7th International Building Physics Conference

IBPC2018

Proceedings

SYRACUSE, NY, USA

September 23 - 26, 2018

Healthy, Intelligent and Resilient Buildings and Urban Environments

ibpc2018.org | #ibpc2018

LightLearn: Occupant centered lighting controller using reinforcement learning to adapt systems to humans

June Young Park, Thomas Dougherty, Hagen Fritz and Zoltan Nagy*

Intelligent Environments Laboratory,
Department of Civil, Architectural and Environmental Engineering,
The University of Texas at Austin, Austin, TX, USA

**Corresponding email: nagy@utexas.edu*

ABSTRACT
Humans spend up to 90% of their time indoors, thus building systems should maintain the indoor environment within the comfort range. In this paper, we present LightLearn, a reinforcement learning based occupant centered lighting controller. The control agent interacts with the occupant non-intrusively, learns her/his preferences, and determines actions for achieving both human comfort and energy saving. We present system hardware, control algorithm, and experimental results of LightLearn for an office space. Compared to the full (9am-5pm) and occupancy based control, LightLearn reduced 83% and 63% of operation time, respectively, by adapting to the occupant.

KEYWORDS
building retrofit, occupant behavior, adaptive control, reinforcement learning

INTRODUCTION
Lighting accounts for approximately 20% of the total commercial building energy consumption (DoE 2011). One promising energy efficient strategy is a daylight linked automatic lighting control system, which switches lights on/off depending on the amount of natural lighting. The challenge for this automatic control system, however, is that typically control set-points are not driven by occupant preference but set to a uniform value. In other words, even though different people may have different preference on their comfort, the conventional automatic control uses fixed threshold values for operation. As a result, occupants often either override control settings indicating their discomfort, or in the worst case, they deactivate their control system (Gunay et al. 2013). Note that if occupant comfort is not satisfied, it could have serious repercussions on not only work productivity but also both physical and mental health for humans (Loftness et al. 2003). To put this in a more general perspective, beyond lighting: the objective of building environmental systems should be to provide comfort for the occupants, who spend up to 90% of their time indoors. However, building control research has largely neglected occupant satisfaction (Park and Nagy, 2018).

The occupant centered control system (OCC) has been introduced to address this discrepancy between building control and occupant comfort (Nagy et al. 2015; Nagy et al. 2016). There are two components of occupant comfort, one of which is the actual physical comfort, i.e., the absence of the feeling of pain, and the other is the perception of control over their indoor environment (Nagy et al. 2016). To accomplish the two parts of comfort, the building control system should be focused, or centered, on the characteristics of the occupant. An ideal system should learn the unique preferences of the occupant, and adapt the control set-points to these preferences to balance the decisions of the occupant with the calculated decision for energy savings. There are several experimental results indicating that OCC approach can save energy without sacrificing occupant comfort (Nagy et al. 2015; Nagy et al. 2016).

In this paper, we introduce LightLearn, a reinforcement learning (RL) controller for lighting in the OCC framework. We describe RL and the OCC framework in general, as well as the specific hardware and control algorithm of LightLearn in the next Section. Finally, we present experimental results and conclude the paper.

METHODOLOGY

Reinforcement learning

RL is a machine learning technique (Sutton & Barto, 1998), in which the learning agent interacts with its environment, and uses feedback from the environment to determine the best possible action given the current state (Figure 1a). RL is formalized as a Markov Decision Process (MDP): It is a tuple of states (S), actions (A), transition probabilities (P), and rewards (R: S×A→R). Each state has a value $V^{\pi}(s)$ which is the expected return for the agent when starting in that state and following the policy, where the policy $\pi : S \rightarrow A$ maps states to actions. The agent's objective is to calculate the optimal policy, i.e., the one that leads to maximum expected return (Nagy et al. 2018a). We used model based RL for LightLearn, where P and R are determined from experimental data, and the MDP is solved for the optimal policy using value iteration (Sutton & Barto, 1998).

RL based Occupant centered building control

The general RL-OCC framework acquires two types of data (Figure 1b): 1) Indoor environment, e.g., temperature, humidity, and luminance. 2) Interaction of the occupant with the building systems, which is directly related to occupant satisfaction, e.g., thermostat, blind, and switch usage. For the RL agent, type 1 data is used for states, and type 2 data is the feedback/reward signal. Ultimately, the RL agent determines the optimal action based on states and reward. This action is then used as e.g., set-point for the building control system.

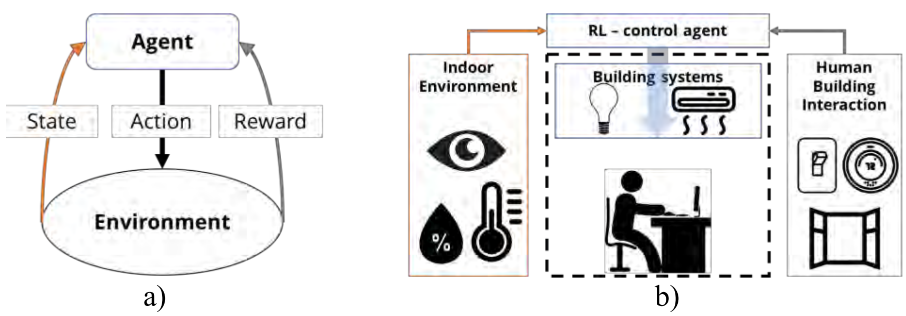


Figure 1. a) Reinforcement learning b) general concept of the RL-OCC framework

LightLearn: An RL-OCC Case Study

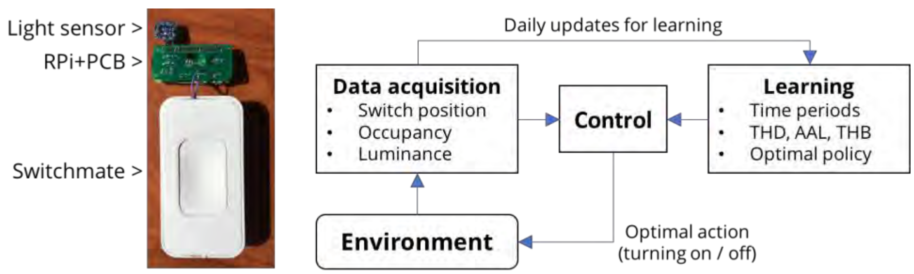
Hardware configuration

The hardware of LightLearn consists of low-cost off-the-shelf components to allow retrofitting any manual on/off light switch (Figure 2a). The controller itself is implemented in Python on a Raspberry Pi Zero W (RPi). The gathered environments are 1) luminance (TSL 2561). 2) A personal BT enabled device (mobile phone, watch, etc) paired with the RPi serves as proxy for occupancy (Nagy et al. 2018b). 3) For both sensing and actuating the light switch, we used a product called Switchmate. It uses a DC motor to move the light switch (on/off), while the occupant can switch the lights via a push button. We also designed a custom printed circuit board to interface and control the Switchmate with the RPi.

Table 1. Data acquisition summary

Variable	Type	Values	Unit	Logging Frequency
Switch position	binary	0/1	-	1 minute
Occupancy	binary	0/1	-	1 minute
Room light level	integer	0-10,000	lux	1 minute

The controller has the three components (Figure 2b): 1) Data acquisition of switch position, occupancy, and light level on the wall each minute (Table 1). 2) Learning process by building the MDP and determining the optimal policy once a day at midnight. (see next section) 3) Live control of the lighting system with the most recent learning results. Each minute, the controller determines the states and actuates the action when new state is different from previous state. A LightLearn was installed in a south facing, one person office, retrofitting the previously manual light switch, and an experiment was conducted for three weeks (3/19 – 4/5).



a) b)

Figure 2. a) LightLearn hardware b) Control algorithm flow chart

State-action Space Design and Learning process

The key when developing a robust RL controller is the design of the state-action space for the MDP. For LightLearn, the possible actions (A) are turning on/off. The states (S) are defined by four parameters: 1) Each day is split into four periods according to the trends of increasing (P1), midday (P2), decreasing (P3), and low ambient light (P4). These periods were identified using the gradient between two consecutive luminance data points. P4 is defined when the gradient is constant. If the majority (>80%) of the gradients are positive or negative, the time periods are set to P1 and P3, respectively. P2 is assigned in between P1 and P3. 2) Occupancy, and 3) switch position (on/off) are additional considerations for the states. 4) The indoor light levels that correspond to the cutoffs between Dark, Comfort, and Bright: if the occupant switches the lights on, we assume that he/she feels dark. Thus, the light level at that moment is the threshold for darkness (THD). If the light levels are below THD, it is considered Dark. Typically, the occupant can switch the light multiple times. Then we average the values to obtain THD. The artificial light level (AAL) is measured by the lux level difference between switch off & on. A last threshold value is the threshold for brightness (THB=THD+AAL), which indicates that it is bright enough for occupant. Light levels above THB are considered to be Bright. Light levels between THD and THB are regarded as Comfort.

Figure 3 shows all possible states for the control agent. For example, when the room is occupied during P1, P2, and P3, the agent can be in the Bright or Comfort state if the lights are on or off, and in the Dark state when the lights are off. In brief, each day can be split into, e.g., at most 26 states, and the objective of the agent is to determine the best action for each of those states. Once states and actions are defined, the probabilities of transitioning from one state to any other state are calculated by historical data. For each state, all the state transitions are counted and divided by the total number of state transitions. Finally, we use the reward values shown in Figure 3. We set negative rewards (-1) for energy wasting, positive rewards (+1) for using energy for occupant comfort, and larger rewards (+2) for energy savings.

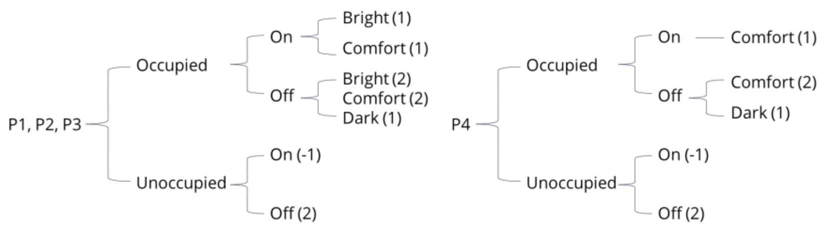
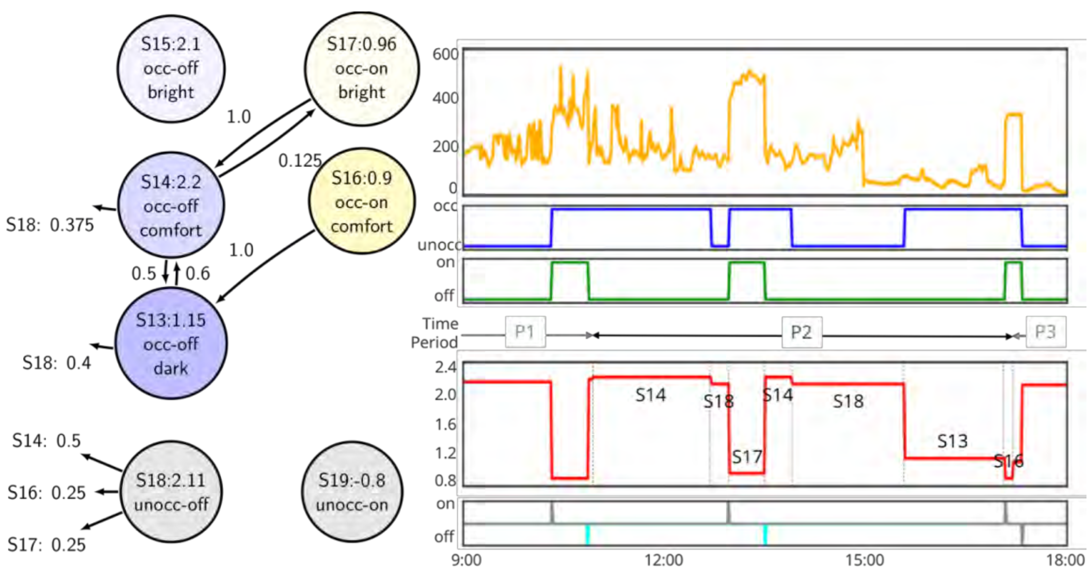


Figure 3. State-Action definition by time periods, occupancy, switch position, and light level; Rewards on each state are annotated in parenthesis.

RESULTS

To illustrate how the controller worked, we show the MDP and the data collection result on March 27th (Figure 4). Historic data (3/19-26) was used to build the MDP (Figure 4a). The controller solved it for the optimal policy for the following day. Specifically, there are 7 states (circles), the transition probabilities are calculated between two states (arrows), and the calculated optimal policy (values) is mapped on each state.

With this optimal policy, the controller actuated the switch to maximize the value based on its current state. For example, the controller determined its state based on the three sensor values (i.e., lux level, occupancy, switch position) in Figure 4b. Based on state transition, the associated values were changed. On March 27th, there were the two control actions: 1) The controller read the current state as S17 (occ-on-bright) and switched off to transition to the state with maximum value (S14, occ-off-comfort) around 13:30. 2) Similarly, the controller read S9 (occ-on-bright) and turned off, transitioning into state S7 (occ-off-comfort) at 10:51 in P1. We emphasize that both actions were taken because the controller transitioned into "Comfort" states.



a) b)

Figure 4. March 27th: a) MDP during P2 b) Data: lux level (yellow), occupancy (blue), switch position (green), value with states (red), and control actions (human-gray, controller-cyan)

The controller calculates the learning outputs every midnight. With these daily outputs, we investigate the evolutions of control parameters: Figure 5a shows how the time period varied by day: the start times of P1 and P4 were relatively constant, because sunrise and sunset time did not significantly change during the experiment. The duration of P2 increased while the durations of P1 and P3 decreased. Because there were no clear patterns for both daylight increasing and decreasing due to the weather condition (cloudy). In figure 5b, we also observe that the THD is around 280 lux until the 7th day. Then, it drops to around 190 lux. This is because the occupant turned on the switch with lower ambient light levels on the cloudy days (7th–10th). This shows how LightLearn adapts its set-points based on occupant interaction.

Similarly, the evolution of the values for states S6, S9, and S10 in Figure 5c demonstrate the adaptation capabilities of LightLearn. Each state has same value (1) on the first day, yet the values of S9 is the lowest until the 8th day, because there was no state transition into S9. However, after the 7th day, the occupant turned on the light often due to cloudy weather, which increased the transition probabilities to the S9 state.

To evaluate energy savings of LightLearn, we calculated the total light-on time and compared with other scenarios (Table 2). Essentially, the operation time serves as proxy for energy consumption. The full operation is the worst-case scenario, assuming the light is on during the normal office schedule (9am-5pm) on weekdays. In addition, the occupancy-based operation was compared. In this case, we assumed that the occupant turns on the light whenever she/he occupies the room. Compared to the full operation and the occupancy-based case, LightLearn reduced the operation time by 83% and 63%, respectively. In brief, the controller successfully saved energy consumption by learning the occupant behavior and the room environment.

Table 2. Operation time (light-on) of each scenario

Scenario	**Full operation (9am-5pm)**	**Occupancy-based**	**LightLearn**
Operation time (hrs)	104	48.8	17.6

DISCUSSIONS

The learning process can be tuned by varying the RL parameters. First, the controller determines the importance of future reward by discount factor (γ). Typically, lighting control requires a very small value ($\cong \gamma$ 0): If excess weight is given to future energy saving, the controller might keep the switch off in a morning and expect to have sufficient daylight level later. This would cause discomfort situation for the occupant. However, in other types of building control problems, i.e., HVAC, higher discount factors might be required (Vazquez-Canteli et al. 2017). Also, the reward assignment can initialize the objective of controller. For the sake of energy saving, we assigned higher rewards (r=2) for the off states (Figure 3). Thus, those states have relatively higher values, and lead the controller to primarily switch off.

CONCLUSION

This paper introduced RL based OCC framework. As a case study, we implemented LightLearn in an office. LightLearn was successfully deployed and collected occupant related data. Ultimately, the controller generated the MDP and calculated the optimal policy every midnight for the operation on the following day. The results indicate that the controller reduced unnecessary operation time by learning unique characteristics of the occupant behavior and the room environment. For future studies, RL-OCC approach should be further evaluated by different building systems (i.e., dimmable lighting, HVAC), environments (i.e., visual and thermal), more occupants, and for longer periods.

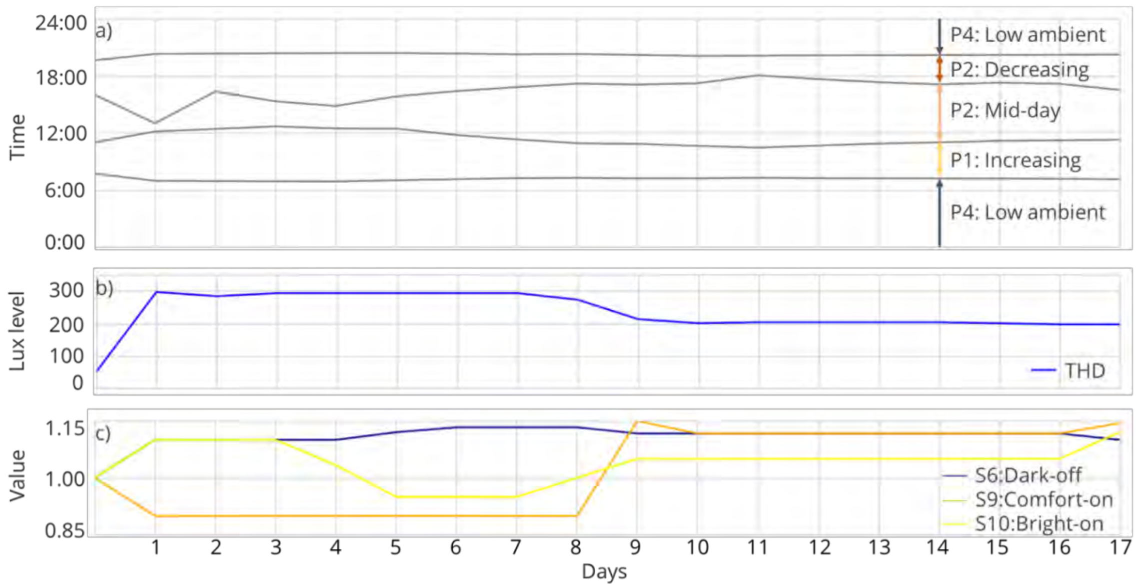


Figure 5. Evolutions of a) time period b) threshold darkness level, and c) values of S6,S9,S10

ACKNOWLEDGEMENT

This research is supported by UT Austin through a Green Fee Award.

REFERENCES

DoE, 2011. Buildings energy databook. *Energy Efficiency & Renewable Energy Department*.

Gunay H.B., O'Brien W. and Beausoleil-Morrison I. 2013. A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices. *Building and Environment*. Vol. *70*, pp. 31-47.

Loftness V., Hartkopf V., Gurtekin B., Hansen D., and Hitchcock R. 2003. Linking energy to health and productivity in the built environment. In: *Proceedings of the 2003 Greenbuild Conference*.

Nagy Z, Park J.Y., and Vazquez-Canteli J. 2018 Reinforcement learning for intelligent environments: A tutorial. in press.

Nagy Z, Vazquez-Canteli J., and Park J.Y. 2018. Using Bluetooth Based Occupancy Estimation for HVAC Set-Back to Reduce Energy Consumption in Buildings. In: *Proceedings of the 2018 ASHRAE Annual Conference.*

Nagy Z., Yong F. Y., Frei M., and Schlueter A. 2015. Occupant centered lighting control for comfort and energy efficient building operation. *Energy and Buildings*, Vol. 94, pp. 100-108.

Nagy Z., Yong F. Y., and Schlueter A. 2016. Occupant centered lighting control: a user study on balancing comfort, acceptance, and energy consumption. *Energy and Buildings*, Vol. 126, pp. 310-322.

Park J.Y. and Nagy Z. 2018. Comprehensive analysis of the relationship between thermal comfort and building control research-A data-driven literature review. *Renewable and Sustainable Energy Reviews*. Vol. 82, pp. 2664-2679.

Sutton R. S., and Barto A. G. 1998. *Reinforcement learning: An introduction* (Vol. 1, No. 1). Cambridge: MIT press.

Vazquez-Canteli J., Kampf J., and Nagy Z. 2017. Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted Q-iteration. *Energy Procedia*. Vol. 122, pp. 415-420.