem, we are prone to collect more online control data for model updating. However, the constraints on agent exploration for discovering unknown samples become the second practical issue. Note that during the business hours, to comfort employees, a control agent is not allowed to explore the possible controlling space randomly. Thus, the trade-off between data exploration and a stable operation raises a chicken-egg problem. To address the issue, a well-designed method for online learning becomes an emergent task.

III. PROPOSED METHOD

As before mentioned, our method includes two learning stages. The two-stage designs allow our framework to work in a real building without any risk and enable our model to be updated adaptively. Below, we explain each stage in detail.

A. Initial stage

The first stage aims to train an initial NN-based building model by using the offline training data. The input of the NN model is composed of s_t, e_t, and a_t. Particularly, the current state s_t is a vector that represents the terminal temperatures of many VAVs at time *t*. The uncontrollable environmental factor e_t is a vector including outdoor temperature, weather, condition, power supply in the office, and the status of devices. The control action a_t represents the AHU temperature setpoint. The output is the future VAV temperature state $s_{t+\Delta T}$ at time $t+\Delta T$. Thus, a training sample is represented as the vector format $[s_t,\ e_t,\ a_t,\ s_{t+\Delta T}]$.

In our work, since we have limited initial training samples, we design a four-layer NN with tanh activation functions to avoid the over-fitting problem. Besides, due to the 31 features in e_t share some dependency, we reassemble these features by principal component analysis (PCA) [18] to remove the

redundancy. After PCA, the feature number is reduced from 31 to 13. The overall model structure is summarized in Fig. 4. The input and out layers have 19 and 5 nodes. The node number for the two hidden layers is 15 and 10. The function of the NN, which model the state transition from t to $t+\Delta T$, can then be defined as

$$\hat{s}_{t+\Delta T} = F(s_t, \mathrm{PCA}(e_t), a_t). \quad (1)$$

In (1), according to the experts' experience, 18 minutes are enough for the terminal VAV temperature converged to the steady state after tuning the AHU setpoint. Thus, we select ΔT=18minutes in our application.

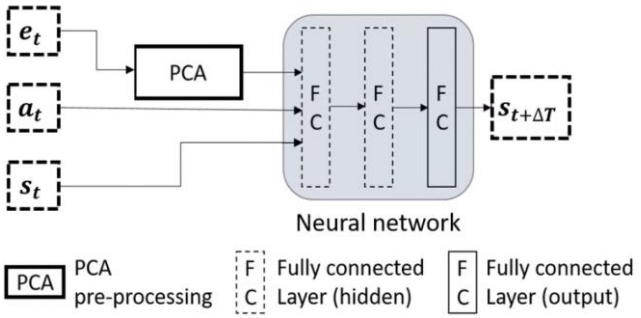


Figure 4. The structure of our proposed simulation model.

For training, a normalization process is necessary. Note that the scale of different sensors can be very different. A preprocessing step to normalize and unified the range of features would make the training procedure easier. Next, to train our NN given the initial dataset, we minimize the following regression loss.

$$Loss = \|s_{t+\Delta T} - F(s_t,\ PCA(e_t),\ a_t)\|^2 \quad (2)$$

However, our experiment in section IV.A shows that even the model is well trained, the model shows an inverse tendency where the predicted VAV temperature become lower when we increase the AHU setpoint. This adverse effect is caused by the data-bias initial dataset. To address the issue, we follow the basic physical properties of a cooling system to generate more synthesized training samples. Through data augmentation, we have more samples for training. Meanwhile, these samples also help to amend the model tendency.

The synthesized samples can be generated by algorithm 1. In the algorithm, a_t^{syn} is the synthesized AHU setpoint. a_t^{real} is the real AHU setpoint from the initial dataset. $\hat{s}_{t+\Delta T}^{syn}$ is the synthesized VAV temperature. $\hat{s}_{t+\Delta T}^{real}$ is the real VAV temperature. In addition, α_1 and α_2 are two controllable coefficients. In our algorithm we set $\alpha_2 = 0.6$ to simulate the fast-active-cooling property and set $\alpha_1 = 0.4$ to simulate the slow-passive-heating property.

TABLE I. ALGORITHM 1. DATA SYNTHESIS

Algorithm 1. Data synthesis

Input: Real dataset
Output: Synthesis dataset

Set $\alpha_1 = 0.4, \alpha_2 = 0.6$.
Set $A^{syn} \leftarrow$ Synthesized AHU setpoints $\{a_1^{syn}, a_2^{syn}, \ldots a_n^{syn}\}$.
Set $U^{real} \leftarrow$ {The initial dataset}
Set $U^{syn} \leftarrow \{.\}$
For every sample $D_t^{real} = [\hat{s}_t^{real},\ e_t^{real},\ a_t^{real}, \hat{s}_{t+\Delta T}^{real}] \in U^{real}$:
  For every synthesized AHU setpoint $a_k^{syn} \in A^{syn}$:
    $\hat{s}_{t+\Delta T}^{syn} =$
    $\begin{cases} \hat{s}_{t+\Delta T}^{real} + \alpha_1 * (a_k^{syn} - a_t^{real}) & if\ (a_k^{syn} - a_t^{real}) > 0 \\ \hat{s}_{t+\Delta T}^{real} + \alpha_2 * (a_k^{syn} - a_t^{real}) & if\ (a_k^{syn} - a_t^{real}) < 0 \end{cases}$
    $D_{tk}^{syn} = [\hat{s}_t^{real},\ e_t^{real},\ \boldsymbol{a_k^{syn}}, \boldsymbol{\hat{s}_{t+\Delta T}^{syn}}]$
    $U^{syn} \leftarrow U^{syn} \cup \{D_{tk}^{syn}\}$
  endfor
endfor

B. Operation stage with online Learning

Ideally, given the temperature state s_t and the current uncontrollable environmental factor e_t, the building model can help to predict the future temperature state $\hat{s}_{t+\Delta T}$ if the building model has been well-trained. Therefore, we can decide the optimal AHU setpoint that can minimize the error between the VAV-temperature prediction and the expected temperature. Following the idea, the optimal AHU setpoint $\mathrm{a}_{t|\ model}^*$ is determined by algorithm 2. In algorithm 2, A is the set of possible temperature set points. The acceptable error is set to be 0.25°C. $\mathrm{a}_{t|\ model}^{1\mathrm{st}}$ and $\mathrm{a}_{t|\ model}^{2\mathrm{nd}}$ represent the setpoints which are expected to achieve the smallest error and the second smallest error. In order to save energy, we would select $\mathrm{a}_{t|\ model}^{2\mathrm{nd}}$ as $\mathrm{a}_{t|\ model}^*$ in the case that the expected error is acceptable and its setpoint is higher than $\mathrm{a}_{t|\ model}^{1\mathrm{st}}$.

TABLE II. ALGORITHM 2. OPTIMAL AHU SETPOINT SELECTION

Algorithm 2. Optimal AHU setpoint selection

Input: $\boldsymbol{s_{t+\Delta T}^*, s_t, e_t, a_t, Acceptable\ error}$
Output: $\mathbf{a}_{t|\ model}^*$

Error_set = {}
For every AHU setpoint $\mathbf{a_t} \in \boldsymbol{A}$
  $\mathrm{Error}_{\mathrm{a_t}} = \|F(s_t, PCA(e_t), a_t) - s_{t+\Delta T}^*\|$
  Error_set = **Error_set** $\cup$ $\{\mathrm{Error}_{\mathrm{a_t}}\}$
$\mathbf{a}_{t|\ model}^{1\mathrm{st\ min}}$, $\mathbf{a}_{t|\ model}^{2\mathrm{nd\ min}}$ = **Argmin**$_{a_t}$(Error_set)
If($\mathbf{a}_{t|\ model}^{2\mathrm{nd\ min}} > \mathbf{a}_{t|\ model}^{1\mathrm{st\ min}}$ ***and*** $\mathbf{Error}_{\mathbf{a}_{t|\ model}^{2\mathrm{nd}}} <$ ***Acceptable error*****):**
  $\mathrm{a}_{t|\ model}^* = \mathrm{a}_{t|\ model}^{2\mathrm{nd\ min}}$
Else:
  $\mathrm{a}_{t|\ model}^* = \mathrm{a}_{t|\ model}^{1\mathrm{st\ min}}$
endif

However, in the initial stage, the learned building model is not yet perfect. To avoid the wrong setpoint assignment, we proposed to dynamically evaluate the model confidence. The evaluation metric can be measured by the accuracy of the model prediction in a weighted average fashion. Denote that $\mathrm{a}_{t-\Delta T}$ is the setpoint action, $e_{t-\Delta T}$ is the uncontrollable environmental factor, $s_{t-\Delta T}$ is the previous temperature state at time $t-\Delta T$, and s_t is the real state observation at time t. The model confidence α_t is then measured by:

$$\alpha_t = \beta exp\left(\frac{-\|F(s_{t-\Delta T}, PCA(e_{t-\Delta T}), a_{t-\Delta T}) - s_t\|^2}{\sigma^2}\right) + (1\text{-}\beta)\ \alpha_{t-\Delta T} \quad (3)$$

In (3), β and σ are tunable parameters; $\alpha_{t-\Delta T}$ is the model confidence at $t-\Delta T$. At the beginning, we set $\alpha_0 = 0.5$.

In the case that the model prediction is not trustable, we borrow the expert's experiment to setup the AHU. Therefore, a hybrid control method is proposed in (4).

$$\hat{\mathrm{a}}_t = \alpha_t * \hat{\mathrm{a}}_{t|\ model} + (1-\alpha_t)\hat{\mathrm{a}}_{t|\ expert}. \quad (4)$$

Based on the confidence value α_t, we determine the ratio to fuse the setpoint determined by our agent ($\hat{\mathrm{a}}_{t|\ model}$) and the setpoint recommended by the experts ($\hat{\mathrm{a}}_{t|\ expert}$). Note that $\hat{\mathrm{a}}_{t|\ model}$ is determined by Algorithm 2 and $\hat{\mathrm{a}}_{t|\ expert}$ is defined in Fig. 3. In fact, the hybrid control plays important role for both training data exploration and energy saving.

When the model confidence α_t is close to 1, our agent aims to find the optimal AHU setpoint for energy saving and keeping space comfortable. When the model goes wrong, possible due to sudden weather or season changes, the term $\hat{a}_{t|\ expert}$ would make sure the setpoint can comfort the staffs without any risk; meanwhile, the term $\hat{a}_{t|\ model}$ would guide the agent for data exploration so that the model can be refined and adapted to the new environment again. Furthermore, the model confidence α_t can be a useful tool to understand the system performance through the on-line learning procedure.

Furthermore, when the proposed agent starts to control the AHU setpoint over time, more new training sample $[s_t, e_t, a_t, s_{t+\Delta T}]$ can be collected. By adding more new samples into the training dataset, our NN model can be online updated over time by minimizing the regression loss similar to (2).

IV. Experimental Results

Our experiments had been set up on the 3rd floor of the building. As shown in Fig. 5, the office area includes six rooms and one open space for working. Our system aims to control the terminal temperature of a 3500 m^2 area covered by the five VAVs denoted by the red rectangle in Fig. 5. At each decision moment, we acquire 37 observations as our system input including 31 uncontrollable environmental sensor feedbacks (e_t), 5 VAV status (s_t), and one AHU setting point (a_t). The description of each input is summarized in Table IV. Our dataset includes two parts: (1) the initial two-week data under the expert's control and (2) the 4-month data under our on-line control. For every 18 minutes, we record a sample in the vector format $[s_t, e_t, a_t, s_{t+\Delta T}]$. For every week, we have 560 samples.

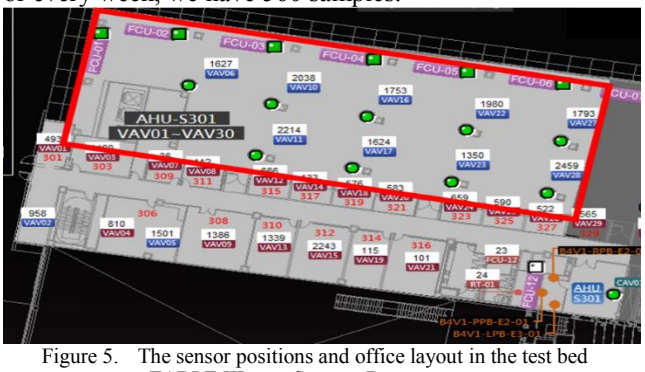


Figure 5. The sensor positions and office layout in the test bed

TABLE III. Sensor Description

Sensor type	Sensor # / input type		Description
Power supplied	2	e_t	Shown as B4V1-PPB-E2-01 and B4V1-LPB-E3-01 in Fig. 5. The sensors read the amount of the supplied power.
FCU on/off	5	e_t	The device status of each FCU.
FCU coil water valve opening	5	e_t	The opening level of the coil water valve in FCU.
FCU return air temperature	5	e_t	The return air temperature of FCU. The sensor is located near the window.
VAV throttle opening	5	e_t	The sensors, set on the ceiling, measure the throttle opening of the VAVs.
VAV on/off	5	e_t	The device status of the VAVs.
VAV temperature	5	s_t	The return air temperature of VAVs
AHU static pressure	1	e_t	The sensor, within the AHU, used to measure the static pressure in the air duct.
AHU set point	1	a_t	The AHU set point for the outlet temperature.
MAU outlet temperature	1	e_t	The sensor in the MAU used to measure the outlet air temperature in an air duct.
Outdoor temperature	1	e_t	The sensor in the MAU used to measure the temperature of outdoor air.
Raining sensor	1	e_t	The sensor outside of the building used to measure the binary weather condition (Sunny=0, Rainy=1).

To evaluate the robustness of the proposed method, we discuss the influence of the proposed data augmentation in the initial learning stage. Later on, we demonstrate the ability of our model for online learning and online control. Finally, we compare our method with some related works and with the expert-based HVAC control.

A. Data augmentation for the initial model training

We use the data collected in the first 10 days for the initial phase. The other 4-day data is used for training the NN model illustrated in Fig. 4; the second week is used for testing. Due to the expert's scheduling, the data collected in the two weeks has the severe biased problem where the AHU temperature settings are distributed around 22.7°C (day time) or 26.7°C (night time). To explain the functionality of data augmentation, we apply a model tendency test to evaluate the learned NN models with and without data augmentation. For the tendency test, we can select one testing sample and input its 31 uncontrollable environmental factors and 5 current VAV statuses into the two NN models. Next, if we tune the AHU setting point from 20°C to 27°C, we can plot the output VAV temperatures estimated by the models accordingly. Here, we use 560 testing samples to generate 560 tendency curves. These curves are finally used to estimate the mean and standard deviation of the tendency curve over the control temperature in Fig. 6.

Without the help of data augmentation, we may find the output VAV temperature decreases when we rise the AHU setpoint. It may violate the physical property. However, if we check the initial training dataset and find the output VAV temperature is around 24.5°C in the night time when the AHU setpoint is around 27°C, we may not feel surprised of the tendency. We should also explain that the VAV temperature is lower mainly because of the low outdoor temperature during the night time. In contrast, by applying the data augmentation, we see the new model tendency of the VAV temperature increases along with the rising AHU temperature.

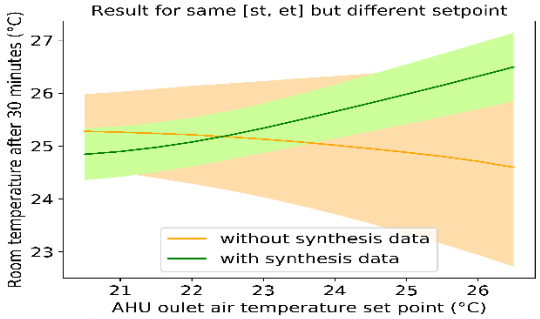


Figure 6. The tendency test of the models trained with and without data augmentation.

To show the control logic of the trained initial model, we perform a verification test by using 12-hour testing date. From the testing data, as shown in Fig. 7, we plot the expected

terminal temperature (orange), the AHU control schedule determined by the expert, and the final terminal temperature distribution (green) after the expert's control. We can find the terminal temperature is much lower than the expected one. In this case, a qualified agent should increase the AHU setpoint to not only save energy but meet the expected temperature. At the same time, we plot the AHU control schedules determined by the two initial NN models. Remark that the optimal AHU setpoint provided by the two models is decide by Algorithm2 illustrated in Table II. By comparison the control schedules, we see the model trained with data augmentation recommends to increase the AHU temperature setpoint when the terminal VAV temperature is lower than the expected one. In contrast, without the help of data augmentation, the model gives an incomprehensible AHU schedule. Even though we do not truly apply the models to control the real HVAC system in this test, the analysis also shows the risk of simply trusting the data without verification during training. From this viewpoint, data preprocessing plays an important role in our design.

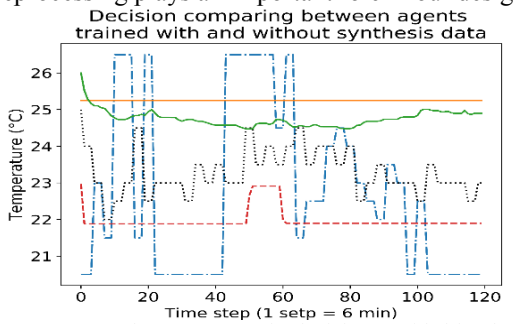


Figure 7. Compare the AHU control schedules provided by the agent trained with and without data augmentation. Orange line: the expected temperature. Green line: the final terminal temperature. Gray/Blue dashed line: model control with/without data augmentation. Red dashed line: the expert's control.

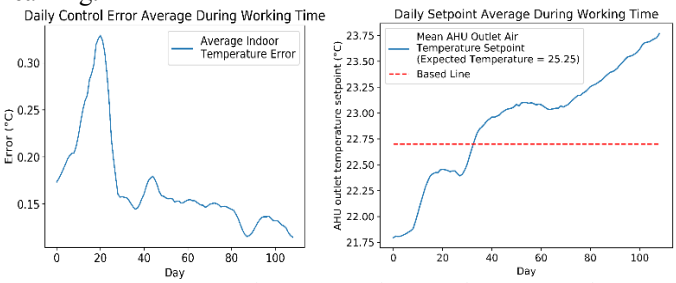
B. Evaluation of on-line learning and control

Next, we test our method in terms of on-line control and learning in the real office. For evaluation, we calculate two metrics, VAV Temperature Error (VTE) and Energy-saving Level (ESL). Denote T_i is the working time (7am~19pm) and n_i is the number of sampling times in the ith day; the VTE_i and ESL_i for the ith day is estimated by (7) and (8) respectively.

$$VTE_i = \frac{1}{n_i}\sum_{t\in T_i}\|s_t - s_t^*\| \quad (7)$$

$$ESL_i = \frac{1}{n_i}\sum_{t\in T_i} a_t \quad (8)$$

A smaller VTE_i means the terminal temperature is more stable and closer to the expected temperature. A higher ESL_i, or equally a higher mean AHU setpoint, indicates less energy consuming. The distributions of VTE_i and ESL_i over 108 days are plotted in Fig. 8. Our results show the VTE_i decreases gradually. It means the prediction of our model is getting better and better owing to the on-line learning. Also, based on the updated environment model, our agent can recommend suitable AHU setpoints to reduce the VTE_i. Moreover, after 30-day on-line learning, Fig. 8(b) implies that our agent can save more energy by increasing the AHU setpoint temperature. After three working months, our agent can ensure the office comfortable and save more energy. In addition, the real office space contains many uncertainties such as human activities, the status of electrical devices, etc. Even so, the result of the 4-month on-line control also shows the robustness of our method in a complex environment. To further understand the function of our online learning and its contribution to energy saving, we plot two AHU control schedule in Fig. 9. One is the schedule output by the initial agent; the other one is the AHU schedule determined by a well-trained agent after 90-day online learning. Again, we can see the agent is able to save more energy and keep the terminal temperature around the expected temperature after online learning.



a. VTE errors over 108 days b. ESL values over 108 days

Figure 8. The evaluation of our on-line control and learning model.

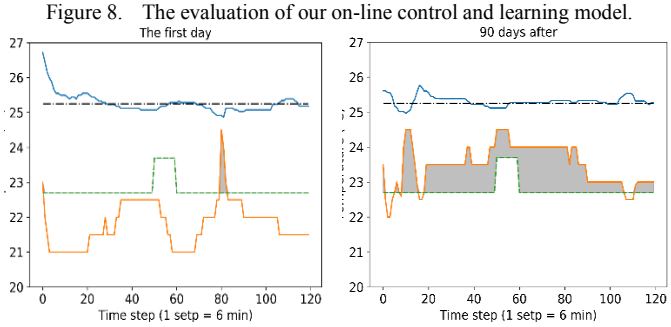


Figure 9. Compare the AHU control schedules on the first day and 90-day later. The period where our AHU setpoint is higher than the expert's schedule is marked by gray. A larger gray area implies a better ability for energy saving. After 90-day online learning, the agent is able to save more energy and keep the terminal temperature around the expected 25.25°C.

C. Performance comparison

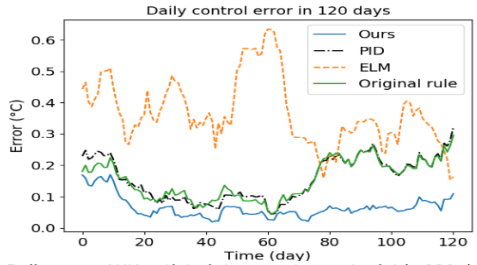
One main contribution of this work is to deploy our online learning and control system in a real building. Though the result is promising, a comparison with conventional methods can further help us to assess our approach. To have a fair comparison and a repeatable test, we need an environment simulator rather than a real environment to test different control methods. In this experiment, instead of using commercial software to simulate a virtual office, we train a deeper NN (5 layers) based on the 4-month dataset to simulate the property of the office. The neural node numbers are 37, 128, 64, 32, and 5 for each layer. The activation function is Leaky-Relu. We select 70% of the 4-month data to train the simulator and 30% data for testing. The mean square error between the outputted VAV temperatures and the group truth of the testing data is 0.08°C. The small testing error may show that the deep NN has captured the office property. In the following test, we treat the trained deep NN as the real environment and apply different control methods on it for comparison.

Based on the simulator, we calculate VTE_i and ESL_i over days for four methods including ELM [7], PID, the

expert's control and ours. The ELM [7] method is also NN-based but has no design for online learning. PID control is the most popular strategy for HVAC control. Based on the error feedback, the method would dynamically and frequently adjust the setpoint without understanding the property of an environment. The VTE_i distribution in Fig. 10(a) shows that our method has smaller control errors. The error variance of our method, the expert control, PID, and EML are 0.0045, 0.0107, 0.0116, 0.0505, respectively. Meanwhile, from the downtrend tendency of our VTE_i distribution, we see our method is able to reduce the error over time due to online learning. These results may mean our model is more robust and could adapt to a dynamic environment.

In addition, Fig. 10(b) shows the ESL_i distributions of different methods over days. If compared with PID and the expert control methods, our method continuously evolves and eventually is able to recommend higher setpoints during the working time. Although EML can achieve higher ESL_i, the method also produces an unstable VTE_i distribution with larger errors. On the contrary, our model can save energy while keeping the office comfortable. The distribution trend of our ESL_i distribution also proves the online learning strategy is functional.

(a)



(b)

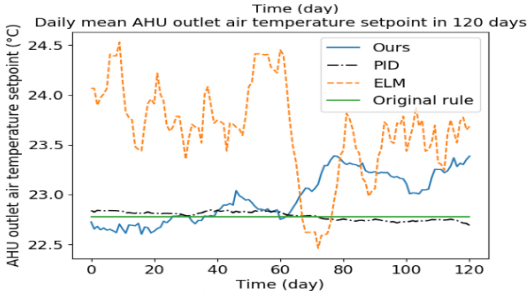


Figure 10. Performance comparison.
(a)The VTE distributions of different control methods over 120 days.
(b)The ESL distributions of different control methods over 120 days.

V. Conclusion

We proposed an online learning method to control an HVAC system in a real building by scheduling the AHU temperature reference point over time. The online learning method includes two phases. The first phase aims to train a building model from a small dataset with biased and limited training samples. To overcome the challenge, we augment synthesized data to embed some physical constraints as priors for training the initial model. In the operation phase, the control agent has been updated day by day. The on-line control in the real building shows that the terminal temperature is more stably close to the expected temperature. At the same time, the determined AHU temperature reference points are much higher than the original expert setting. It proves that our agent can utilize energy in a more efficient manner while maintaining the environment AC quality. The long-term evaluation (more than 100 days) also show the robustness of our system regarding the learning ability and environment adaptively.

Acknowledgement

This work was supported by the Center for Innovative Research on Aging Society (CIRAS), National Chung Cheng University and the Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by Ministry of Education (MOE) in Taiwan.

References

[1] N.Wu, Z.Li, K.Barkaoui, X.Li, T.Murata, andM.Zhou, "IoT-based smart and complex systems: A guest editorial report," *IEEE/CAA J. Autom. Sin.*, 2018.
[2] D. W. U.Perera, C.Pfeiffer, andN.-O.Skeie, "Control of temperature and energy consumption in buildings - A review," *Igarss 2014*, vol. 5, no. 1, pp. 1–5, 2014.
[3] M.Han *et al.*, "A review of reinforcement learning methodologies on control systems for building energy A review of reinforcement learning methodologies on control systems for building energy," *Work. Pap. Transp. Tour. Inf. Technol. microdata Anal.*, pp. 1–26, 2018.
[4] M.Maasoumy, A.Pinto, andA.Sangiovanni-Vincentelli, "Model-Based Hierarchical Optimal Control Design for HVAC Systems," 2012.
[5] A.Aswani, N.Master, J.Taneja, A.Krioukov, D.Culler, andC.Tomlin, "Energy-efficient building HVAC control using hybrid system LBMPC," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2012.
[6] Y.Ma, F.Borrelli, B.Hencey, B.Coffey, S.Bengea, andP.Haves, "Model predictive control for the operation of building cooling systems," *IEEE Trans. Control Syst. Technol.*, 2012.
[7] G. T.Costanzo, S.Iacovella, F.Ruelens, T.Leurs, andB. J.Claessens, "Experimental analysis of data-driven control for a building heating system," *Sustain. Energy, Grids Networks*, vol. 6, pp. 81–90, 2016.
[8] Y.Chen, Y.Shi, andB.Zhang, "Modeling and optimization of complex building energy systems with deep neural networks," in *Conference Record of 51st Asilomar Conference on Signals, Systems and Computers, ACSSC 2017*, 2018.
[9] Z.Jun andZ.Kanyu, "A particle swarm optimization approach for optimal design of PID controller for temperature control in HVAC," in *Proceedings - 3rd International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2011*, 2011.
[10] F.Marini andB.Walczak, "Particle swarm optimization (PSO). A tutorial," *Chemom. Intell. Lab. Syst.*, 2015.
[11] D.Subbaram Naidu andC. G.Rieger, "Advanced control strategies for heating, ventilation, air-conditioning, and refrigeration systems - An overview: Part I: Hard control," *HVAC and R Research*. 2011.
[12] E. H.Mamdani, "Advances in the linguistic synthesis of fuzzy controllers," *Int. J. Man. Mach. Stud.*, 1976.
[13] C.Camacho andE.Bordons, *Model Predictive Control*, 2nd ed. Springer-Verlag London Ltd., 2007.
[14] S.Liu andG. P.Henze, "Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: Part 1. Theoretical foundation," *Energy Build.*, 2006.
[15] S.Liu andG. P.Henze, "Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory Part 2: Results and analysis," *Energy Build.*, vol. 38, no. 2, pp. 142–147, 2006.
[16] T.Wei, Y.Wang, andQ.Zhu, "Deep Reinforcement Learning for Building HVAC Control," 2017, pp. 1–6.
[17] "Online Self-learning for Smart HVAC Control and Energy Saving Demo Website," 2019. [Online]. http://acm.ee.ccu.edu.tw:2020/. [Accessed: 19- Jul- 2019]
[18] L.Smith, "A tutorial on Principal Components Analysis," *Commun. Stat. - Theory Methods*, 2002.