

Entorno de Simulación basado en DEVS para Agentes de Aprendizaje por Refuerzo aplicado a la Generación y Administración de Energías Renovables

Ezequiel Beccaria, Veronica Bogado, and Jorge A. Palombarini

Facultad Regional Villa María - UTN, Villa María, X5900 HLR Argentina; CIT Villa María - CONICET-UNVM, Villa María, Córdoba, X5900 HLR Argentina
{ebeccaria, vbogado, jpalombarini}@frvm.utn.edu.ar
<http://www.frvn.utn.edu.ar>

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 particular, el uso de energías alternativas representa un problema con enormes desafíos, donde existen procesos críticos que necesitan monitoreo y control en tiempo real de un ambiente altamente incierto. En este trabajo, se presenta una solución para entrenar este tipo de agentes con entornos modelados y simulados usando DEVS. 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.

Keywords: Aprendizaje por Refuerzo, DEVS, Problemas de Decisión, Energías Renovables

1. Introducción

El *Reinforcement Learning (RL)* [24] 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 de RL es la necesidad de contar con un entorno de entrenamiento para llevar a cabo el aprendizaje de los agentes [24]. Este entorno permite al agente RL 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 RL 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 RL, con el fin de evitar la incertidumbre, el riesgo y las posibles pérdidas económicas de una mala correlación del mismo.

Este trabajo, se propone un entorno de entrenamiento para agentes RL basado en Discrete Event System Specification (DEVS) [9]. Dicho entorno general se aplica 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.

El resto de este trabajo se organiza de la siguiente forma. En la sección 2, se realiza una descripción detallada de los trabajos relacionados. Luego, en la sección 3 se presenta la definición del problema. En la sección 4, se desarrolla un caso de estudio describiendo la implementación computacional y los parámetros utilizados para las simulaciones y, por último, en la sección 6, se presentan algunas conclusiones sobre el actual trabajo y las expectativas a cubrir en trabajos futuros.

2. Trabajos Relacionados

Respecto de la temática abordada en la presente propuesta, no existen antecedentes directos en los cuales se aborde de manera integral el modelado y simulación generativa del proceso de generación, almacenamiento y administración del consumo de energía eléctrica a partir de biogás. Sin embargo, en [6] se define un marco para la definición de entornos de entrenamiento no formales para agentes RL, una interface común entre estos (agente y entorno) y una gran cantidad de entornos de prueba pre-implementados. En [4] se define un marco metodológico para el desarrollo de entornos de entrenamiento formales para el entrenamiento de agentes RL, el cual es utilizado para el desarrollo del caso de estudio propuesto en este trabajo.

Existen algunos antecedentes relacionados con el modelado y simulación del proceso de generación de biogás a partir de desechos orgánicos, de manera acotada a la evaluación de la producción de metano en relación a la variabilidad del sustrato empleado. En ese sentido, en [18] se emplea un modelo basado en Redes Neuronales (RN) para determinar la performance de un biorreactor a escala de laboratorio. En [20], [26] y [5], un modelo similar se emplea para desarrollar una metodología de análisis del proceso de producción de biogás, donde se utiliza un algoritmo de optimización basado en Algoritmos Genéticos [12] para identificar las variables relevantes en la predicción del flujo de biogás. Con el mismo objetivo, en otros trabajos se han desarrollado modelos analíticos enfocados en mejorar la estabilidad y el desempeño del proceso de producción de biogás para aumentar la eficiencia de operación de planta. Así, en [21], se ha desarrollado

un modelo simple con el objetivo de representar adecuadamente las dinámicas de la Digestión Anaeróbica (DA) a partir del ajuste de tres sustratos principales (proteínas, carbohidratos y lípidos). En [15], se emplea un enfoque basado en simulación para estudiar los efectos del uso de residuos sólidos municipales incinerados en la producción de biogás empleando la ecuación de Gompertz para simular el rendimiento obtenido. Por otra parte, en [10], se predice el comportamiento observado de la DA a partir de un modelo bio-cinético basado en los balances de masa del sustrato, la biomasa y la producción de metano. En [27], se utilizan procesos de optimización aplicados al sustrato con el cuál se alimenta el biodigestor para obtener mejores rendimientos de metano. En [23] y [2], se aplica un modelo matemático con parámetros variables en el tiempo para describir el comportamiento dinámico de la DA de residuos animales. Por otra parte, en [17], se realiza la simulación de un proceso de DA de biomasa agrícola para analizar costos de producción. En [1], se lleva a cabo un estudio teórico de la DA para predecir la cantidad de biogás generado a partir de desechos orgánicos agrícolas, y en [19], se simula el proceso de DA a partir de un modelo de primer orden, para predecir el rendimiento de metano empleando como sustrato residuos orgánicos municipales y barro biológico. En [11], se desarrolla un complemento en MATLAB para realizar la optimización, el análisis de estabilidad y el control de una planta industrial de biogás.

Por otro lado, con respecto al modelado y simulación de gestión energética, en [13] se realiza el modelado y simulación de los distintos componentes de una red eléctrica inteligente de manera tal de tomar decisiones de diseño adecuadas y en [7] se emplea DEVS para simular la operación de una granja de turbinas de viento y asistir en la toma de decisiones relativa a su mantenimiento. En [8] se define un modelo de generación de energía solar y en [3] se define un modelo de simulación para la generación de energía a partir de biogás producido a partir de desechos biológicos. Ambos modelos basados en DEVS, serán utilizados como parte del entorno de simulación propuesto en este trabajo.

3. Definición del Problema

En la actualidad, el paulatino agotamiento de los combustibles fósiles y la necesidad global de una reducción en las emisiones de gases de efecto invernadero han atraído especial interés a los procesos de producción de energía no contaminante de manera sustentable y sostenible en el tiempo, como es el caso de la digestión anaeróbica de residuos biológicos [25]. El uso de energías renovables tiene como condicionante la incertidumbre a la hora de contar con la disponibilidad de la energía generada de manera constante (un generador eólico no puede generar energía en días de poco viento). El biogás producido a partir de desechos animales no esta exento de este problema.

A diferencia de la energía eólica, el costo de almacenar el biogás para ser utilizado en el momento adecuado no es prohibitivo, pero esto presenta el inconveniente de tener que administrar el uso del mismo, para contar con esa energía en el momento indicado para optimizar costos de producción. La dinámica de

los distintos procesos industriales hace que esta tarea no sea trivial, presentando la necesidad de contar con metodologías de control en tiempo real para la administración de estos procesos.

Para poder utilizar *RL* como marco de referencia para la solución de este tipo de problemas, es necesario contar con un entorno de entrenamiento con el cual el agente interactúe y aprenda (ver Fig. 1). El *Entorno* implica todo lo que no es directamente controlable por el Agente. Ambas entidades interactúan continuamente, el Agente ejecutando acciones A_t y el Entorno respondiendo a dichos estímulos con nuevos estados S_t y premios o castigos R_t por sus acciones, que el agente debe tratar de maximizar con el paso del tiempo [24].

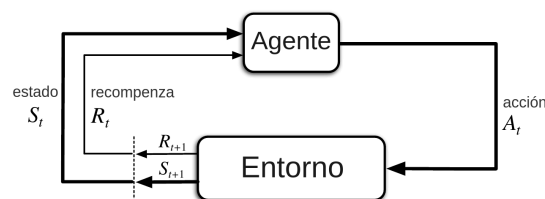


Figura 1. Interacción entre Agente y Entorno en RL

Debido a la gran cantidad de interacciones que necesita un agente RL para poder aprender una política cercana a la óptima y el riesgo de experimentar acciones que pueden implicar perjuicios económicos y sociales, desarrollar el entrenamiento del agente en un entorno real se vuelve una tarea sumamente inviable. Esta situación, presenta la necesidad de contar con entornos de entrenamiento simulados que tengan un alto grado de correlación con el entorno real y, así, reducir la incertidumbre y el riesgo de sufrir posibles pérdidas económicas.

El formalismo DEVS se presenta como un gran candidato para lograr esto, ya que permite definir modelos específicos teniendo en cuenta su adaptación a cualquier ambiente en donde existen componentes que interactúan entre sí siguiendo un esquema de caja negra. Asimismo, DEVS es una herramienta para el modelado y simulación basado en la teoría de sistemas y provee conceptos bien definidos de acoplamiento entre componentes, construcción modular y jerárquica de componentes siguiendo el paradigma orientado a objetos, el cual facilita el reuso de componentes de manera natural. Para realizar este trabajo, se sigue una metodología *bottom-up* donde primero se definieron los DEVS atómicos, las relaciones entre ellos y, finalmente, los DEVS acoplados [9].

En este contexto, utilizar DEVS para modelar y simular un entorno de entrenamiento para agentes RL implica el mapeo de cada uno de los componentes definidos en la Fig. 1 a uno, o más, modelos DEVS. Los detalles de este mapeo se pueden encontrar en [3, 4], estos son omitidos por cuestiones de espacio.

4. Caso de Estudio

A la hora elegir la metodología para llevar a delante el caso de estudio, lo cual implica el desarrollo del entorno de simulación, la verificación del mismo, la selección y entrenamiento del agente RL, y la posterior evaluación del mismo, se siguieron los paso definidos en [4].

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 RL 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).

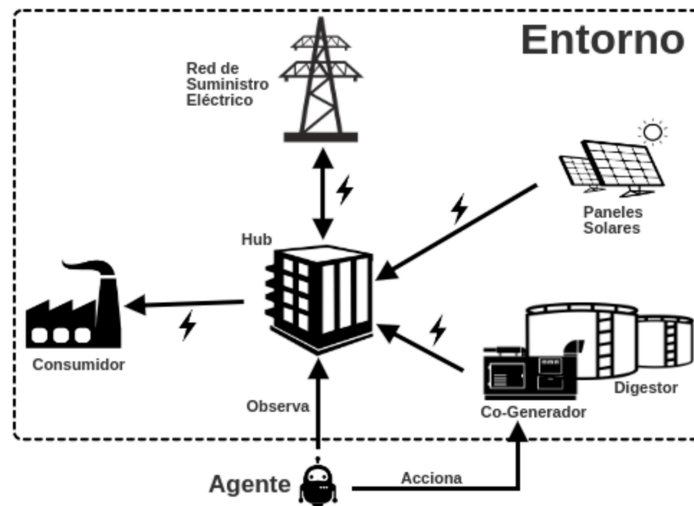


Figura 2. Entorno de control para la gestión de energías renovables.

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. 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 [8] 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 [3]. 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_{Wh} .
- 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:

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

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.

4.1. Parámetros Experimentales

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 provisto por el ente de distribución eléctrica de la zona “Empresa Provincial de Energía de Córdoba” (EPEC)¹.

La cantidad de material bio-degradable que llega al digestor para la producción de biogás (una media diaria de 2,5 T) 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 [3]. En cuanto a la producción de energía fotovoltaica, el modelo utilizado esta definido en [8]. Los datos climáticos para determinar la cantidad de energía solar producida se obtuvieron de una estación meteorológica local.

¹ <https://www.epec.com.ar/>

Como algoritmo de aprendizaje para el entrenamiento del agente RL, se utiliza Proximal Policy Optimization (*PPO*) [22]. PPO ha demostrado a lo largo de su uso en numerosos problemas, un excelente compromiso entre capacidad de aprendizaje y eficiencia en la cantidad de tiempo necesario para encontrar una política de acción cercana a la óptima. Los parámetros utilizados por el algoritmo de aprendizaje se encuentran definidos en la tabla 1.

Tabla 1. Definición de Hiper-parámetros para PPO

Hiper-parámetros	<i>PPO</i>
Capas ocultas (Tamaño)	2 (32)
Activación	ReLU - [16]
Tamaño de salida	2+1
Activación salida	Softmax(2) + Identity(1)
Método de optimización estocástica	<i>ADAM</i> - [14]
Hilos	16
Ratio de aprendizaje	$3e-4$
Ratio de descuento	0.99
Normalización de entrada	Si
Normalización de recompensa	Si

Como línea 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 generador para disminuir el consumo energético de red.

5. Resultados

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. En la Figura 3 y en la Tabla 2, se pueden visualizar los resultados obtenidos. 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 (Línea Base) y de un 17% con respecto al agente aleatorio. En los restantes perfiles de consumo, la mejora relativa de desempeño es de un 0.5% y 0.19% respectivamente. Esto es debido a que los niveles de consumo bajos y los horarios de operatoria diurnos de los perfiles 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.

6. Conclusiones y Trabajo Futuro

En este trabajo, se propuso un entorno de simulación basado en DEVS para entrenar agentes RL aplicado a la generación y administración de biogás con

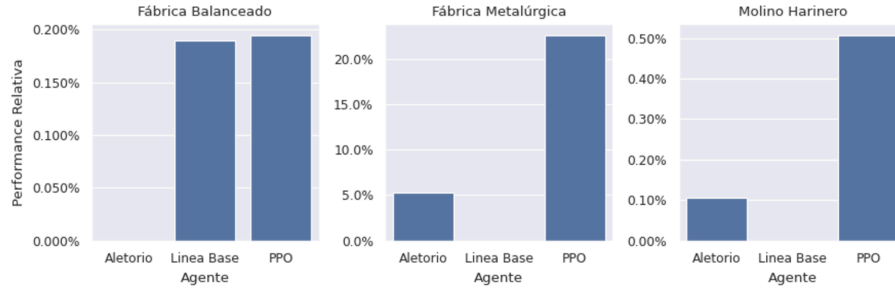


Figura 3. Desempeño relativo por agente para cada perfil de consumidor.

Tabla 2. Mejora de desempeño relativa entre agentes.

Perfil Consumidor	Random	Fixed	PPO
Fábrica Balanceado	0.0 %	0.0018 %	0.0019 %
Fábrica Metalúrgica	0.05 %	0.0 %	0.22 %
Molino Harinero	0.001 %	0.0 %	0.005 %

el fin de facilitar el proceso de toma de decisiones en tiempo real. Esta solución permite acelerar los tiempos de entrenamiento del agente RL, mejorando su performance a menor costo, combinando las ventajas de usar RL 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 RL, 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 RL, para así lograr una mejor política de administración de la energía renovable generada.

Referencias

- [1] Spyridon Achinas y Gerrit Jan Willem Euverink. "Theoretical analysis of biogas potential prediction from agricultural waste". En: *Resource-Efficient Technologies 2.3* (2016), págs. 143-147.
- [2] BK Bala. "System dynamics modelling and simulation of biogas production systems". En: *Renewable energy 1.5-6* (1991), págs. 723-728.
- [3] Ezequiel Beccaria, Veronica Bogado y Jorge A Palombarini. "A devs-based simulation model for biogas generation for electrical energy production". En: *2018 IEEE Biennial Congress of Argentina (ARGENCON)*. IEEE. 2018, págs. 1-8.
- [4] Ezequiel Beccaria, Veronica Bogado y Jorge Andres Palombarini. "A DEVS Based Methodological Framework for Reinforcement Learning Agent Training". En: *IEEE Latin America Transactions 19.4* (2021), págs. 679-687.
- [5] Tetyana Beltramo y col. "Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm". En: *Biosystems Engineering 143* (2016), págs. 68-78.
- [6] Greg Brockman y col. *OpenAI Gym*. 2016. eprint: [arXiv:1606.01540](https://arxiv.org/abs/1606.01540).
- [7] Eunshin Byon y col. "Simulation of wind farm operations and maintenance using discrete event system specification". En: *Simulation 87.12* (2011), págs. 1093-1117.
- [8] Carlos M. Chezzi y col. "Modelo DEVS para Evaluación de Asignación de Energía Solar para Vivienda Estándar". En: *V Congreso Nacional de Ingeniería Informática - Sistemas de Información*. Facultad Regional Santa Fe - Universidad Tecnológica Nacional. 2017, págs. 890-898.
- [9] A. I. Concepcion y B. P. Zeigler. "DEVS Formalism: A Framework for Hierarchical Model Development". En: *IEEE Transactions on Software Engineering 14.2* (feb. de 1988), págs. 228-241. ISSN: 0098-5589. DOI: 10.1109/32.4640.
- [10] M Fedailaine y col. "Modeling of the anaerobic digestion of organic waste for biogas production". En: *Procedia Computer Science 52* (2015), págs. 730-737.
- [11] Daniel Gaida y col. "MATLAB toolbox for biogas plant modelling and optimization". En: *Progress in Biogas II-Biogas production from agricultural biomass and organic residues* (2011), págs. 67-70.

- [12] David E Goldberg. “Genetic algorithms in search, optimization, and machine learning, 1989”. En: *Reading: Addison-Wesley* (1989).
- [13] Moath Jarrah. “Modeling and simulation of renewable energy sources in smart grid using DEVS formalism”. En: *Procedia Computer Science* 83 (2016), págs. 642-647.
- [14] Diederik P Kingma y Jimmy Ba. “Adam: A method for stochastic optimization”. En: *arXiv preprint arXiv:1412.6980* (2014).
- [15] HM Lo y col. “Modeling biogas production from organic fraction of MSW co-digested with MSWI ashes in anaerobic bioreactors”. En: *Bioresource Technology* 101.16 (2010), págs. 6329-6335.
- [16] Andrew L Maas, Awni Y Hannun y Andrew Y Ng. “Rectifier nonlinearities improve neural network acoustic models”. En: *Proc. icml*. Vol. 30. 1. 2013, pág. 3.
- [17] Maizirwan Mel y col. “Simulation study for economic analysis of biogas production from agricultural biomass”. En: *Energy Procedia* 65 (2015), págs. 204-214.
- [18] Vijay V Nair y col. “Artificial neural network based modeling to evaluate methane yield from biogas in a laboratory-scale anaerobic bioreactor”. En: *Bioresource technology* 217 (2016), págs. 90-99.
- [19] A Nielfa, R Cano y M Fdz-Polanco. “Theoretical methane production generated by the co-digestion of organic fraction municipal solid waste and biological sludge”. En: *Biotechnology Reports* 5 (2015), págs. 14-21.
- [20] H Abu Qdais, K Bani Hani y N Shatnawi. “Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm”. En: *Resources, Conservation and Recycling* 54.6 (2010), págs. 359-363.
- [21] Anna Schneider. “Dynamic modeling and simulation of biogas production based on anaerobic digestion of gelatine, sucrose and rapeseed oil”. Tesis doct. Jacobs University Bremen, 2016.
- [22] John Schulman y col. “Proximal Policy Optimization Algorithms”. en. En: (2017), pág. 12.
- [23] Iv Simeonov, V Momchev y D Grancharov. “Dynamic modeling of mesophilic anaerobic digestion of animal waste”. En: *Water Research* 30.5 (1996), págs. 1087-1094.
- [24] *Sutton & Barto Book: Reinforcement Learning: An Introduction*. 2018.
- [25] Peter Weiland. “Biogas production: current state and perspectives”. En: *Applied microbiology and biotechnology* 85.4 (2010), págs. 849-860.
- [26] Kaan Yetilmezsoy y col. “Development of ann-based models to predict biogas and methane productions in anaerobic treatment of molasses wastewater”. En: *International journal of green energy* 10.9 (2013), págs. 885-907.
- [27] Martin Zaefferer, Daniel Gaida y Thomas Bartz-Beielstein. “Multi-fidelity modeling and optimization of biogas plants”. En: *Applied Soft Computing* 48 (2016), págs. 13-28.