AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1703.2024
Publicado / Published: 10/2024

Control Energético Óptimo De Una Vivienda Con Múltiples Zonas Térmicas A Través De Aprendizaje Por Refuerzos Profundo

Energetic Optimal Control Of A Dwelling With Multiple Thermal Zones Through Deep Reinforcement Learning

Presentación: 8 y 9 de octubre de 2024

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

Director:
Alejandro ARENA

Codirector:
Facundo BROMBERG

Resumen

La implementación del aprendizaje por refuerzos profundo (DRL) ha avanzado significativamente en diversos campos científicos, superando muchas dificultades inherentes a su uso. Sin embargo, han surgido desafíos específicos en cada área. En el control de sistemas de climatización en edificios, se han identificado limitaciones de escalabilidad que dificultan su aplicación en entornos con múltiples zonas térmicas o numerosos agentes. Para abordar este problema, este trabajo presenta un método de control para múltiples agentes en múltiples zonas térmicas de una vivienda. Este método facilita la escalabilidad mediante la implementación de una política de control basada en una red neuronal profunda con parámetros completamente compartidos, utilizada por todos los agentes. Esta aplicación representa el estado del arte en sistemas multiagentes totalmente cooperativos, asegurando una comunicación efectiva entre los agentes para un control óptimo de la vivienda. La implementación de este método en una vivienda de interés social en la provincia de Mendoza demuestra su efectividad en escenarios complejos. Se discuten las limitaciones encontradas y se sugieren futuras líneas de investigación.

Palabras clave: Sistema Multiagentes, Aprendizaje por Refuerzos Profundo, Automatización, Inteligencia Artificial.

Abstract

The implementation of deep reinforcement learning (DRL) has significantly advanced across various scientific fields, overcoming many inherent difficulties. However, specific challenges have emerged in each area. In the control of HVAC systems in buildings, scalability limitations have been identified, hindering its application in environments with multiple thermal zones or numerous agents. To address this issue, this work presents a control method for multiple agents in multiple thermal zones of a dwelling. This method facilitates scalability by implementing a control policy based on a deep neural network with fully shared parameters, used by all agents. This application represents the state of the art in fully cooperative multi-agent systems, ensuring effective communication among agents for optimal control of the dwelling. The implementation of this method in a social housing unit in the province of Mendoza demonstrates its effectiveness in complex scenarios. The limitations encountered are discussed, and future research directions are suggested.

Keywords: Multi-Agent Systems, Deep Reinforcement Learning, Automation, Artificial Intelligence.

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1703.2024
Publicado / Published: 10/2024

Introducción

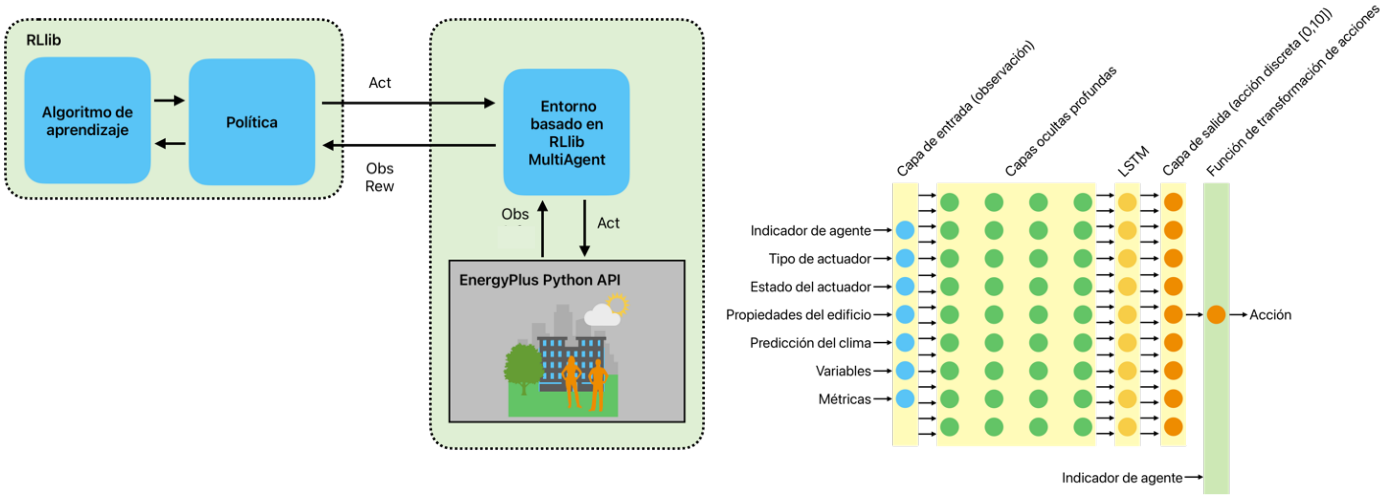
El desarrollo de un controlador automático de los equipos de climatización no es una tarea sencilla debido a la complejidad del entorno, como las características del edificio y su relación con el clima y sus habitantes (Zhang, Kannan, Kuppannagari, & Prasanna, 2019). En este contexto, el aprendizaje por refuerzos profundo (DRL) ha surgido como una metodología prometedora capaz de adaptarse a entornos dinámicos y estocásticos, utilizando redes neuronales profundas y redes neuronales recurrentes para ejercer el control óptimo de dispositivos.

La primera aplicación orientada a la gestión energética de edificios mediante DRL fue la de (Wei, Wang, & Zhu, 2017), quienes aplicaron DRL para el control discretizado del flujo másico de aire de diferentes zonas térmicas, logrando ahorros energéticos de entre el 22% y el 71% en comparación con controles convencionales. Otros autores, como (Gao & Li, 2019), se enfocaron en el control de las temperaturas de control del termostato, sin considerar la operación interna del equipo de climatización, sino la temperatura objetivo para la zona térmica.

Sin embargo, en estos trabajos, y hasta donde conocemos, las aplicaciones se basan en el entrenamiento de agentes simples (control de una sola variable), múltiples agentes que responden a una política cada uno, múltiples agentes que responden a una política que predice la acción conjunta óptima, políticas individuales para cada zona térmica o en el control a nivel de edificio (considerando las múltiples zonas térmicas como una sola desde el punto de vista del control). En todos estos casos, la escalabilidad se presenta como un problema que requiere aún ser estudiado con profundidad.

Grupo	**Variables**
Indicador del agente	Número entero único por cada agente.
Tipo de actuador	Número entero único por cada tipo de actuador.
Estado del actuador	Valor actual de control del actuador que controla el agente.
Propiedades del edificio/zona térmica	Área de planta; relación de aspecto; relación superficies de ventanas y muros en cada orientación; energía máxima para calefacción; energía máxima para refrigeración.
Predicción del clima	Precipitación; presión atmosférica; temperatura de bulbo seco; humedad relativa; velocidad y dirección del viento.
Variables	Del sitio: temperatura de bulbo seco, presión atmosférica, velocidad y dirección del viento, humedad relativa, radiación en el plano horizontal. De la zona térmica: temperatura media del aire, humedad relativa, conteo de personas. Otras variables: hora, día de la semana, día del año, si está o no lloviendo, si hay o no luz solar.
Métricas	Energía eléctrica, energía para calefacción y energía para refrigeración.

Tabla 1. Espacio de observaciones clasificado por grupo de variables.



(a) MDP (b) Política

Figura 1. El esquema en (a) muestra la implementación de la metodología de control basada en DRL, mientras que en (b) se expone un visual simplificada de la arquitectura de la política.

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1703.2024
Publicado / Published: 10/2024

En este trabajo se presenta una metodología para la gestión energética en viviendas con múltiples zonas térmicas basada en el aprendizaje por refuerzos profundo. En particular, se estudia el control coordinado de los niveles de temperatura de los termostatos para calefacción y refrigeración, así como el flujo másico de aire acondicionado suministrado por los equipos ubicados en cada una de las zonas térmicas de una vivienda. Esta metodología contribuye al desarrollo de controladores eficientes que operen en escenarios complejos.

Metodología

El aprendizaje por refuerzos (RL) se lo puede implementar como un proceso de decisión de Markov (MDP), en donde uno o más agentes son capaces de observar su entorno y aplicar acciones sobre este. El objetivo de cada agente es maximizar su recompensa a largo plazo, la cual la obtiene al aplicar una política. La política es una función que le dice al agente qué acción ejecutar dada una observación. Ésta se optimiza a través de un algoritmo de aprendizaje, que utiliza la interacción de los agentes con su entorno para aprender a tomar las decisiones correctas a través de un mecanismo de prueba-y-error. En el DRL, a diferencia de otros métodos de RL, la política viene representada por una red neuronal profunda (DNN) y el proceso de aprendizaje trata de ajustar los parámetros (o pesos) de la DNN.

En la figura 1(a) se presenta el MDP utilizado, en donde se pueden observar diferentes elementos que lo conforman. En entorno es implementado con la herramienta de simulación energética de edificios EnergyPlus versión 23.2.0 (Crawley, y otros, 2001) para una predicción efectiva del comportamiento del edificio y se lo integra a partir de su interfaz de programación de aplicaciones (API) con RLlib (Liang, y otros, 2018), utilizado como marco de trabajo para la implementación en DRL.

El modelo de EnergyPlus envía observaciones (Obs) al entorno implementado, el cual se encarga de calcular la recompensa y comunicar esta información a RLlib para poder realizar el entrenamiento de la DNN. Luego, RLlib devuelve una acción que es transformada y adaptada a los diferentes actuadores implementados en EnergyPlus. Este proceso se genera de forma continua, permitiendo la optimización de la política a través de un algoritmo de aprendizaje.

La política utilizada en este trabajo se esquematiza en la figura 1(b). Ésta se la configuró como una DNN de 5 capas con 256 neuronas cada una, todas interconectadas y activadas con una función Unidad Lineal Rectificada (ReLU). A esta DNN se le adicionó una capa recurrente de 256 celdas de Memoria Corta a Largo Plazo (LSTM), lo cual mejora el rendimiento de la política por tratarse de un entorno con una dinámica temporal. La política implementada es compartida por todos los agentes y todas las zonas térmicas, considerando que los agentes son totalmente cooperativos. Para ello es necesario establecer los espacios de observación y acción de cada agente de forma homogénea.

El espacio de observación se lo compone de variables habitualmente utilizadas en el control térmico de edificios, como la temperatura media del aire, la humedad relativa y la temperatura exterior, entre otras. Todas las variables de estado que los agentes pueden observar se encuentran en la tabla 1. En total se cuentan 173 variables de estado, entre las que hay 24 horas de predicción climática de 6 variables atmosféricas, lo cual es suficiente para la toma de decisiones (Henze & Schoenmann, 2003). En particular se destacan dos de ellas. La primera se lo denomina “indicador de agente” y es un número entero que representa a un agente en particular, permitiéndole a la política de pesos totalmente compartidos tener en cuenta qué agente la está consultando. La segunda es el tipo de actuador, lo que permite que agentes que controlan actuadores iguales en diferentes zonas térmicas puedan compartir el conocimiento experimentado.

El espacio de acciones, que representa la capa de salida de la DNN, se lo establece en un espacio discreto de 11 acciones. Luego, una función transforma la acción para que pueda ser utilizada correctamente por el actuador correspondiente. De esta manera, la Ecuación (1) adapta la temperatura requerida para calefacción en el rango de [18 °C; 22 °C], la requerida para refrigeración en [23 °C; 27 °C], y el flujo másico de aire en el rango de [0 m3/s; 0.5 m3/s], siendo a' la acción, a, transformada entre los valores inferior, $\underline{x}$, y superior, $\overline{x}$, del rango.

$$a' = \underline{x} + \frac{a}{10}\left(\overline{x} - \underline{x}\right) \tag{1}$$

El algoritmo de aprendizaje utilizado es Optimización de Políticas Próximas (PPO), el cual se lo considera como el estado del arte de los algoritmos de aprendizaje y se ha visto que en el ámbito de los edificios funciona mejor que otros. El factor de descuento de las recompensas futuras se establece en un valor de 0.8 y la tasa de aprendizaje en 0.001.

La función de recompensa se calcula como se anuncia en la Ecuación (2).

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1703.2024
Publicado / Published: 10/2024

$$r = -(\beta)\left(\frac{\Sigma_t^k E_t}{(k-t)*E_{ref}}\right) - (1-\beta)\left(\frac{1}{1-e^{-0.13(PPD_{av}-45)}}\right) \quad (2)$$

Donde β es un parámetro de ponderación que realiza un balance entre el consumo de energía y el confort alcanzado, E_t es el consumo de energía de cada paso de tiempo t, E_{ref} es el valor de energía que normaliza el término. PPD_{av} se lo calcula como se establece en la Ecuación (3), en la cual PPD_t es el parámetro del modelo de (Fanger, 1970) que indica el porcentaje de disconfort esperado para un estado dado en el paso de tiempo t, y PPD_{ref} es el valor de normalización, utilizando 100% por ser el máximo posible.

$$PPD_{av} = \frac{\Sigma_t^k PPD_t}{(k-t)*PPD_{ref}} \quad (3)$$

En este trabajo se ha considerado como caso de estudio una vivienda de interés social que integra estrategias bioclimáticas para el ahorro de energía de climatización, la cual ha sido modelada en EnergyPlus versión 23.2.0. Es un edificio diseñado por el Instituto Provincial de la Vivienda (IPV) de la provincia de Mendoza, en el que se han identificado 3 zonas térmicas climatizadas: la sala de estar-comedor, la habitación principal y una segunda habitación. La climatización se ha modelado como una bomba de calor ideal para cada espacio que cuentan con una potencia nominal de 2500 W, a las cuales se le puede regular el flujo másico de aire y los niveles de temperatura de este. En total el entorno cuenta con nueve agentes, tres por cada zona térmica acondicionada.

La política entrenada se la compara contra un controlador convencional basado en las reglas, las cuales establecen un nivel de temperatura para calefacción de 17 °C entre las 23 a las 7 horas y de 20 °C para el resto del tiempo. La temperatura para refrigeración se establece en 28 °C para el rango horario de 23 a 7 horas y de 25 °C para el resto del día. El flujo másico es variable, calculado en cada paso de tiempo para que se llegue a la condición deseada y con un máximo de 0.5 m^3/s.

Resultados

Se entrenó el modelo del edificio durante 720 000 pasos de tiempo, que representan 13.7 años de simulación. Cada año utilizó una base climática obtenida con Meteonorm versión 7.3 de diferentes sitios de la Argentina y el mundo. Luego, se evaluó el modelo entrenado para el clima de Mendoza, el cual no fue utilizado durante el entrenamiento.

El rendimiento de la política entrenada se evalúa desde dos enfoques. Por un lado, se considera el confort dentro de la vivienda cuando esta se encuentra ocupada. En la Tabla 2 se presenta el tiempo, expresado en porcentaje anual, en el cual se haya disconfort térmico en las diferentes zonas térmicas. Se puede observar como la política entrenada consigue igualar prácticamente las condiciones de confort en la vivienda. Si se consideran el total de horas en el año que no se obtuvo confort, la política propuesta mejora en un 4% este indicador con respecto al control de referencia, según los lineamientos descritos en el estándar ASHRAE 55-2004 (American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2004).

Por otra parte, se analizan los consumos de energía necesarios tanto para la calefacción como para la refrigeración de las diferentes zonas térmicas de la vivienda. En la Figura 2 se presentan los resultados del requerimiento energético y de potencia para cada controlador, el propuesto basado en DRL y el de referencia basado en reglas (RB). El propuesto presenta un aumento del 6% anual del consumo de energía, mientras que el requerimiento de potencia crece en un 7% para operar los mismos equipos de climatización. Este aumento de potencia sucede por un mayor flujo másico de aire refrigerado por unidad de tiempo.

	Tiempo del año con mucho frío [%]		**Tiempo del año con frío [%]**		**Tiempo del año con calor [%]**		**Tiempo del año con mucho calor [%]**	
	DRL	*RB*	*DRL*	*RB*	*DRL*	*RB*	*DRL*	*RB*
Estar-comedor	0%	0%	0%	0%	2%	2%	0%	0%
Habitación principal	14%	13%	32%	31%	0%	0%	0%	0%
Segunda habitación	15%	15%	34%	33%	0%	0%	0%	0%

Tabla 2. Tiempo de disconfort térmico para las diferentes zonas térmicas en valores anuales porcentuales de las horas ocupadas.

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1703.2024
Publicado / Published: 10/2024

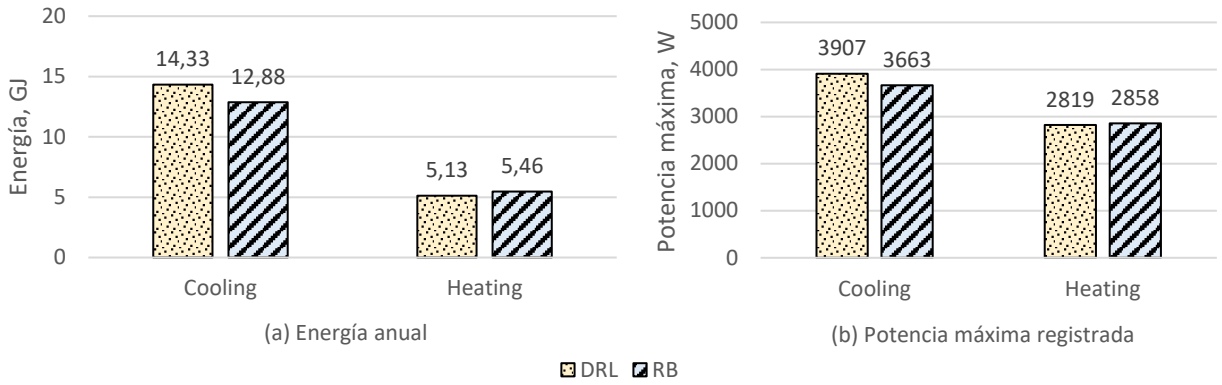


Figura 2. Demanda de energía y potencia máxima anuales para los controles DRL y en RB.

Discusión

Los consumos energéticos encontrados son mayores a los que el sistema convencional planteado requiere. Sin embargo, se encontró una mejor gestión del confort dentro de la vivienda. Para poder mejorar el rendimiento del controlador, es necesario realizar un entrenamiento más prolongado con un ajuste de los hiperparámetros de aprendizaje, como la tasa de aprendizaje, el factor de descuento de las recompensas futuras y el tamaño del lote de entrenamiento. Este último en particular, es importante para poder considerar las diferentes estaciones del año.

La metodología propuesta, basada en una política totalmente compartida por los agentes, es la primera vez que se aplica en el área del control de los edificios. Su desarrollo fue propuesto por (Gupta, Maxim, & Mykel, 2017) y brinda resultados efectivos en entornos totalmente cooperativos, como es el caso de este trabajo. Adicional al indicador de agente sugerido por los autores del trabajo adicional, aquí se ha propuesto un segundo indicador que sirve para categorizar a los agentes en grupos, según el actuador que controlan. Sin embargo, esta característica es necesario estudiarla en mayor profundidad.

Por otra parte, los espacios de observación y acción propuestos, junto con la configuración de la política, han permitido el planteo de una política escalable y adaptable a múltiples zonas térmicas con múltiples agentes. Esto permite dar un paso más en el desarrollo de un modelo generalizado, capaz de funcionar en diferentes entornos sin necesidad de contar con entrenamientos prolongados en modelos virtuales y permitiendo la implementación directa en escenarios reales.

En trabajos futuros se debe estudiar con mayor profundidad no solo el efecto del indicador de tipo de actuador, sino también el balance entre el consumo de energía y el confort a través del factor de ponderación β, que aún no es clara su efectividad (Zhang & Lam, 2018). Además, se deben considerar efectos como el de la ventilación natural y el control de sombras, y un perfil de ocupación realista. Estos tres se han considerado como se lo hace habitualmente, a través de calendarios horarios, pero sin un comportamiento estocástico, lo cual sería un escenario más realista.

Por último, es necesario evaluar la criticidad de los eventos de temperaturas extremas dentro del edificio. Es necesario implementar estrategias que le aseguren a las personas dentro de los edificios que durante la operación automática de los sistemas de climatización no se llegará a escenarios con temperaturas extremas.

Conclusiones

En este trabajo se presentó una metodología de control innovadora basada en DRL para la gestión energética de viviendas con múltiples zonas térmicas. Su aplicación a un caso de estudio simulado en EnergyPlus presentó mejoras en el confort de los habitantes con un incremente en el consumo de energía para la climatización del hogar. El control se aplicó sobre los valores de temperaturas del termostato y el flujo másico de aire acondicionado para equipos ubicados en tres zonas térmicas diferentes.

La integración del indicador único de agente y del indicador del tipo de actuador que este controla permitieron una gestión eficiente de múltiples zonas térmicas a través de una sola política para el hogar. Esta aplicación es la primera en el campo de investigación del control de edificios con DRL.

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1703.2024
Publicado / Published: 10/2024

En trabajos posteriores se avanzará con las limitaciones planteadas en la discusión de este trabajo. Principalmente se planea abordar las temáticas de generalización de la política y su ensayo en escenarios reales. La metodología aquí propuesta permite que se puedan utilizar diferentes entornos, diferentes cantidades de actuadores (que varían según el edificio) y un espacio de acción adecuado para los diferentes dispositivos que pueden aparecer en una vivienda. Un entrenamiento por aprendizaje de currículum (es decir, un entrenamiento por etapas en donde se van integrando de a poco diferentes aspectos o dispositivos a controlar que aumentan la complejidad de la tarea) permitiría obtener una coordinación de los equipos de climatización con la ventilación natural y el control de sombras.

Referencias

American Society of Heating, Refrigerating and Air-Conditioning Engineers. (2004). ANSI/ASHRAE Standard 55: thermal environmental conditions for human occupancy. Atlanta.

Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W., Huang, Y., Pedersen, C. O., . . . Glazer, J. (2001). EnergyPlus: creating a new-generation building energy simulation program. Energy and Buildings, 33(4), 319-331. doi:10.1016/S0378-7788(00)00114-6

Fanger, P. O. (1970). Thermal comfort. Analysis and applications in environmental engineering.

Gao, G., & Li, J. (2019). Energy-Efficient Thermal Comfort Control in Smart Buildings via Deep Reinforcement Learning. arXiv: Systems and Control.

Gupta, J. K., Maxim, E., & Mykel, K. (2017). Cooperative multi-agent control using deep reinforcement learning. Autonomous Agents and Multiagent Systems: AAMAS 2017 Workshops, Best Papers (págs. 66-83). São Paulo, Brazil: Springer International Publishing. Obtenido de https://ala2017.cs.universityofgalway.ie/papers/ALA2017_Gupta.pdf

Henze, G. P., & Schoenmann, J. (2003). Evaluation of Reinforcement Learning Control for Thermal Energy Storage Systems. HVAC&R Research, 259-275. doi:10.1080/10789669.2003.10391069

Liang, E., Liaw, R., Nishihara, R., Moritz, P., Fox, R., Goldberg, K., . . . Stoica, I. (2018). RLlib: Abstractions for Distributed Reinforcement Learning. International Conference on Machine Learning (ICML). Stockholm.

Wei, T., Wang, Y., & Zhu, Q. (2017). Deep Reinforcement Learning for Building HVAC Control. DAC '17: Proceedings of the 54th Annual Design Automation Conference 2017, (págs. 1-6). doi:10.1145/3061639.3062224

Zhang, C., Kannan, R., Kuppannagari, S. R., & Prasanna, V. K. (2019). Building HVAC Scheduling Using Reinforcement Learning via Neural Network Based Model Approximation. (A. f. Machinery, Ed.) BuildSys '19: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 287 - 296. doi:10.1145/3360322.3360861

Zhang, Z., & Lam, K. P. (2018). Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system. BuildSys '18, (págs. 148-157). Shenzen, China. doi:10.1145/3276774.3276775