13th IEA Heat Pump Conference
April 26-29, 2021 Jeju, Korea

Energy saving pre-cooling pattern search of an inverter air conditioner using a deep reinforcement learning algorithm

Myung Sup YOON[a,*], Won Sik YOON[a]

[a]Energy Technology Center, Korea Testing Laboratory, 87 Digital-ro 26-gil, Guro-gu, Seoul, 08389, Republic of Korea

Abstract

In this study, we experiment with the energy saving pre-cooling operation of an inverter air conditioner using a deep reinforcement learning algorithm under a fixed going out time duration. An air conditioner with a rated cooling capacity of 6500 W is used. An indoor building load of 4500 W sensible heat and an outdoor temperature of 35°C are continuously applied throughout the study. During the going out time, three remote control temperature options (21, 26, 31°C) are given for simplicity, and a remote controlling pattern (set temperature sequences) that minimizes power consumption upon arrival home while satisfying the arrival target temperature condition is searched using deep reinforcement learning. For a given going out time duration, the power consumptions of the temperature operation pattern found by the computer deep reinforcement learning algorithm and a human are compared. Although the artificial intelligence searches for the power consumption operation pattern over a long period (7 ~ 9 days), depending on the learning environment, it identifies a more accurate and more energy saving and an unexpected operation pattern than by human analysis.

Keywords: Air conditioner; Deep reinforcement learning; Pre-cooling; AI; Energy saving; Smart HVAC; Set temperature sequences

1. Introduction

Due to the development of artificial intelligence (AI) technology, many of us will soon live in more comfortable and energy efficient smart homes, with pre-cooling/pre-heating before returning home that incorporates data from smart vehicles. With the development of high efficiency inverter air conditioners at part-load condition, more energy can be saved by operating an air conditioner gradually before arrival than to operate an air conditioner suddenly on arrival if certain conditions are met [1]. Studies on building energy prediction, management, and heating, ventilation, and air conditioning (HVAC) system control using deep learning technology have been actively conducted [2–20]. However, research that applies deep learning technology to household air conditioners is scarce. Some AI platform companies such as Google Nest and global electronics companies are developing deep learning technology for household air conditioning, but they tend not to publicly release their research findings. Precision temperature control for an air conditioner and optimal operation point search of air conditioner components for maximum energy efficiency (EER) using a deep reinforcement learning algorithm have been conducted in previous research [21–22]. However, there is no air conditioner deep learning research for energy saving pre-cooling operation during the time at which all household occupants are away from home.

Figure 1 shows a conceptual diagram of a comfortable and energy saving HVAC pre-cooling or pre-heating system in conjunction with location information of occupants. The arrival time varies depending on the means of transportation and traffic conditions of the day. Pre-cooling or pre-heating should also vary according to the arrival time of family members, outdoor atmospheric conditions, and the internal building loads of the day. If

* Corresponding author.
E-mail address: msyun95@ktl.re.kr

sufficient data exist for these variables, or a well-defined environment exists, it seems that optimal operation is possible using a deep learning algorithm.

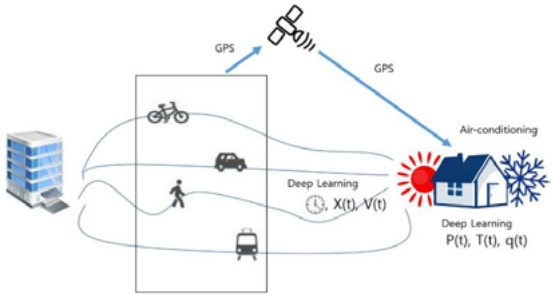


Fig. 1. Concept of pre-cooling or pre-heating using deep learning

2. Experimental Approach

2.1. Problem setup

In this study, in order to minimize the AI learning time, a deep reinforcement learning algorithm is applied under certain assumptions in a well-defined test chamber environment [23–24]. First, cases where the fixed going out time is 2.5 and 3.5 hours are calculated. Second, the building load (q_{build}) during non-occupancy is assumed to be constant at 4500 W of constant sensible heat, and the outdoor side temperature is assumed to be constant at 35°C. Third, it is assumed that there is no latent heat consideration. Fourth, in order to minimize learning obstacles in the dead band zone, which is a feature of the product control, only three signals of set temperature (21, 26, 31°C) are considered. Fifth, the initial temperature (T_0) of the indoor room temperature (T_{in}) in most of the learning episodes is near 26°C. Sixth, temperature variations are considered only when a single agent (actor) exists as one resident (Fig. 2(a)). Seventh, linear distance movement is assumed in all episodes during learning (Fig. 2(b)).

Figure 2(a) shows the temperature change of a room (dashed lines) over the going out time duration ($t_2 - t_1$). Three set temperatures (T_{set}) of the remote controller (21, 26, 31°C) can be given in various sequences over the going out duration at about every 2 minutes. If T_{set} is fixed as 26°C throughout the problem, then the indoor room temperature will ideally be horizontal. To simplify the problem, two cases of the going out time duration (2.5 h and 3.5 h) are scrutinized. Figure 2(b) shows the remaining distance to the destination over time. Linear approximation for the simplicity of the problem is assumed regardless of the various paths according to the traffic and traffic conditions (Fig. 1). A non-linear remaining distance (Ls(t)) pattern also can be considered if there is time-dependent GPS information of multiple agents (family members). Here, only linear profiles are assumed, and $t_{arrival(1)}$ corresponds to 2.5 h and $t_{arrival(2)}$ to 3.5 h, respectively.

The purpose of this study is to find T_{set} sequences over the going out time duration to minimize energy consumption (kWh) using a deep reinforcement algorithm. Of course, T_{in} should be in the target temperature (24 ~ 28°C) on the arrival time (t_2), regardless of the temperature variations during the going out time.

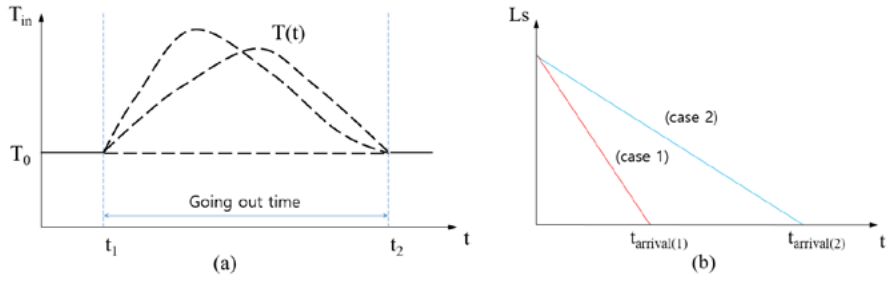


Fig. 2. Variations of (a) indoor temperature (T_{in}) and (b) remaining distance (Ls)

2.2. Air conditioner selection

Air conditioners differ in precision control from model to model and from brand to brand [1]. Minimized dead band zone and precision controlled products are good for reinforcement learning (RL), but unfortunately, the latest models are locked from obtaining IR signals for controlling remote signals using Arduino [21]. Therefore, the old inverter products of Fig. 3 and Table 1 are used throughout this study.

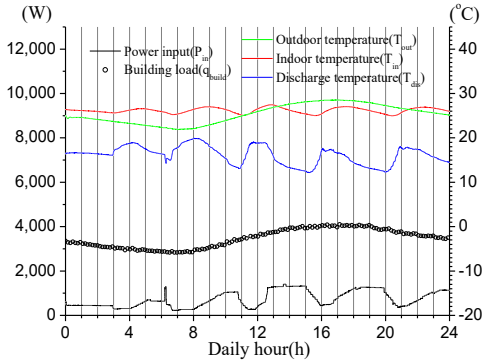


Fig. 3. Operating characteristics of the air conditioner under test (reproduced from Yoon et al. [1, 21])

Table 1. Specifications of test air-conditioner (reproduced from Yoon et al. [21])

Contents	Name plate values
Air-conditioner type	Split wall mount
Compressor type	Inverter type
Electricity	Single phase, 220 V, 60 Hz
Rated cooling capacity	6500 W
Rated power input	1750 W
Max. power input	2900 W
Min. power input	500 W
Refrigerant	R410a, 2 kg

3. Methodology

3.1. Artificial neural network (ANN) model

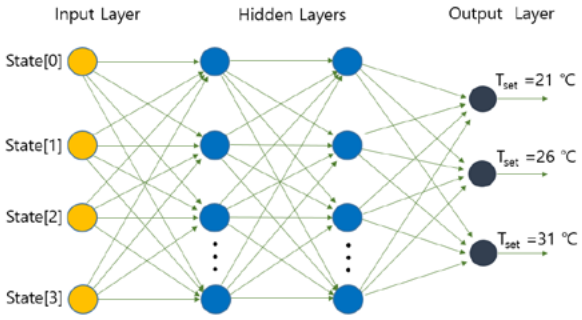


Fig. 4. ANN model for the inverter air conditioner pre-cooling

For the deep reinforcement learning (RL) algorithm, we use one input layer of four nodes, two hidden layers of 24 nodes each, and one output layer of three nodes. The input layer consists of four elements of state [(T_{in}(t) –26), dT_{in}/dt, Pws(t), t]. Each element represents indoor temperature (T_{in}), indoor temperature derivative (dT_{in}/dt), and cumulative power consumption (Pws(t)) at time t, and time t. Here, Pws is same as the Pwh divided by 3600 seconds (Pws = Pwh / 3600). If the movement change due to traffic is non-linear, the above state variable t can be changed into the distance state variable of Ls(t), or it can be replaced with x(t) and y(t), that is, GPS coordination information.

3.2. Reward strategy

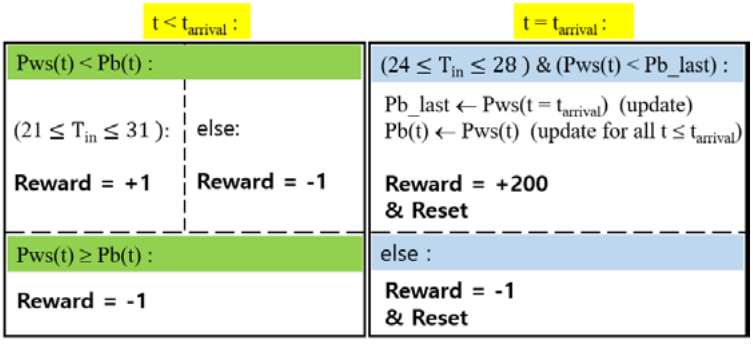


Fig. 5. Regions for the rewards

Figure 5 shows the reward strategy for this problem set-up. Pws(t) is the cumulative power consumption [Ws] measured at each time step, Pb(t) is the cumulative power consumption as a baseline at time t, and Pb_last is the baseline cumulative power consumption at the last time step. If the indoor room temperature (T_{in}) satisfies the target temperature (24–28°C) at the arrival and the Pb($t_{arrival}$) records the minimum power consumption, then a big reward (+200) is given and Pb(t) and Pb_last are updated as new data in this episode, that is, baseline information is updated as the minimum power consumption case. In the first episode (episode = 0), Pb_last starts with 1.e+25, and Pb(t) starts with null values. Before arrival ($t<t_{arrival}$), the target temperature is looser as (21–31°C) under the Pws(t) < Pb(t) condition at each time step, and reward is 1 if satisfied. The reward +1 region induces an air conditioner operation pattern that consumes less power than the previous episode in each time step and consequently induces the reduced cumulative power consumption at the last time of arrival. In other regions, except for the reward +200 and +1 regions, –1 reward is given. Every time, steps have real time progresses according to the linear movement of Fig. 2(b).

3.3. Deep reinforcement learning algorithm coding

The problem definition is:

- State: [(T_{in}–26), dT_{in}/dt, Pws(t), t]
- Action: [21°C, 26°C, 31°C]
- Reward: +200 pt, when T_{in} in the range of (26 ± 2.0)°C and Pws is reduced than ever existing minimum power consumption value (Pb_last) at the arrival time
- Reward: +1 pt, when T_{in} in the range of (26 ± 5.0)°C and Pws(t) is reduced than Pb(t)
- Penalty: –1 pt, in other regions except positive reward regions
- Each episode ending and reset: when t = $t_{arrival}$

The coding flow is shown in Fig. 6. The basic algorithm is the same as established deep Q-learning, with experience replay [25], but it differs by the addition of the reward and reset condition according to Figs. 5 and 6. In Fig. 6, time progresses as real time (t = t + dt) until t = $t_{arrival}$. The initializing condition of T_{in} = (26 ± 0.05)°C is specially applied only to Section 4.4 to compare with the Section 4.3 case of non-exact temperature initializing (fixed time initializing case). The simulation time end is judged by human sense when it seems impossible to achieve more energy saving even with enough repetitive episode training.

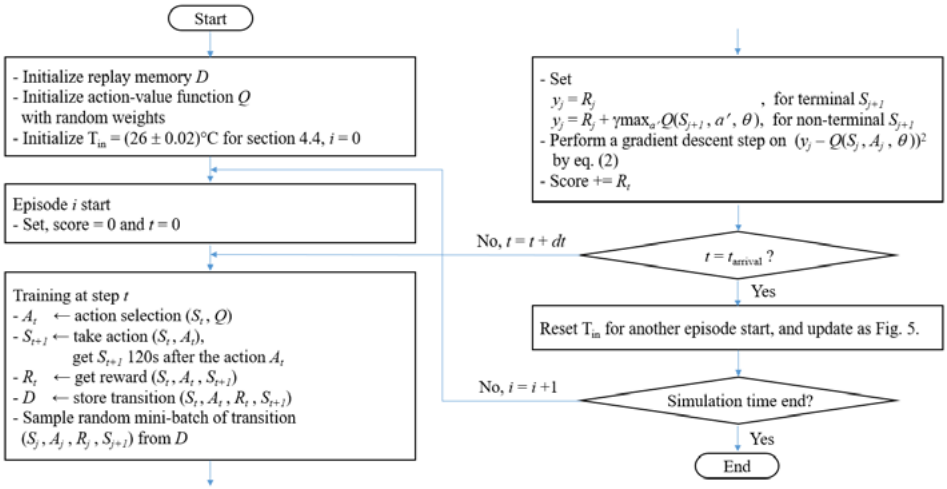


Fig. 6. Deep reinforcement learning coding flow

4. Experiments and results

4.1. Without machine learning (human operation)

Before operation using RL, the amount of minimum power consumption is checked during 2.5 hours and 3.5 hours of going out time by a human. Half an hour is initialized to a set temperature of 26°C before going out at 2.5 hours. Each case of Figs. 7(a), (b), and (c) shows a slightly different T_{in} initializing graph even though the T_{set} temperatures are the same at 26°C. This is generally due to the rough control behavior of air conditioning products, including dead band control. Table 2 summarizes the manual control method. In Fig. 7(b), in order to bring the temperature at t = 3.0 to the target temperature of exactly 28°C, the target range is (24–28°C), T_{set} = 31°C is extended to t = 1.83 h as opposed to 1.33 h in the case of Fig. 7(a). As a result, in Fig. 7(b), the power consumption is reduced to Pwh = 2930 Wh rather than Pwh = 3441 Wh in Fig. 7(a). In Fig. 7(c), instead of lowering the temperature by T_{set} = 26, it is lowered to T_{set} = 21°C. But there is no difference in how the product actually cools due to the (T_{set} –T_{in}) difference detection control.

Figure 8 shows the manual control pattern when the going out time is 3.5 hours. An hour is initialized to a set temperature of 26°C before going out. In Fig. 8(c), T_{set} = 26°C is continuously maintained and changes to 31°C at t = 4.1 h, so that the temperature is near T_{in} = 28°C at the arrival point. Figure 8(c) shows a relatively large electric consumption of 4761 Wh due to poor air conditioner operation.

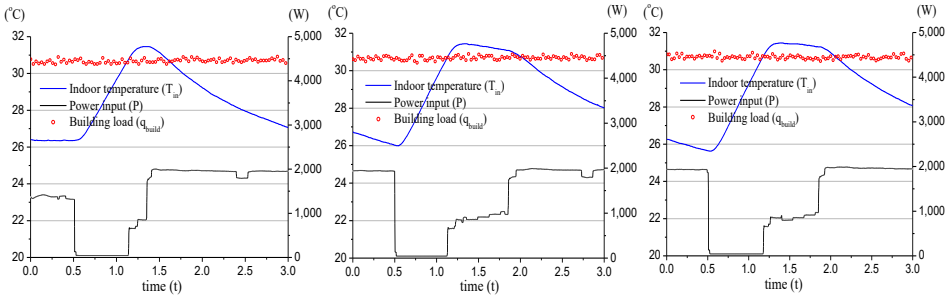


Fig. 7. Human operation during 2.5 h (power accumulation starting at t = 0.5 h): (a) Pwh = 3441, (b) Pwh = 2930, (c) Pwh = 2951

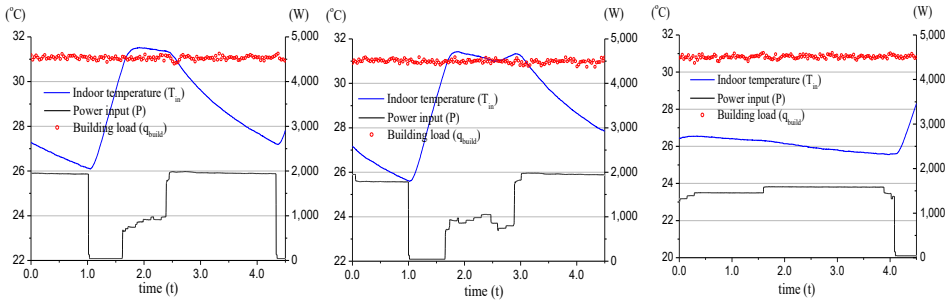


Fig. 8. Human operation during 3.5 h (power accumulation starting at t = 1.0 h): (a) Pwh = 4520, (b) Pwh = 4318, (c) Pwh = 4761

Table 2. Human operation case results

Going out time	Operation case	Operation method	Pwh during going out
2.5 h (Fig. 7)	Fig. 7(a)	26°C(t=0), 31°C(t=0.5h), 26°C(t=1.33h)	3441 Wh
	Fig. 7(b)	26°C(t=0), 31°C(t=0.5h), 26°C(t=1.83h)	2930 Wh
	Fig. 7(c)	26°C(t=0), 31°C(t=0.5h), 21°C(t=1.83h)	2951 Wh
3.5 h (Fig. 8)	Fig. 8(a)	26°C(t=0), 31°C(t=1.0h), 26°C(t=2.5h) , 31°C(t=4.3h)	4520 Wh
	Fig. 8(b)	26°C(t=0), 31°C(t=1.0h), 26°C(t=3.0h)	4318 Wh (4437 Wh)
	Fig. 8(c)	26°C(t=0), 31°C(t=4.1h)	4761 Wh

In the case of Fig. 8(b), due to the manual operation error of the initial temperature adjustment, the initial temperature at the beginning of the power measurement (t =1.0) starts at 25.6°C, which is advantageous for calculating the accumulated power consumption. The time of 430 seconds, which is 26.0°C after t = 1.0, is compensated for the power consumption under the part-load operation assumption (see the power consumption around t = 2.0 in Fig. 8(b)) of 1 kW (430 × 1000 / 3600 = 119 Wh). The 4437 (= 4318 + 119) Wh in the parentheses is a reasonably minimally compensated value for the human operation error.

4.2. Deep reinforcement learning (2.5 hours going out)

The air conditioner operation of 2.5 hours' going out time is solved by RL. It is operated throughout 7.5 days of 59 episodes (Figs. 9 and 10). When we see the entire results graphs (Fig. 9) of 7.5 days, it is difficult to see what the meaningful results are. Therefore, the scores and the cumulative power consumption at each episode are taken by a quick glance from Fig. 10. The red dots in Fig. 10 refer to the results where the episode acquired a +200 reward at the arrival time. The red dots of power consumption decrease until episode 42 and no more episode cases, which gains a +200 reward, appear until the last episode (episode 59).

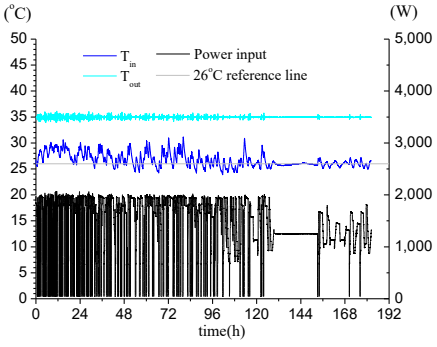


Fig. 9. Total of 7.5 days' deep reinforcement learning for the 2.5 h going out problem

Among the red dot episodes, a representative of three episodes cases (episodes 6, 34, 42) are carefully analyzed (Fig. 11). Commonly, an episode is forcibly stabilized by the 26°C remote control setting during the initial 0.5 hours. However, in the case of Fig. 11(a), the initial value of 26°C is not stably reflected because 0.5 hours are insufficient to stabilize in some cases. In Fig. 12(a), the action outputs for the 21, 26, and 31°C settings are even. In the case of episode 34 (Fig. 12(b)), however, most of the action output maintains the set temperature of 21°C or 26°C, with the exception of the 31°C action output at the 51st time step. In episode 42 (Fig. 12(c)), only the remote control signal of 26°C is continuously maintained. This means that it is more effective to keep the air conditioner at the same initial temperature for a relatively short going out time, which is 2.5 hours in this case. When we compare the human operation case of Table 2, episode 42 of deep reinforcement learning finds smaller power consumption (Pwh = 2766 Wh) than the human (Pwh = 2930 Wh). A human is prone to missing the idea of maintaining the same temperature, but AI finds the answer exactly.

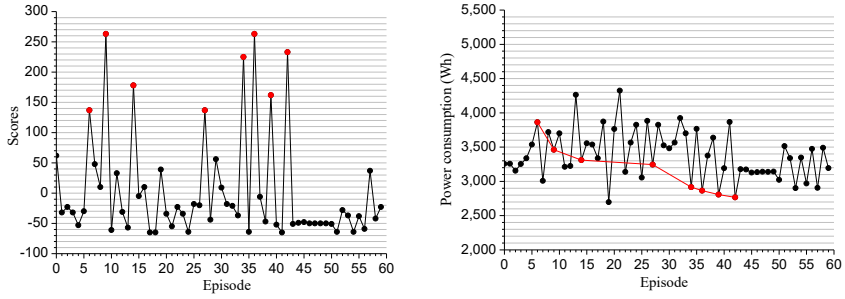


Fig. 10. Variations of (a) scores and (b) power consumption at each episode for the 2.5 h going out problem

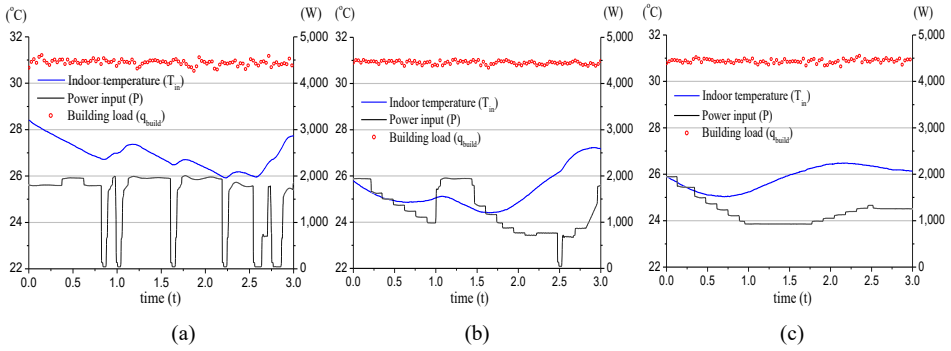


Fig. 11. The RL operation during 2.5 h (power accumulation starting at t = 0.5 h): (a) episode 6, (b) episode 34, (c) episode 42

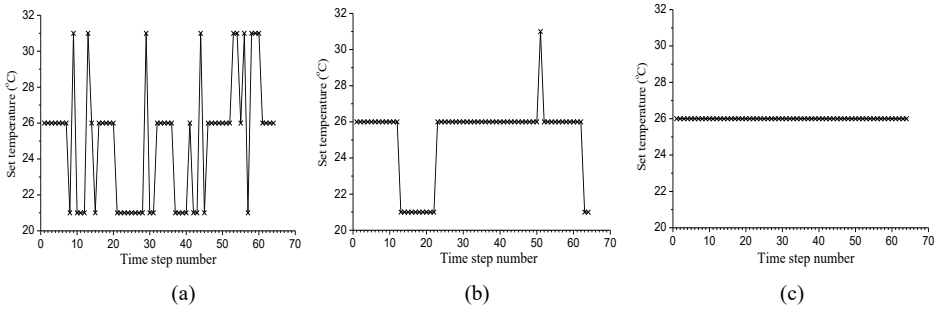


Fig. 12. Set temperature sequences for: (a) episode 6, (b) episode 34, (c) episode 42

4.3. Deep reinforcement learning (3.5 hours going out)

The air conditioner operation of 3.5 hours' going out time is solved by RL. It is operated throughout nine days of 47 episodes (average 0.19 days per one episode, that is, one hour initializing and 3.5 hours training). Because the going out time is longer than the 2.5 hours' going out case, it takes more training time (nine days) than Section 4.2 described (7.5 days of 59 episodes, average 0.127 day per one episode, that is, 30 minutes' initializing and 2.5 hours' training).

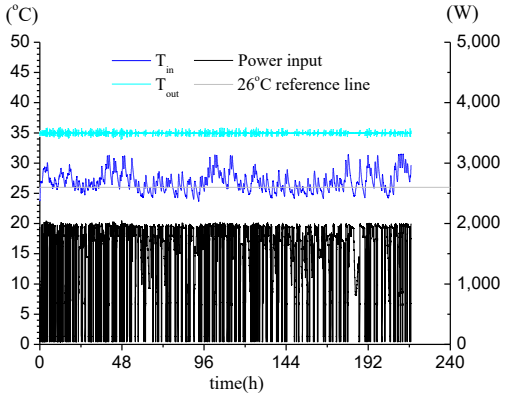


Fig. 13. Total of nine days' deep reinforcement learning for the 3.5 h going out problem

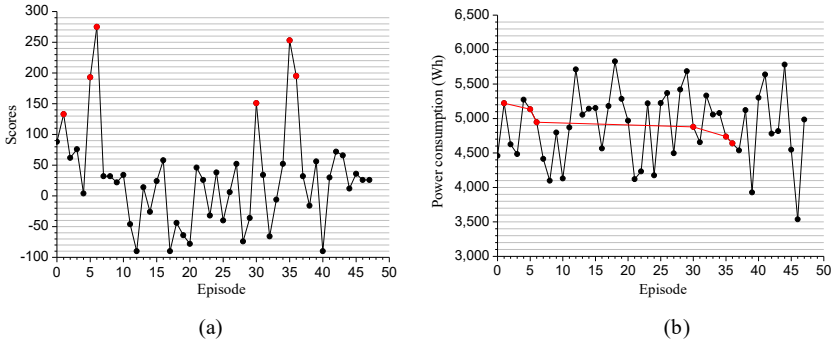


Fig. 14. Variations of (a) scores and (b) power consumption at each episode for the 3.5 h going out problem

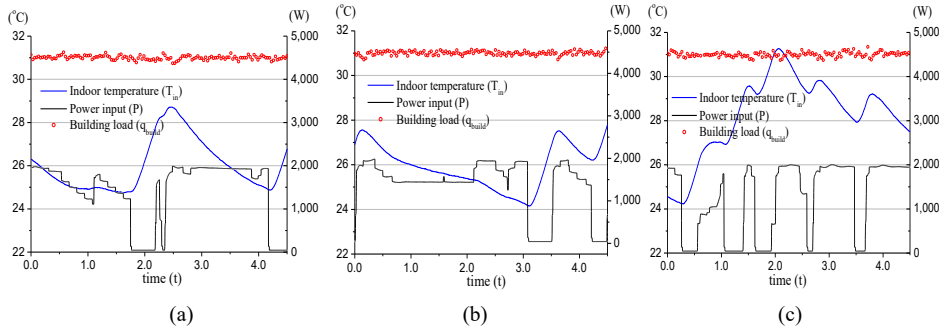


Fig. 15. The RL operation during 3.5 h (power accumulation starting at t = 1.0 h): (a) episode 35, (b) episode 36, (c) episode 42

The RL finds red dots, which satisfies minimized power consumption and in the range of target temperature on the arrival, at episodes 1, 5, 6, 30, 35, and 36 until episode 47. Figures 15 and 16 show the power consumption and remote control signal analysis at episodes 35 and 36 of the energy saving red spot and at episode 42, where there is no energy saving effect. Episodes 35, 36, and 42 record the power consumption of 4739, 4639, and 4780 Wh, respectively. Compared with the results of human analysis in Fig. 8 and Table 2, the cases of human operation (4437 or 4520 Wh) save more energy than the result of RL (4639 Wh). Considering the time it took (9 days), the RL results look comparatively poor.

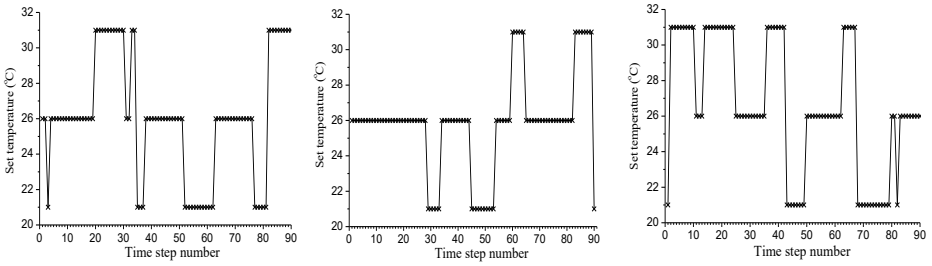


Fig. 16. Set temperature sequences for: (a) episode 6, (b) episode 34, (c) episode 42 (no red dot position of Fig. 14)

4.4. Deep reinforcement learning (3.5 hours going out) with accurate initial condition (T_0 =26 ℃)

The former RL work of Section 4.3 (3.5 hours going out) does not fare well compared with the RL case of Section 4.2 when the going out time is 2.5 hours. The reason for the difference can be guessed by comparing Figs. 11 and 15. In Fig. 11, as the episode progresses, it converges to the 26°C remote control setting value at the arrival time, and in the next episode the learning starts stably at the initial temperature around 26°C. However, in Figs. 15(b) and (c), learning starts slightly away from the initial value of 26°C. The importance of a consistent initial value for RL is well known in the grid world problem or in Atari games [25–26].

With this in mind, 3.5 hours' going out problem is re-solved by modifying the program to start the RL in each episode when the indoor room temperature (T_{in}) is within (26 ± 0.05)°C. If the initial value is under 26°C, then the 31°C remote signal is sent. In the opposite case, a 21°C remote signal is transmitted before starting each episode of learning. The RL is operated fully throughout seven days of 39 episodes, as in Fig. 17.

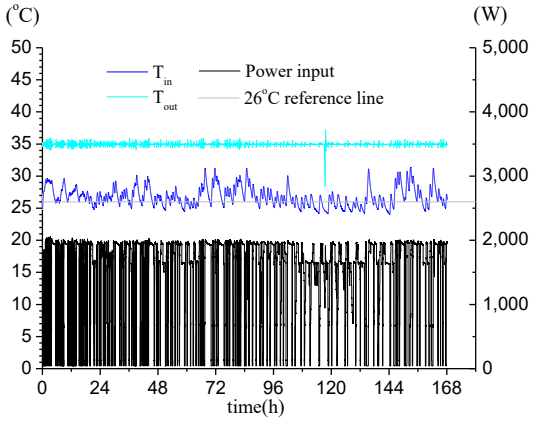


Fig. 17. Total of seven days' deep reinforcement learning for the 3.5 h going out problem

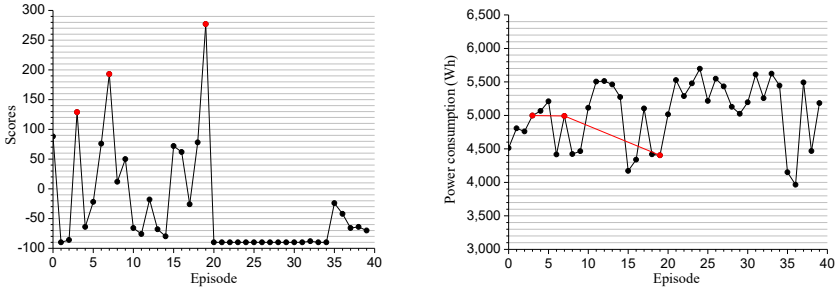


Fig. 18. Variations of (a) scores and (b) power consumption at each episode for the 3.5 h going out problem with accurate initial condition of T_{in} = 26°C

Although learning is performed for seven days, the meaningful RL practically ends within 3.6 days of 19 episodes, as in Fig. 18. At episodes 3, 7, and 19, a reward of +200 (Fig. 5) is recorded, and the power consumption is 4998, 4991, and 4405 Wh, respectively. When compared with the former case of Fig. 14, Fig. 18 converges fast (episode 19) and finds less power consumption (4405 Wh) than the Fig. 14 case (episode 36, 4639 Wh). In Fig. 18(a), it is not easy to get even a smaller reward of +1 (Fig. 5) after episode 19, since the results of episode 19 seem to be optimal.

Compared with the manual result of 4437 Wh in Table 2 and Fig. 8(b), the RL result of 4405 Wh in episode 19 saves more energy. What is more noticeable is that unlike the common sense used in the air conditioner operation method in the operation case of Fig. 8(b) in Table 2 (minimize operation of the air conditioner when going out and then drive again before arrival), RL finds an energy saving method that cannot be found by a human (Fig. 19(c)). Conversely, by comparing the results of Fig. 19(c) and Fig. 8(b), the strategy of RL is analyzed from a human point of view. In Fig. 8(b), after the room temperature reaches 31°C or higher at t = 1.7 h, the air conditioner runs with a part-load of 1 kW to satisfy the 31°C set temperature. After t = 3.0 h, in order to satisfy the final target temperature 28°C (the target range is 24–28°C), the temperature starts to drop by full-load operation. On the other hand, in Fig. 19(c), since the temperature must drop after reaching 31°C anyway in some moment, the temperature drops immediately without delay at the point of 31°C. In addition, energy saving is achieved by aiming for the effect of turning off the air conditioner in (1.8 ~ 2.0) hours and (3.25 ~ 3.5) hours.

In summary, this kind of RL approach is surely useful to find an energy saving operation method when we run an air conditioner. However, we also need more studies about when such an algorithm should be stopped, because in this problem there is no determined answer like the grid world problem or the precision temperature control problem. The grid world problem [26] has a clear goal of reaching the target position, and the air conditioner precision temperature control problem [21] also has an obvious target range of temperature. However, this study has no deterministic minimum energy consumption value. If we have found an energy saving operation method such as in this study, we need to think carefully about how to verify that this minimum solution is a real optimum.

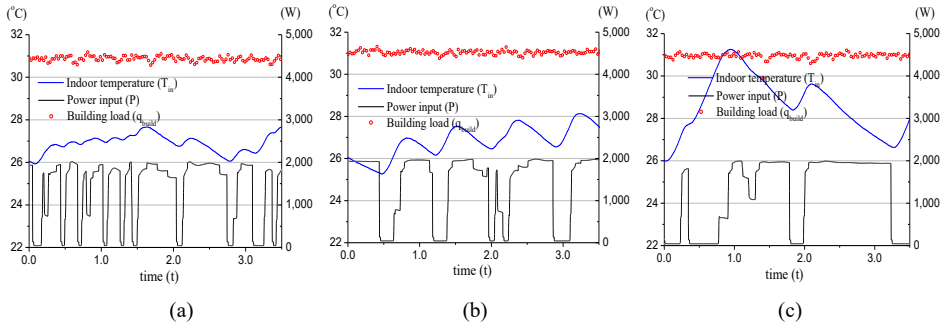


Fig. 19. The RL operation during 3.5 h (power accumulation starting at t = 0 h): (a) episode 3, (b) episode 7, (c) episode 19

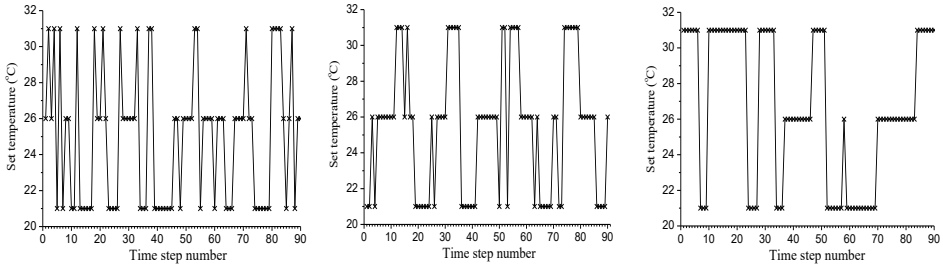


Fig. 20. Set temperature sequences for: (a) episode 3, (b) episode 7, (c) episode 19

5. Conclusion

Reinforcement learning (RL) using a DQN (Deep Q-learning Network) method is applied to a household inverter air conditioner energy saving operation pattern search while occupants are out. Four state variables of [(T_{in}(t) −26), dT_{in}/dt, Pws(t), t] and three actions of 21, 26, 31°C remote control are used in the DQN model. To simplify the problem, some assumptions are made. The indoor building load (q_{build}) and outdoor temperature (T_{out}) are set to be constant at 4500 W and 35°C. Then, assuming that the moving path of the mover (person going out) is linear, only the going out time ($t_{arrival}$) is involved in the problem. For the two going out times of 2.5 and 3.5 hours, the case of manual operation by a human and the case of artificial intelligence operation using RL are compared. The objective of this problem setup is to find remote controller signal sequences during the going out time that save electricity and satisfy the target temperature (24 ~ 28°C) on arrival. In applying RL, it is important to maintain a fixed initial room temperature (T_0 = 26° C) at each episode for fast and stabilized RL results. For the 3.5 hours' going out time problem, two cases of initial condition treatments are given: i) fixed stabilization time as one hour regardless of the start temperature, and ii) accurate initial temperature condition. The latter identifies the energy saving pattern twice as fast as the former, and the result is also better (less energy consumption).

Compared to manual operation based on human judgement, RL finds an exact method that is easy for humans to miss. When the going out time is 2.5 hours, RL derives continuous operation of the air conditioner by fixing the remote controller temperature as 26°C. And when going out time is 3.5 hours, RL shows an operation strategy that induces longer periods of off-mode or part-load mode. In the future, this work can be expanded by merging non-linear movement information with multi-agents and more realistic environmental conditions.

Acknowledgement

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) (P0014268, Smart HVAC demonstration support).

References

[1] M. S. Yoon, and W. S. Yoon, "Study on the smart operation approach of an air conditioner when occupants go out," in Proceedings of the 13th IEA Heat Pump Conference (HPC 2020), 2021.

[2] J. V. Canteli, S. Ulyanin, J. Kämpf, and Z. Nagy, "Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities," vol. 45, Sustainable Cities and Society, 2019, pp. 243–257.

[3] K. Grolinger, A. L'Heureux, M. Capretz, and L. Seewald, "Energy forecasting for event venues: Big data and prediction accuracy," vol. 112, Energy and Buildings, 2016, pp. 222–233.

[4] Z. Zhang, A. Chong, Y. Pan, C. Zhang, S. Lu, and K. P. Lam, "A deep reinforcement learning approach to using whole building energy model for HVAC optimal control," in Proceedings of the ASHRAE and IBPSA-USA SimBuild Building Performance Modeling Conference, 2018.

[5] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun. L. Zhang, Y. Zhang, and T. Jiang, "Deep Reinforcement learning for smart home energy management," vol. 7, IEEE IoT Journal, 2020, pp. 2571–2762.

[6] T. Liu, C. Xu, Y. Guo, and H. Chen, "A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction," vol. 107, Int. J. Refrigeration, 2019, pp. 39–51.

[7] D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," in Proceedings of the IEEE Industrial Electronics Society (IECON), 2016, pp. 7046–7051.

[8] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," vol. 212, Applied Energy, 2018, pp. 372–385.

[9] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," vol 33, Renewable and Sustainable Energy Reviews, 2014, pp. 102–109.

[10] M. Manivannan, B. Najafi, and F. Rinaldi, "Machine learning-based short-term prediction of air-conditioning load through smart meter analytics," vol. 10, Energies, 2017, pp. 1905.

[11] S. K. Jeong, C. H. Han, L. Hua, and W. K. Wibowo, "Systematic design of membership functions for fuzzy logic control of variable speed refrigeration system," vol. 142, Applied Thermal Engineering, 2018, pp. 303–310.

[12] H. Yan, Y. Xia, X. Xu, and S. Deng, "Inherent operational characteristics aided fuzzy logic controller for a variable speed direct expansion air conditioning system for simultaneous indoor air temperature and humidity control," vol. 158, Energy and Buildings, 2018, pp. 558–568.

[13] Z. Li, X. Xu, S. Deng, and D. Pan, "A novel neural network aided fuzzy logic controller for a variable speed (VS) direct expansion (DX) air conditioning (A/C) system," vol. 78, Applied Thermal Engineering, 2015, pp. 9–23.

[14] X. Xu, Z. Zhong, S. Deng, and X. Zhang, "A review on temperature and humidity control methods focusing on air-conditioning equipment and control algorithms applied in small-to-medium-sized buildings," vol. 162, Energy and Buildings, 2018, pp. 163–176.

[15] A. Kusiak, and G. Xu, "Modeling and optimization of HVAC systems using a dynamic neural network," vol. 42, Energy, 2012, pp. 241–250.

[16] P. V. Fazenda, K. Veeramachaneni, P. Lima, and U-M. O'Reilly, "Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems," vol. 6, J. of Ambient Intelligence and Smart Environments, 2014, pp. 675–690.

[17] K. Yan, C. Zhong, Z. Ji, and J. Huang, "Semi-supervised learning for early detection and diagnosis of various air handling unit faults," vol. 181, Energy and Buildings, 2018, pp. 75–83.

[18] W. Vallandares, M. Galindo, J. Gutierrez, W. C. Wu, K. K. Liao, J. C. Liao, K. C. Lu, and C. C. Wang, "Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm," vol. 155, Building and Environment, 2019, pp. 105–117.

[19] C. C. Cheng, and D. Lee, "Artificial intelligence-assisted heating ventilation and air conditioning control and the unmet demand for sensors: Part 1. Problem formulation and the hypothesis," vol. 19, Sensors, 2019, pp. 1131.

[20] S. J. Cox, D. Kim, H. Cho, and P. Mago, "Real time optimal control of district cooling system with thermal energy storage using neural networks," vol. 238, Applied Energy, 2019, pp. 466–480.

[21] M. S. Yoon, W. S. Yoon, and J. S. Lee, "Application of a deep reinforcement learning algorithm in household inverter air-conditioner temperature control," in Proceedings of the 13th IEA Heat Pump Conference (HPC 2020), 2021.

[22] M. S. Yoon, W. S. Yoon, M. K. Seo, S. Y. Ryu, and J. S. Lee, "Air conditioner component optimum opeartion point search through a deep reinforcement learning algorithm," in Proceedings of the 20th International Conference on Control, Automation and Systems (ICCAS 2020), 2020.

[23] M. S. Yoon, J. Lim, T. S. Qahtani, and Y. Nam, "Experimental study on comparison of energy consumption between constant and variable speed air-conditioners in two different climates," in Proceedings of the Asian Conference on Refrigeration and Air-conditioning (ACRA), 2018, E342.

[24] J. Lim, M. S. Yoon, T. S. Qahtani, and Y. Nam, "Feasibility study on variable-speed air conditioner under hot climate based on real-scale experiment and energy simulation," vol. 12, Energies, 2019, pp. 1489.

[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.

[26] R. S. Sutton, and A. G. Barto, Reinforcement learning: An introduction (2nd edition). MIT Press, Cambridge, MA, 2018.