DOI: https://doi.org/10.33414/ajea.1086.2022

Hacia un modelo de GEE con consideración del confort térmico y las estrategias bioclimáticas del edificio basado en MADRL

Towards a BEM model that considers thermal comfort and bioclimatic building strategies based on MADRL

Presentación: 4 y 5 de Octubre de 2022

Doctorando:
Germán Rodolfo Henderson
Grupo CLIOPE – Energía, Ambiente y Desarrollo Sustentable, Universidad Tecnológica Nacional Facultad Regional Mendoza
german.henderson@frm.utn.edu.ar

Director:
Alejandro Arena

Codirector:
Facundo Bromberg

Resumen

El consumo actual de energía primaria en el sector residencial es del 22% a nivel global y es la responsable de emitir 33 Gt de CO2eq. La gestión energética del edificio (GEE) permite hacer un uso eficiente de la energía y su consecuente disminución. El aprendizaje por refuerzos profundo (DRL) ha tenido logros significativos en otros campos y se ha comenzado a aplicar al de la GEE. En este trabajo se presenta un modelo de Multi-Agentes basado en DRL para la GEE en viviendas residenciales. El modelo busca la optimización del uso de la energía y el aumento del confort de los habitantes a partir del control de elementos bioclimáticos y la selección de la temperatura de climatización. Se evalúa el rendimiento del modelo inteligente contra otro convencional, basado en reglas. Los resultados encontrados hasta el momento son alentadores y se abren muchos caminos para la profundización en la investigación.

Palabras clave: Bioclimática, Aprendizaje por Refuerzos Profundo, Automatización, Vivienda.

Abstract

The current consumption of primary energy in the residential sector is 22% globally and is responsible for emitting 33 Gt of CO2eq. Building energy management (GEE) allows efficient use of energy and its consequent reduction. Deep reinforcement learning (DRL) has made significant gains in other fields and has begun to be applied to GEE. In this paper, a multi-agent model based on DRL for GEE in residential dwellings is presented. The model seeks to optimize the use of energy and increase the comfort of the inhabitants through the control of bioclimatic elements and the selection of the air conditioning temperature. The performance of the intelligent model is evaluated against a conventional one, rule based. The results found so far are encouraging and open many paths for further research.

Keywords: Bioclimatic, Deep Reinforcement Learning, Automation, Dwelling.

Introducción

Las estrategias bioclimáticas en la construcción son mecanismos que permiten realizar un uso eficiente de la energía al utilizar los recursos climáticos, como el sol y el viento. Su utilización permite disminuir la necesidad energética del edificio. Sin embargo, no siempre producen los ahorros esperados debido a que, en general, se requiere de usuarios activos que hagan funcionar los mecanismos, los cuales no siempre lo hacen por pereza o bien por otros factores que influyen en un funcionamiento distinto del esperado. Sistemas de gestión energética en edificios (GEE) que integren este tipo de consideraciones son necesarios para poder reducir el consumo de energía del sector residencial. Por otra parte, una de las consecuencias del cambio climático es la necesidad de que cada región deba revisar las estrategias bioclimáticas empleadas (Flores-Larsen, Filippín y Barea, 2019).

En 2016, las empresas de tecnología Google y Facebook quebraron la línea de conocimiento con el desarrollo de nuevos métodos de inteligencia artificial (IA) que han evolucionado en tecnologías de aprendizaje por refuerzos profundo (DRL) que combinan las redes neuronales profundas con el aprendizaje por refuerzos (Silver, 2016, y Tian y Zhu, 2016). El desarrollo del DRL en el sector de los edificios propone nuevas formas de resolver las problemáticas planteadas junto con otras relacionadas con la GEE.

En los últimos años se han desarrollado muchas investigaciones que buscan, a través del control de la temperatura de set point, el control de la temperatura del fluido caloportante del sistema de climatización, del control de flujo másico y otros, la optimización del uso de energía y el aumento del confort de los habitantes (Brandi et al., 2020; Coraci et al., 2021; Dermardiros, Bucking y Athienitis, 2019). Los resultados encontrados son alentadores, por ejemplo, Brandi et al. (2020) hallaron ahorros de energía entre el 5% y el 12%. Coraci et al. (2021) con un modelo de control similar obtuvo consumos un poco mayores, pero tasas de disconfort mucho menores, entre 75% y 48% para los escenarios de referencia. Esto último implica un mayor consumo energético para lograr mayores niveles de confort. Otras aplicaciones permiten además optimizar la integración de los recursos renovables distribuidos, como la energía solar fotovoltaica, y el almacenamiento de energía (Brandi, Fiorentini y Capozzoli, 2022; Brandi, Gallo y Capozzoli, 2022). Se ha observado que en general todos estos modelos dejan de lado las estrategias bioclimáticas, las cuales son tratadas de forma independiente.

La automatización de aberturas para el control de la ventilación natural encuentra aplicaciones de DRL como la de An et al. (2021), que encuentran mejores resultados que el control de ventanas con modelos basados en reglas (RB) al tratar de mitigar la contaminación del aire en los espacios interiores. Sin embargo, los autores no consideran la temperatura ni velocidad del viento en el interior, parámetros importantes del confort térmico. Han et al. (2020) opera el estado de ventanas para mejorar el confort térmico de los habitantes encontrando mejoras del 90% en comparación con los datos registrados en el edificio al cual se aplicó el modelo.

Este trabajo presenta los avances relacionados a la implementación de un modelo de recomendaciones para el control inteligente de una vivienda basado en DQN, un método de DRL, para la optimización del uso de energía y la maximización del confort de los habitantes. El modelo utiliza múltiples agentes, existiendo uno para operar cada uno de los elementos activos y pasivos de la vivienda. Para su evaluación se lo compara contra un modelo RB que realiza el control en el mismo entorno.

Desarrollo

Metodología

IA puede ser clasificada en tres grandes grupos: aprendizaje supervisado, no supervisado y por refuerzos. El aprendizaje por refuerzos (RL) se basa en el concepto de prueba-y-error, similar a como el humano aprende (Ertel, 2017).

Este concepto se puede definir como el Proceso de Decisión Markoviano (MDP). En el MDP un agente es definido como una entidad capaz de realizar acciones sobre el entorno para motivar los cambios de estado en este. El entorno es todo lo que está fuera del agente y con lo que el agente interactúa. El estado es una colección de propiedades que definen al entorno, y que el agente puede ver. La recompensa es una función del par estado-acción y le dice al agente que tan bueno es realizar esa acción en ese estado. El objetivo del agente es maximizar la recompensa obtenida a largo plazo en el entorno. Para ello utiliza algoritmos de aprendizaje similares al presentado en la ecuación (1), donde α es la tasa de aprendizaje y γ es la tasa de descuento de las recompensas futuras. Como concepto final se encuentra la política, que es un mapeo de las acciones a realizar para los diferentes estados del entorno (Sutton y Barto, 2018).

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t) \right] \quad (1)$$

Dos modelos son empleados en este trabajo. Por un lado, un modelo de MADRL denominado DQN (del inglés, Deep Q Network, que significa Redes Q Profundas) el cual utiliza redes neuronales profundas para estimar los valores de recompensa esperados en un par estado-acción (Mnih et al., 2013). Por otro lado, un modelo RB es utilizado para poder comparar y evaluar el desempeño del anterior.

Formulación del problema: el entorno

El entorno se lo ha definido este trabajo como una habitación, la cual ha sido confeccionada en EnergyPlus API Python versión 21.1.0. El entorno se puede visualizar en la figura 2, que muestra una habitación prismática. Los mecanismos bioclimáticos para el caso de estudio son la ventilación natural y la ganancia solar o control de sombras. Por otra parte, la climatización activa se realiza por medio de un calentador y un refrigerador. Se ha limitado la potencia de estos equipos activos a 400 W para el caso planteado.

La localización utilizada es la de Mendoza, Argentina (latitud -33.0 y longitud -69.0). El sitio se caracteriza por ser un oasis, localizado en el oeste del país a los pies de la cordillera de los Andes. Presenta un clima semi-árido con pocas precipitaciones y una gran cantidad de días claros. La amplitud térmica entre el día y la noche, y las brizas nocturnas, son utilizadas como estrategia de enfriamiento de la masa térmica de los edificios. Los datos meteorológicos fueron obtenidos con Meteonorm versión 7.3.

El proceso de aprendizaje del MADRL

Se ha planteado un sistema de 4 agentes, cada uno de ellos operando uno de los dispositivos accionables: temperatura de set-poit dual (calefacción y refrigeración), apertura de ventana norte, apertura de ventana sur y persiana exterior en la ventana norte. El entorno creado en EnergyPlus API Python envía observaciones a cada uno de los agentes, los cuales devuelven acciones luego de consultar la política aprendida hasta el momento. Luego de 30 pasos de tiempo (cada paso de tiempo se lo ha establecido en 15 minutos), se envía un bache de experiencia al buffer. El algoritmo de aprendizaje DQN, o entrenador de políticas, utiliza este buffer para entrenar las redes neuronales profundas de cada una de las políticas. Luego de recibir un bache de experiencias, se actualizan las políticas que utilizan los agentes para tomar decisiones. La utilización de un buffer de experiencias permite evitar el sesgo de la red neuronal, como así también permite la utilización de nuevas experiencias para el aprendizaje.

Se utilizó una configuración similar en todas las redes neuronales, solo cambiando la capa de entrada (espacio de estados) y de salida (espacio de acciones) de cada una de ellas. La configuración general utiliza 4 capas ocultas de 64 neuronas totalmente conectadas y con una función de activación ReLU (Rectified Lineal Unit). La tasa de aprendizaje utilizada fue de $lr = 0.001$, tasa de descuente $\gamma = 0.99$ la capacidad del buffer de 50 000. Estos parámetros se decidieron utilizar luego de un afinamiento realizado con la herramienta Tune de Ray (Liaw et al., 2018). Se definen a continuación los espacios de estados y acciones y las funciones de recompensa de cada uno de los agentes.

El entrenamiento de los agentes se realizó utilizado la librería RLlib de Ray (Liang et al., 2018). Se utilizó una configuración de servidor-cliente en modo local, lo que permitió trabajar directamente con la API de EnergyPlus en Python.

Agente DSP: Control del set-point dual. Este decide qué acciones realizar sobre el entorno en cuanto al control de la temperatura del espacio interior. Para ello establece los valores de temperatura deseados de refrigeración y calefacción.

- *Espacio de acciones*: El espacio de acciones viene definido por el rango de temperaturas que el agente puede establecer. Se definieron los rangos de 18 °C a 28 °C para la refrigeración y de 17 °C a 27 °C para la calefacción, y estableciendo siempre la condición de que la temperatura de calefacción sea al menos 1 °C menor que la de refrigeración. Se establecen saltos discretos de 1 °C para este control. La cantidad de acciones es la combinación de todas las posibilidades, dando como resultado un espacio de acciones de $nA = 66$.
- *Espacio de estados*: El espacio de estados se lo ha determinado por las variables atmosféricas del sitio y del interior de la vivienda. Se han considerado un total de 9 variables: hora, paso de tiempo actual, temperatura de bulbo seco del sitio, temperatura de bulbo seco del espacio interior, velocidad del viento, dirección del viento, radiación solar global, radiación solar en el plano de la persiana y humedad relativa.
- *Función de recompensa*: El diseño de la función de recompensa es uno de los desafíos más importantes que hay que afrontar a la hora de implementar un modelo de RL (Sutton y Barto, 2018). Para el desarrollo de la función de recompensa en este trabajo se revisaron las utilizadas en los trabajos de Dalamagkidis et al. (2007), Dermardiros et al. (2019), Yoon y Moon (2019), Brandi et al. (2020), Han et al. (2020), Park et al. (2021) y Coraci et al. (2021). En general, todos los autores utilizan penalidades como señal para el agente.

Sin embargo, Coraci et al (2021) utiliza, además de las penalidades, una recompensa nula cuando no hay habitantes en el edificio o bien cuando el estado tiene una temperatura interior dentro del rango deseado. La función de recompensa de este agente se la definió como se muestra en las ecuaciones (2), donde E_i es la energía requerida para calefaccionar o refrigerara la habitación hasta la temperatura seteada. T_{dn} es el límite inferior de confort y T_{up} es el límite superior.

Agente NWP: Accionamiento de persiana.

- *Espacio de acciones*: Se utilizan acciones discretas y binarias, es decir, que la persiana puede ser abierta o cerrada de manera completa solamente. Esto reduce el espacio de acciones a 2.
- *Espacio de estados*: El espacio de estados se lo define por 4 variables de interés para este elemento, las cuales son: hora, paso de tiempo, temperatura interior y temperatura exterior.
- *Función de recompensa*: Para el funcionamiento de la persiana, lo que se espera es que esta se accione para impedir la ganancia solar en momentos de calor dentro de la vivienda, pero permitir su ingreso cuando hace falta calentar el ambiente interior. Por otra parte, se espera que la persiana se accione en las noches de días fríos para evitar las pérdidas radiativas, mientras que se espera se encuentre abierta en las noches de días de calor. Por otra parte, no es necesario conocer cuanta energía va a aportar este sistema pasivo, sino que solo saber que va a aportar algo ya es suficiente. Por ello se puede controlar a partir de un sensor de iluminación sencillo. La función de recompensa se presenta en las ecuaciones (3).

Agentes NW y SW: Accionamiento de ventana. El accionamiento de las ventanas tiene relación con la ventilación natural producida en la habitación. La distinción entre la ventana norte y sur es fundamental, ya que las presiones que causan la ventilación son distintas en cada una de ellas, como así también la capacidad que tienen para ventilar. Sin embargo, los espacios de acciones y estados y la función de recompensa de ambas son las mismas.

- *Espacio de acciones*: Al igual que con la persiana, se utilizan acciones discretas y binarias, es decir, que la ventana puede ser abierta o cerrada de manera completa solamente, dando como resultado un espacio de acciones de 2.
- *Espacio de estados*: El espacio de estados viene definido por 7 variables: hora, paso de tiempo, temperatura interior, temperatura exterior, velocidad de viento, dirección de viento y humedad relativa.
- *Función de recompensa*: La función de recompensa utilizada es un poco más compleja que las anteriores, utilizando como parámetros el ΔT definido anteriormente para el agente DSP y las condiciones de temperatura obtenidas en el interior y las deseadas y las condiciones exteriores de temperatura, tal como se lo presenta en las ecuaciones (4).

El modelo basado en reglas (RB)

Las reglas que sigue el modelo RB son las expresadas en las ecuaciones (5) a (8). Los rangos de operación del refrigerador y el calefactor de la ecuación (8) corresponden a los límites definidos para el rango de confort en los diferentes periodos de tiempo (T_{dn} y T_{up}).

Resultados

Se realizaron entrenamientos con valores de β igual a 2 y 5, factor que pondera el valor que diferentes usuarios le asignan al requerimiento energético. Para ambos casos el modelo propuesto de MADRL basado en DQN se desempeñó mejor que el RB. En la figura 4 se puede observar como las recompensas obtenidas son en general iguales o mayores que el RB. El modelo RB obtiene recompensas mucho menores para días fríos, mientras que para días calurosos o dentro del rango de confort establecido, se comporta tan bien como el modelo planteado.

Los desempeños encontrados para los diferentes modelos fueron de 2.45, 3.26 y 3.47 horas de confort por cada kWh requerido para RB, DQN ($\beta = 2$) y DQN ($\beta = 5$) respectivamente. Esto se ve reflejado en los ahorros de energía de un 36% y 41% de los modelos de DQN con respecto al RB. Sin embargo, se puede apreciar una pérdida de horas de confort en del 15% y 16% respectivamente, lo cual puede observarse en la figura 5.

$$r_{DSP_i} = \begin{cases} -\frac{\beta \times E_i}{\Delta T_i^2}, \text{ para } \Delta T_i \neq 0 \\ 0, \text{ para } \Delta T_i = 0 \end{cases}, \begin{cases} \Delta T_h = T_{dn} - T_o, \text{ para calefacción} \\ \Delta T_c = T_o - T_{up}, \text{ para refrigeración} \end{cases}, \; r_{DSP} = r_{DSP_h} + r_{DSP_c} \quad (2)$$

$$\begin{cases} si\ T_i > T_{up} \begin{cases} r_{nwb} = +1, \ para\ perciana\ baja \\ r_{nwb} = 0, \ para\ perciana\ alta \end{cases} \\ si\ T_i < T_{dn} \begin{cases} r_{nwb} = -1, \ para\ perciana\ baja \\ r_{nwb} = 0, \ para\ perciana\ alta \end{cases} \\ si\ T_{up} \leq T_i \leq T_{dn} \ \rightarrow \ r_{nwb} = 0 \end{cases} \quad (3)$$

$$\begin{cases} si\ T_o > T_{up} \begin{cases} si\ T_i > T_{up} \begin{cases} w_{rew} = -(T_i - T_{up})^2, \text{ para ventana abierta} \\ w_{rew} = 0, \text{ para ventana cerrada} \end{cases} \\ si\ T_i < T_{dn} \begin{cases} w_{rew} = +1, \text{ para ventana abierta} \\ w_{rew} = 0, \text{ para ventana cerrada} \end{cases} \\ si\ T_{up} \le T_i \le T_{dn} \ \rightarrow \ w_{rew} = 0 \end{cases} \\ si\ T_o < T_{dn} \begin{cases} si\ T_i > T_{up} \begin{cases} w_{rew} = +1, \text{ para ventana abierta} \\ w_{rew} = 0, \text{ para ventana cerrada} \end{cases} \\ si\ T_i < T_{dn} \begin{cases} w_{rew} = -(T_{dn} - T_i)^2, \text{ para ventana abierta} \\ w_{rew} = 0, \text{ para ventana cerrada} \end{cases} \\ si\ T_{up} \le T_i \le T_{dn} \ \rightarrow \ w_{rew} = 0 \end{cases} \\ si\ T_{up} \le T_o \le T_{dn} \ \rightarrow \ w_{rew} = 0 \end{cases} \quad (4)$$

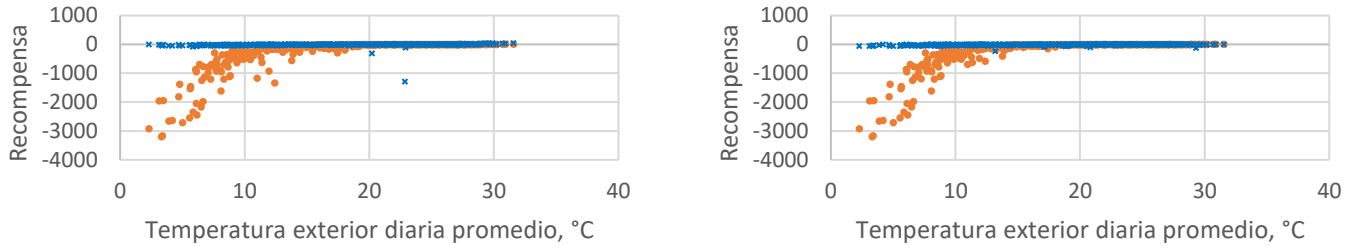
$$\text{Ventana: } Si \begin{cases} (T_i \ge T_{sp} + \Delta T) \text{ y } \begin{cases} (T_i > T_0) \ \Rightarrow \text{Abrir ventana} \\ (T_i \le T_0) \ \Rightarrow \text{Cerrar ventana} \end{cases} \\ (T_i \le T_{sp} - \Delta T) \text{ y } \begin{cases} (T_i > T_0) \ \Rightarrow \text{Cerrar ventana} \\ (T_i \ge T_0) \ \Rightarrow \text{Abrir ventana} \end{cases} \end{cases} \quad (5)$$

$$\text{Persiana: } Si \begin{cases} (T_i \ge T_{sp} + \Delta T) \text{ y } \begin{cases} (B_w = 0) \ \Rightarrow \text{Abrir persiana} \\ (B_w > 0) \ \Rightarrow \text{Cerrar persiana} \end{cases} \\ (T_i \le T_{sp} - \Delta T) \text{ y } \begin{cases} (B_w = 0) \ \Rightarrow \text{Cerrar persiana} \\ (B_w > 0) \ \Rightarrow \text{Abrir persiana} \end{cases} \end{cases} \quad (6)$$

Refrigerador: de 23 a 7 horas 28 °C y de 8 a 22 horas 25 °C (7)

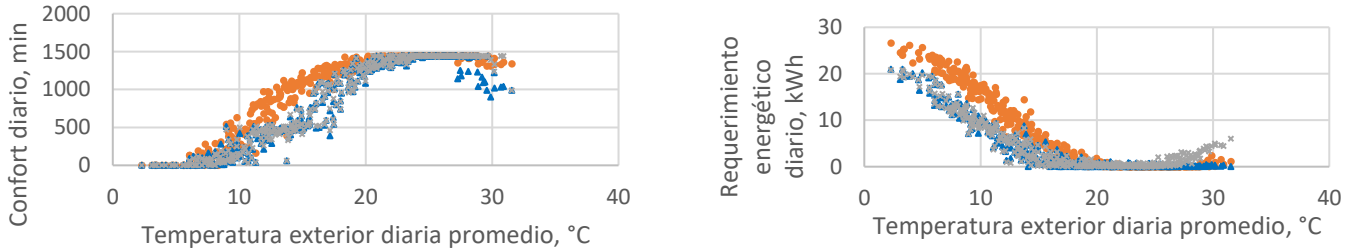
Calefactor: de 23 a 7 horas 17 °C y de 8 a 22 horas 20 °C (8)

La figura 5 se ha realizado en función de la temperatura promedio exterior diaria, ya que es un indicador de cómo han sido en general las variables climáticas en cada día. Se puede ver como el requerimiento energético se ve más afectado por la necesidad de calefacción mas que de refrigeración, alcanzando valores máximos de 26 kWh y 6 kWh respectivamente. Estas condiciones dependen tanto del clima como del tipo constructivo de vivienda, aspectos que se tendrán en cuenta explorar en futuras etapas de la investigación.



(a) (b)

Figura 4: distribución de la recompensa acumulada diaria con respecto a la temperatura exterior promedio diaria, para los métodos RB (círculos naranjas) y DQN (cruces azules) y una ponderación de (a) $\beta = 2$ y (b) $\beta = 5$.



(a) (b)

Figura 5: Dispersión del (a) tiempo de confort acumulado diario y (b) el requerimiento energético diario, con respecto a la temperatura exterior diaria promedio para los modelos RB (círculos naranjas) y DQN con ponderaciones de $\beta = 2$ (cruces grises) y $\beta = 5$ (triángulos azules).

Conclusiones y discusión

Se ha implementado un modelo basado en MADRL para el control inteligente de las estrategias pasivas y activas de climatización de una vivienda. Se encontraron ahorros energéticos de entre un 31% y 41% con respecto a un modelo RB y disminución en el confort de un 16%. Estos resultados siguen la conducta que se encontró en Coraci et al. (2021), en donde un gran aumento del confort requirió un pequeño aumento del requerimiento energético. Aquí, una gran disminución del requerimiento energético implicó una disminución en el confort.

Se ha podido observar como la variación del parámetro β que pondera el valor de la energía disminuye los ahorros energéticos encontrando cantidades un poco mayores en el confort de los habitantes.

En trabajos futuros se incluirán mayores variables en las observaciones de los agentes para permitir la predicción del clima y su afectación al desempeño energético y térmico de la vivienda. Por otra parte, se incluirán también otros elementos, como el almacenamiento energético y la generación de energía local a partir de fuentes renovables como la fotovoltaica. Será importante en trabajos futuros evaluar el comportamiento de los agentes en diferentes entornos generados por diferentes tipos constructivos y tamaños de vivienda, como así también diferentes regiones climáticas.

Referencias

An, Y., Xia, T., You, R., Lai, D., Liu, J., & Chen, C. (2021). *A reinforcement learning approach for control of window behavior to reduce indoor PM2. 5 concentrations in naturally ventilated buildings*. Building and Environment, 200, 107978. https://doi.org/10.1016/j.buildenv.2021.107978

Brandi, S., Fiorentini, M., & Capozzoli, A. (2022). *Comparison of online and offline deep reinforcement learning with model predictive control for thermal energy management*. Automation in Construction, 135, 104128. https://doi.org/10.1016/j.autcon.2022.104128

Brandi, S., Gallo, A., & Capozzoli, A. (2022). *A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings*. Energy Reports, 8, 1550-1567. https://doi.org/10.1016/j.egyr.2021.12.058

Brandi, S., Piscitelli, M. S., Martellacci, M., & Capozzoli, A. (2020). *Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings*. Energy and Buildings, 224, 110225. https://doi.org/10.1016/j.enbuild.2020.110225

Coraci, D., Brandi, S., Piscitelli, M. S., & Capozzoli, A. (2021). *Online implementation of a soft actor-critic agent to enhance indoor temperature control and energy efficiency in buildings*. Energies, 14(4), 997. https://doi.org/10.3390/en14040997

Dalamagkidis, K., Kolokotsa, D., Kalaitzakis, K., & Stavrakakis, G. S. (2007). *Reinforcement learning for energy conservation and comfort in buildings*. Building and environment, 42(7), 2686-2698. https://doi.org/10.1016/j.buildenv.2006.07.010

Dermardiros, V., Bucking, S., & Athienitis, A. K. (2019). *A simplified building controls environment with a reinforcement learning application*. Proceedings of the 16th IBPSA Conference, 956-964. https://doi.org/10.26868/25222708.2019.211427

Ertel, W. (Ed. 2). (2017). *Introduction to Artificial Intelligence*. Springer Cham. https://doi.org/10.1007/978-3-319-58487-4

Flores-Larsen, S., Filippín, C., & Barea, G. (2019). *Impact of climate change on energy use and bioclimatic design of residential buildings in the 21st century in Argentina*. Energy and Buildings, 184, 216-229. https://doi.org/10.1016/j.enbuild.2018.12.015

Han, M., May, R., Zhang, X., Wang, X., Pan, S., Da, Y., & Jin, Y. (2020). *A novel reinforcement learning method for improving occupant comfort via window opening and closing*. Sustainable Cities and Society, 61, 102247. https://doi.org/10.1016/j.scs.2020.102247

Liang, E., Liaw, R., Moritz, P., Nishihara, R., Fox, R., Goldberg, K., Gonzalez, J. E., Jordan, M. I., Stoica, I. (2018). *RLlib: Abstractions for Distributed Reinforcement Learning*. Proceedings of the35thInternational Conference on MachineLearning, Stockholm, Sweden. https://doi.org/10.48550/arXiv.1712.09381

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., & Stoica, I. (2018). Tune: A research platform for distributed model selection and training. arXiv preprint arXiv:1807.05118. https://doi.org/10.48550/arXiv.1807.05118

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). *Playing atari with deep reinforcement learning*. arXiv preprint arXiv:1312.5602. https://doi.org/10.48550/arXiv.1312.5602

Schiavon, S., & Lee, K. H. (2013). *Dynamic predictive clothing insulation models based on outdoor air and indoor operative temperatures*. Building and Environment, 59, 250-260. https://doi.org/10.1016/j.buildenv.2012.08.024

Park, B., Rempel, A. R., Lai, A. K., Chiaramonte, J., & Mishra, S. (2021). *Reinforcement Learning for Control of Passive Heating and Cooling in Buildings*. IFAC-PapersOnLine, 54(20), 907-912. https://doi.org/10.1016/j.ifacol.2021.11.287

Yoon, Y. R., & Moon, H. J. (2019). *Performance based thermal comfort control (PTCC) using deep reinforcement learning for space cooling*. Energy and Buildings, 203, 109420. https://doi.org/10.1016/j.enbuild.2019.109420