AJEA - Actas de Jornadas y Eventos Académicos de UTN
VI Jornadas de Intercambio y Difusión de los Resultados de Investigaciones de los Doctorandos en Ingeniería
DOI: https://doi.org/10.33414/ajea.1035.2022

Entornos de Entrenamiento para Agentes de Aprendizaje por Refuerzo en Discrete Event System Specification

Training environments for Reinforcement Learning Agents in Discrete Event System Specification

Presentación: 4 y 5 de Octubre de 2022

Doctorando:
Ezequiel Beccaria
Facultad Regional Villa María – Universidad Tecnológica Nacional - Argentina
ebeccaria@frvm.utn.edu.ar

Director:
Jorge Andres Palombarini

Codirectora:
Verónica Bogado

Resumen

La dinámica y complejidad de los entornos industriales actualmente han llevado a la necesidad de soluciones que permitan capturar la interacción en tiempo real para tomar decisiones sobre el control de los procesos involucrados. El Aprendizaje por Refuerzo es un enfoque promisorio, que se aplica en problemas de decisión secuencial, donde la complejidad radica en la interacción agente-entorno y la incertidumbre subyacente del entorno, pero requiere de una simulación que refleje el proceso bajo control (entorno) y su dinámica para entrenar el agente. En este trabajo, se presenta una solución para entrenar este tipo de agentes con entornos modelados y simulados usando Discrete Event System Specification. El mismo se aplica al problema de generación y administración de una energía alterna, biogás producido por un digestor y usado por diferentes perfiles de consumidores industriales.

Palabras clave: Aprendizaje por Refuerzo, Discrete Event System Specification, Problemas de Decisión, Energías Renovables

Abstract

Nowadays, dynamics and complexity of industrial environments have led to the need for solutions that allow capturing the interaction in real-time to make decisions about the control of the involved processes. Reinforcement Learning is a promising approach to solve sequential decision problems, where the complexity lies in the agent-environment interaction and the underlying uncertainty of the environment. This requires a simulation that reflects the process under control (environment) and

its dynamics to train the agent. In this work, a novel solution is presented to train this type of agent with modeled and simulated environments using Discrete Event System Specification. The same applies to the problem of generation and administration of alternative energy, biogas produced by a digester and used by different industrial consumer profiles.

Keywords: Reinforcement Learning, Discrete Event System Specification, Secuential Decision Process, Renewable Energy

Introducción

El Aprendizaje por Refuerzo (AR) (Sutton et al., 2018) se ha convertido en uno de los campos de más rápido crecimiento como metodología para brindar capacidades de aprendizaje a los agentes de *Inteligencia Artificial (IA)* que deben encontrar políticas de acción para diferentes *Problemas de Decisión Secuencial (PDS)* complejos. Un aspecto limitante en el uso del AR es la necesidad de contar con un entorno de entrenamiento para llevar a cabo el aprendizaje de los agentes (Sutton et al., 2018) .

Este entorno permite al agente AR experimentar estados diferentes del PDS y la consecuencia de las acciones disponibles en ellos, y así poder inferir una política de acción que maximice la recompensa del agente. Para aplicar soluciones de AR a nivel industrial, no basta pensar en mecanismos que aseguren el correcto desempeño del agente, sino también, en el nivel de correlación entre el entorno de entrenamiento, donde se realiza el aprendizaje, y el proceso real, es decir, es necesario reducir la brecha entre el proceso real y el entorno simulado. En la actualidad, no se ha prestado mucha atención al desarrollo de entornos de entrenamiento formales para el entrenamiento de agentes AR, con el fin de evitar la incertidumbre, el riesgo y las posibles pérdidas económicas de una mala correlación del mismo.

En (Beccaria et al., 2021) fue definido un marco metodológico para entrenar agentes AP de forma sistemática con el objetivo de representar problemas de control donde existan eventos exógenos dependientes del tiempo. Es decir, captura no solo la dinámica del proceso a controlar, sino también, los eventos externos que influyen sobre este. Para ello, se propone el uso de Procesos de Decisión Generalizados Semi-Markovianos (Generalized Semi-Markov Decision Process – GSMDP, Younes et al., 2004) como formalismo para el modelado del problema de control.

En la Figura 1, se definen gráficamente los aspectos involucrados en el problema de decisión y cómo estos se representan en la solución propuesta. Así, el Entorno se captura en un Modelo GSMDP que describe el funcionamiento del proceso de control a simular: el conjunto S de estados, un conjunto E de eventos, los relojes de activación C_e de los eventos E, la función transición P (s_{t+1} | s_t , e_t), la función de recompensa r(s_t , e_t , s_{t+1}) y la función F_e (τ|s) encargada de determinar el momento de activación para cada evento e en el estado s.

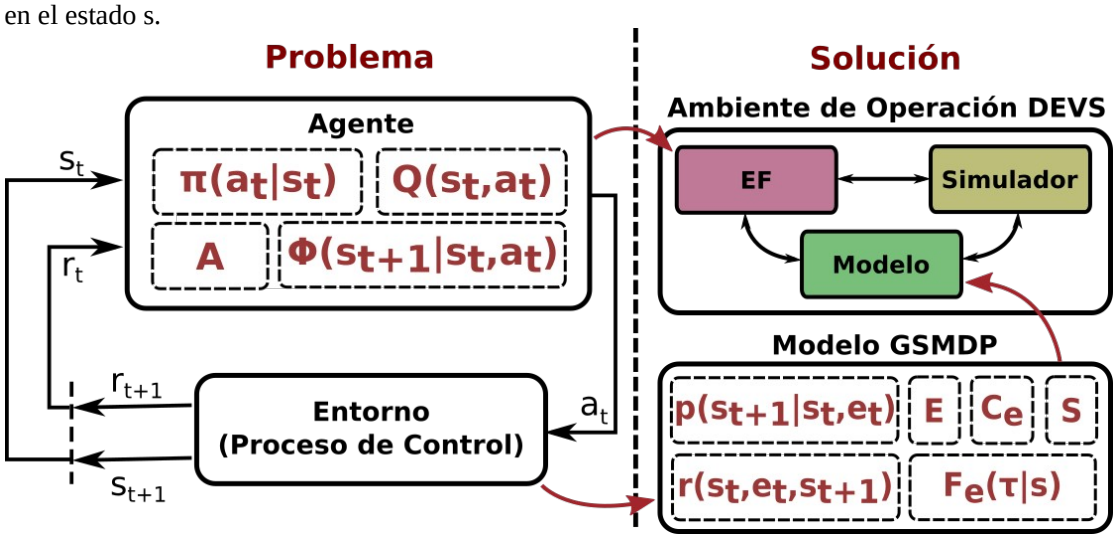

Figura 1: Entorno de entrenamiento para agentes basado en DEVS

Una vez definido el modelo del entorno, es necesario poder ejecutar una simulación del mismo para que el agente pueda experimentar acciones y observar los efectos de las mismas en términos de transiciones de estados y recompensas recibidas. Al utilizar el formalismo Discrete EVent System Specification (DEVS) (Concepcion et al., 1988) como motor de simulación del entorno o proceso bajo control, se debe realizar una transformación previa del Modelo GSMDP definido, a un modelo DEVS (Modelo). Para esto, se utilizan las reglas de transformación definidas en (Rachelson, 2009). El Agente integra el mecanismo de decisión y aprendizaje, siendo embebido en otro modelo DEVS incluido dentro del marco experimental (EF).

En este trabajo, se presenta una aplicación del marco metodológico definido en (Beccaria et al., 2021) al problema de generación y administración de energía generada a través del proceso de digestión anaeróbica, y así reducir el consumo eléctrico tradicional para distintos perfiles de consumidores industriales.

Desarrollo

Como caso de estudio, se presenta un esquema de oferta-demanda de energía eléctrica, donde distintos perfiles de consumidores industriales utilizan energía generada mediante medios renovables (biogás y solar), para reducir el consumo eléctrico de red. En este esquema, un agente AR es el responsable de administrar el biogás generado, para reducir el costo del consumo eléctrico total de los distintos perfiles de consumidores industriales (Figura 2). El *consumidor* es una de las piezas variables de este esquema, dado que cada uno de estos, presenta variaciones en el nivel de la energía que requieren para operar, como así también en los horarios donde desarrollan su actividad.

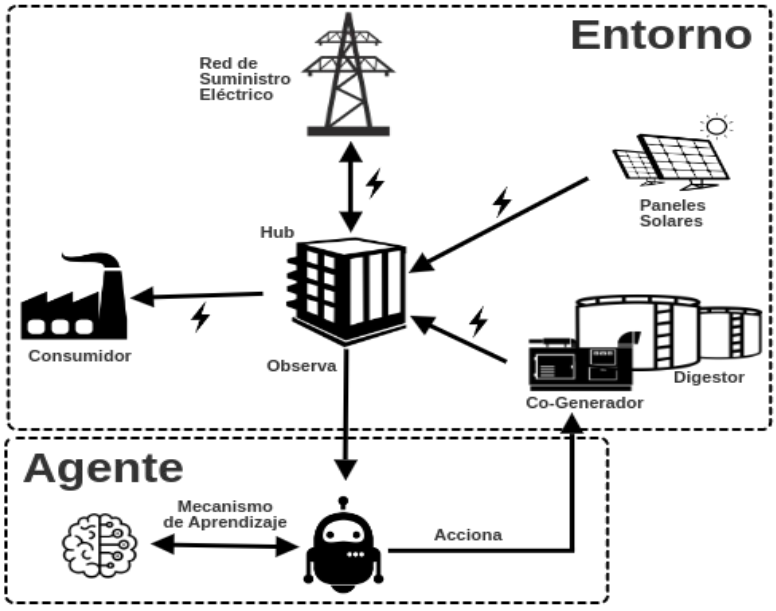

Figura 2: Entorno de control para la gestión de energías renovables.

El componente *Hub/Controlador*, es donde llega la energía generada por los medios renovables y en caso de existir un excedente, vende el mismo a la red de suministro eléctrico. En cuanto a los medios de generación de energía renovables, la granja de paneles solares (Chezzi et al., 2017) no tiene ningún mecanismo de control, y el nivel de energía generado es dependiente del clima, la fecha y hora simulada. En cuando a la energía generada a partir de biogás, se simula un arribo diario de materia biológica al digestor y la correspondiente producción del biogás utilizando el modelo definido en (Beccaria et al., 2018). Este biogás almacenado es utilizado para encender un co-generador que produce energía eléctrica para ser utilizada por el consumidor. El objetivo del agente es determinar el mejor momento para encender el co-generador, en base al estado global del entorno, y así utilizar el gas almacenado para minimizar el consumo eléctrico de red total del consumidor. El estado del entorno de entrenamiento esta compuesto por:

- La energía consumida por el consumidor CE_{Wh}.
- La energía pico consumida por el consumidor CP_W.
- La energía producida por la granja de paneles solares S_{Wh}.
- La energía producida por el co-generador a partir del biogás almacenado cuando este esta encendido CO_{Wh}. En caso contrario el valor es igual a 0 (cero).
- El volumen de biogás almacenado m^3.
- Fecha y hora de la simulación.

La recompensa percibida por el agente en cada etapa de decisión (cada 1 hora de simulación) es:

$$r_t = b_1 (S_{Wh}+CO_{Wh}) - (c_1 CE_{Wh} + c_2 max(CP_W)) \quad (1)$$

donde b_1 es cuanto paga el distribuidor de electricidad por la energía inyectada a la red, c_1 es el costo de venta de electricidad por parte del distribuidor y c_2 un costo adicional por el máximo de energía pico utilizada por el consumidor.

Se poseen 2 años de datos de consumo de los distintos perfiles de consumidores industriales, por lo que se utiliza el primer año de los datos para llevar a cabo el entrenamiento del agente, y el segundo año es utilizado como conjunto de prueba para evaluar la política de acción aprendida por el mismo.

Los distintos perfiles de consumo industrial utilizados son: una fábrica de alimento balanceado para mascotas, un molino harinero y una fábrica metalúrgica. La cantidad de energía consumida por cada uno de los perfiles de consumidores utilizados en la simulación fueron provistos por el ente de distribución eléctrica de la zona “Empresa Provincial de Energía de Córdoba” (EPEC - https://www.epec.com.ar/).

La cantidad de material bio-degradable que llega al digestor para la producción de biogás (una media diaria de 2.5T) es simulada a partir de los datos de desperdicios diarios que genera una planta de procesamiento de carne porcina. Adicionalmente, se simula la producción de energía renovable a partir de una planta de generación de energía fotovoltaica de 100 m^2, que complementa la energía generada con el biogás producido por el digestor.

Los parámetros utilizados a la hora de determinar la capacidad de generación y almacenamiento de biogás, como así también, la capacidad de generación de energía eléctrica a partir del biogás producido son los definidos en (Beccaria et al., 2018). En cuanto a la producción de energía fotovoltaica, el modelo utilizado esta definido en (Chezzi et al., 2017). Los datos climáticos para determinar la cantidad de energía solar producida se obtuvieron de una estación meteorológica local. Como algoritmo de aprendizaje para el entrenamiento del agente AR, se utiliza Proximal Policy Optimization (PPO) (Schulman et al., 2017). Como linea base de desempeño, también se evaluaron un agente aleatorio y un agente con una política base. En esta última, cada vez que el consumo energético del consumidor se dispara, y existan reservas de biogás, encenderá el co-generador para disminuir el consumo energético de red.

Resultados

En la Figura 2 y en la Tabla 1, se pueden visualizar los resultados obtenidos, en cada fila se pueden visualizar los resultados de cada uno de los agentes en cada uno de los escenarios, el porcentaje mostrado en cada celda representa la mejora relativa con respecto al agente de peor desempeño por fila.

Al evaluar la política de acción aprendida luego de llevar a cabo el entrenamiento del agente PPO, el mismo demostró tener un desempeño superior al resto de los agentes en todos los escenarios. Al aprender una política de acción para el perfil *Fábrica Metalúrgica*, la mejora de desempeño relativa del agente PPO es de un 22% en el costo del consumo energético con respecto al agente de menor desempeño (Linea Base) y de un 17% con respecto al agente aleatorio. En los restantes perfiles de consumo, la mejorar relativa de desempeño es de un 0.5% y 0.19% respectivamente. Esto es debido a que los niveles de consumo y los horarios de operatoria diurnos de estos perfiles, donde la mejora relativa es menor, hacen que se reduzca el margen de acción posible por parte de un agente debido a la existencia de la planta de generación de energía fotovoltaica complementaria, que durante el día aporta energía al sistema.

Perfil Consumidor	**Aleatorio**	**Linea Base**	**PPO**
Fábrica Balanceado	0.0 %	0.18 %	**0.19 %**
Fábrica Metalúrgica	5 %	0.0 %	**22 %**
Molino Harinero	0.1 %	0.0 %	**0.5 %**

Tabla 1: Mejora de desempeño relativa entre agentes

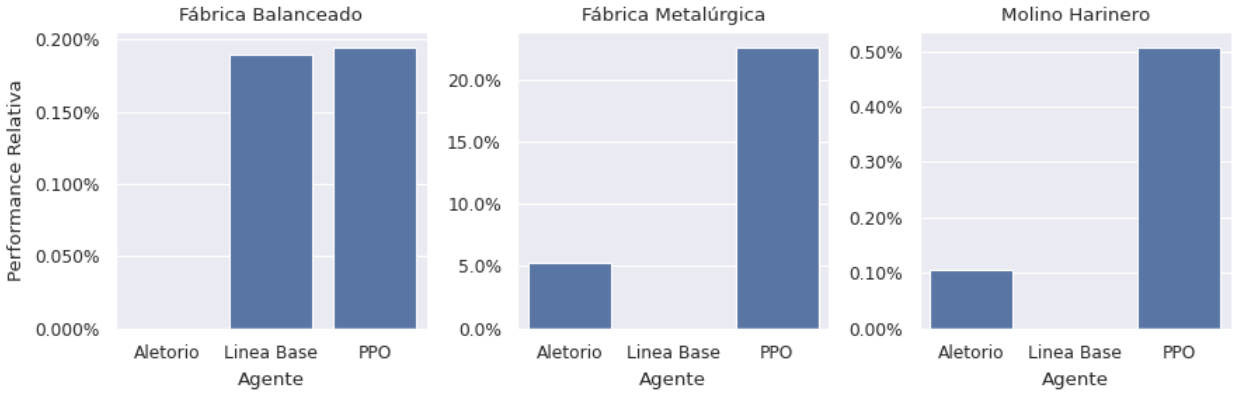

Figura 3: Desempeño relativo por agente para cada perfil de consumidor.

Conclusiones

En este trabajo, se propuso un entorno de simulación basado en DEVS para entrenar agentes AP aplicado a la generación y administración de biogás con el fin de facilitar el proceso de toma de decisiones en tiempo real. Esta solución permite acelerar los tiempos de entrenamiento del agente AP, mejorando su performance a menor costo, combinando las ventajas de usar AP en un proceso donde la toma de decisiones secuencial se da en tiempo real con DEVS como herramienta para especificar y simular dicho proceso y su dinámica. En particular, en el contexto de energías renovables, esta propuesta permite analizar diferentes escenarios de uso de energías alternativas, en este caso biogás, modificando las demandas de consumo, la capacidad de producción de metano, como así también pudiendo escalar el sistema al incorporar un mayor número de componentes, por ejemplo, componentes que representen la producción de biogás a partir de materia orgánica diversa. Asimismo, permite mejorar la gestión de dichos procesos mediante la modificación de los parámetros.

Si bien existen propuestas basadas en modelos más precisos para predecir la cantidad de biogás producido por digestión anaeróbica, el objetivo principal de este trabajo es desarrollar un modelo de simulación que permita capturar lo mejor posible el proceso que se está simulando, es decir, su complejidad, considerando no solo una variable sino varias simultáneamente como, por ejemplo, la producción de biogás, el consumo, factores externos (eventos) que pueden afectar, entre otros. Es importante destacar ésto ya que este entorno de simulación es el entorno con el que interactúa el agente AP, del cual aprende. Al ser un enfoque basado en DEVS provee ventajas como la definición de elementos de simulación específicos del dominio (problemática energética), el desacople entre el modelo, el simulador (interno) y el marco experimental, facilitando la reutilización de componentes de simulación y su evolución en el tiempo. Esto permite una construcción modular y jerárquica del proceso que se está controlando, pudiendo capturar tanta complejidad como se necesite para que el agente aprenda una política. Todas estas características definen una herramienta flexible para tomar decisiones relacionadas a la generación, almacenamiento y uso de energías alternativas.

En trabajos futuros, se pretende desarrollar una herramienta de software que incluya el marco general de entrenamiento de agentes RL usando modelos de simulación basados en DEVS para facilitar su usabilidad en diferentes problemas. En particular, respecto a la problemática de energías renovables, se pretende trabajar con otros algoritmos de AP, para así lograr una mejor política de administración de la energía renovable generada.

Referencias

Beccaria, E., Bogado, V., & Palombarini, J. A. (2018). “A devs-based simulation model for biogas generation for electrical energy production”. *2018 IEEE Biennial Congress of Argentina (ARGENCON)* (pp. 1-8). IEEE.

Beccaria, E., Bogado, V., & Palombarini, J. A. (2021). “A DEVS Based Methodological Framework for Reinforcement Learning Agent Training”. *IEEE Latin America Transactions*, *19*(4), 679-687.

Chezzi, C. M., Bordón, F., Lerman R., Tymoschuk A. R. (2017). “Modelo DEVS para Evaluación de Asignación de Energía Solar para Vivienda Estándar”. *V Congreso Nacional de Ingeniera Informática - Sistemas de Información (CONAIISI)*. Facultad Regional Santa Fe - Universidad Tecnológica Nacional. 890-898.

Concepcion, A. I., & Zeigler, B. P. (1988). “DEVS formalism: A framework for hierarchical model development”. *IEEE Transactions on Software Engineering*, *14*(2), 228-241.

Rachelson, E. (2009). “*Temporal markov decision problems”*. *Thesis* (Ph. D. in Artificial Intelligence), Université Paris.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). “Proximal policy optimization algorithms”. *arXiv preprint arXiv:1707.06347*.

Sutton, R. S., & Barto, A. G. (2018). “Reinforcement learning: An introduction”. *MIT press*.

Younes, H. L., & Simmons, R. G. (2004, July). “Solving generalized semi-Markov decision processes using continuous phase-type distributions”. *AAAI* (Vol. 4, p. 742).