Hierarchical deep reinforcement learning for hydrogen supply chain management

Geunseo Songa, b , Vahid Khalighc, J. Jay Liuc, Jonggeol Naa, b,\*

aDepartment of Chemical Engineering and Materials Science, Ewha Womans University, Seoul 03760, Republic of Korea

bGraduate Program in System Health Science and Engineering, Ewha Womans University, Seoul 03760, Republic of Korea

cDepartment of Chemical Engineering, Pukyong National University, Busan 48513, Republic of Korea

jgna@ewha.ac.kr

Abstract

For an effective transition from fossil fuel-based energy sources to renewable energy sources, it is crucial to accompany research on the design and optimization of supply chain for new energy sources. The traditional tool for optimizing supply chain management (SCM), the mathematical programming (MP) method, has limitations in terms of computation time and cost as the scale and complexity of the supply chain increase. Therefore, we need a new powerful optimization methodology that enables real-time decision making, considers the interaction among various components within the supply chain, and accommodates the uncertainties in demand and energy supply. In this study, we proposed deep reinforcement learning (DRL) as a new tool to overcome the limitations of MP and satisfy the conditions required for optimizing SCM, as mentioned earlier. Furthermore, we aim to compare a single-agent reinforcement learning (SARL) system with a multi-agent reinforcement learning (MARL) system. Our model achieves successful performance by converging to a value like the optimum of the MP.

**Keywords**: Hydrogen supply chain, Hierarchical deep reinforcement learning, Multi-agent, Operation scheduling, Optimization

* 1. Introduction

As hydrogen receives attention as an energy source that can replace fossil fuels, research on hydrogen energy is being actively conducted. Fundamental research on hydrogen is important, but for effective conversion of energy sources, hydrogen supply chain (HSC) design and optimization must be accompanied.

Some studies have considered the operation scheduling of a small-scale hydrogen supply chain (HSC) consisting of multiple hydrogen refueling stations (HRSs) and hydrogen distribution system (HDS) in the retail hydrogen market using a mathematical programming (MP) approach (Mohammad H. Shames, 2022). Previously, MP approach was a major tool to optimize the supply chain management (SCM), however, as the scale and complexity of the supply chain increase, optimizing it with MP leads to an exponential increase in computation time and cost. Also, there are limitations in considering fluctuations in demand, energy generation, and other factors within the supply chain. Therefore, in order to apply the supply chain model to the real-world scenarios, it is necessary to build a complex model and find a new powerful tool capable of optimizing it.

In this work, we propose reinforcement learning (RL) as a new methodology to optimize the operation of a comprehensive hydrogen supply chain encompassing production, transportation, storage, and distribution. RL is a category of machine learning algorithms specifically designed for sequential decision-making, offering a promising solution to tackle these challenges (Boute et al. 2022). RL is used in many fields of chemical engineering, such as scheduling (Lee, J.M. 2004) and finding molecular structures (Sanchez-Lengeling, B. 2017), but it has not yet been widely applied in the field of supply chain management. When applying the DRL method for HSC optimization, not only is it advantageous to expand the supply chain, but it can also consider the relationship of each part of the hydrogen supply chain and cope with various fluctuations.

* 1. Methodology
		1. RL approach for hydrogen supply chain optimization

We developed a RL model to optimize the operation of a comprehensive hydrogen supply chain, encompassing production, transportation, storage, and distribution of hydrogen. Formally, RL is composed of five elements: agent, environment, state *S*, action *A*, and reward *R*. During the learning process, the agent observes states from the environment, selects actions under a certain policy, and receives corresponding rewards. The goal of RL is to find the optimal policy of actions that maximizes the ultimate cumulative reward. The process is mathematically formalized as a Markov Decision Process (MDP). The MDP functions as a flexible framework for goal-directed learning, defined as a tuple $M = (S, A, P(s\_{t+1}, r| s, a), R, γ)$.



**Figure 1.** Schematic representation of reinforcement learning (RL) model architecture. The hydrogen supply chain (HSC), which serves as the environment, encompasses production, transportation, storage, and distribution, representing interactions between each part

The schematic diagram of the RL model for optimizing the operation of the hydrogen supply chain is depicted in Figure 1. In production part, hydrogen can be produced from various primary energy sources, such as green domestic hydrogen through the electrolysis process, steam methane reforming (SMR) using natural gas as a feedstock, obtained as a by-product in a refinery, or imported in the form of NH₃. The produced hydrogen is transported and stored through various means depending on the phase and is distributed to different regions according to demand. The RL agent interacts with the environment and determines the time, location, type and capacity of different facilities to minimize the net present value of the overall cost.

* + 1. Agent architecture

a

b

**Figure 2.** (a) Single agent hierarchical action space architecture. In this system, actions are composed of three levels, and the agent receives a single reward at the end of each episode. (b) Hierarchical multi-agent architecture. This system consists of three agents, and at the end of each episode, the three agents receive different rewards individually.

In this section, we propose two types of agents. A RL model for hydrogen SCM optimization must incorporate a hierarchical decision-making process as it determines the following sequentially: 1) hydrogen production quantity and methods based on demand, 2) transportation means and storage locations based on the type of the produced hydrogen, and 3) distribution according to the hydrogen load in each region. The agent illustrated in Figure 2a is a single agent capable of hierarchical decision-making through hierarchical action space, while the agent represented in Figure 2b satisfies this through multi-agent interactions.

* + - 1. Single agent architecture

The RL agent is a single agent with a hierarchical action space, and as shown in Figure 2a, every action consists of three levels of decisions. In the first level decision process, the agent determines the region for hydrogen production and the quantity of hydrogen produced based on four production methods, including electrolysis, SMR, refinery, and NH₃. In the second level, the agent decides on the type of transportation such as tube-trailer, truck, pipeline, and tanker truck based on whether the produced hydrogen is in gaseous or liquid or liquid carrier form. Also, in this level, the agent decides the location for hydrogen storage. Finally, in the third level, the agent determines the amount of hydrogen to distribute from storage tanks to each region based on the hydrogen demand.

After an episode concludes, the decided action pair is conveyed to the environment, inducing a state change, and the agent receives a corresponding reward. The objective of hydrogen SCM optimization is to minimize the net cost incurred in the investment and operation of the hydrogen supply chain. Therefore, the reward is defined as in equation (1). Since the agent has a single reward function, it can exhibit the best performance in minimizing the incurred cost as long as it converges well.

|  |  |
| --- | --- |
| $$Reward=-(Genlnv+TransInv+StrInv+HRSInv+GenOp+TransOP+StrOP+HRSOp)$$ | (1) |

* + - 1. Multi agent architecture

The RL agent depicted in Figure 2b forms a multi-agent system consisting of three agents. In contrast to a single agent, this system involves interactions not only between the agent and the environment but also among the agents themselves. Therefore, each agent not only predicts the value of its individual actions but also learns to take actions by considering the actions of other agents. The first agent determines the quantity of hydrogen based on each production method and, at the end of the episode, receives the corresponding reward, which is formulated in equation (2). The second agent decides on the hydrogen transportation means and storage locations, receiving the reward corresponding to equation (3). Finally, the third agent determines the amount of hydrogen to distribute to each region and receives the reward specified in equation (4).

In the real world, entities within each part of the supply chain have different objectives. Multi-agent RL (MARL) system allows for multi-objective optimization as each agent operates by observing the actions of other agents to maximize its own reward. The operating costs of the supply chain may be somewhat higher than in a single-agent system, but this can be interpreted as a well-reflected outcome of the real-world complexities entwined with relationships among multiple stakeholders.

|  |  |
| --- | --- |
| $$Reward=-(Genlnv+GenOp)$$ | (2) |
| $$Reward=-(Translnv+TransOp+StrInv+StrOp)$$ | (3) |
| $$Reward=-(HRSlnv+HRSOp)$$ | (4) |

3. Results



**Figure 3.** Episode reward mean

**Figure 4.** Demand satisfaction of hydrogen refueling stations

To assess the validity of RL methodology compared to conventional mathematical programming (MP) techniques for optimizing the supply chain, we designed a simple hydrogen supply chain model and conducted optimization. Figure 3 shows that the episode reward mean value converges well as the agents are learned. The HSC model consists of a wholesale market that sells three types of hydrogen, a hydrogen distribution system, and two hydrogen refueling stations. The system involves purchasing hydrogen in accordance with demand and selling it to hydrogen vehicles. By satisfying the hydrogen load as depicted in Figure 4, we confirmed the capability of demand-driven decision-making using RL methodology.

* 1. Conclusion and future works

In this study, we have developed a RL framework to optimize a comprehensive hydrogen supply chain model, including production, transportation, storage, and distribution. RL agents found optimal policies to interact with the supply chain environment to determine the type and location of each facility, and when to minimize the costs incurred. Moreover, RL models overcome the computational limitations of conventional MP methods, enabling real-time decision-making. Furthermore, the performance of the results trained through RL converged with the optimal values obtained through MP, demonstrating the viability of the methodology. In the future, we aim to analyze the differences in agent decision-making results to build and train a single agent hierarchical behavioral space RL framework and a hierarchical multi-agent RL framework. Furthermore, we intend to build a model that considers the unique dynamics of the domestic hydrogen market, conduct a case study, and analyze the trade-off between the economy and the environment.

References

L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, 2021. A Review of Deep Reinforcement Learning

for Smart Building Energy Managent, IEEE Internet of Things Journal, 8, 9426901

Mohammad H. Shams, Haider Niaz, J.Jay Liu, 2022. Energy management of hydrogen refueling

stations in a distribution system: A bilevel chance-constrained approach, Journal of Power Sources, 533, 231400

David Silver, et al. 2018. A general reinforcement learning algorithm that masters chess, shogi,

and Go through self-play, Science 362.6419, pp. 1140-1144.

Younghoon Lee and Seunghoon Lee, 2022. Deep reinforcement learning based scheduling within

production plan in semiconductor fabrication, Expert Systems with Applications 191, p.116222.

Laura Stops and Roel Leenhouts and Qinghe Gao and Artur. 2022. Flowsheet synthesis through hierarchical reinforcement learning and graph neural networks, eprint arXiv:2207.12051.

Zhou Fan and Rui Su and Weinan Zhang and Yong Yu. 2019. Hybrid Actor-Critic Reinforcement Learning in Parameterized Action Space, arXiv:1903.01344.

Boute, Robert N., Joren Gijsbrechts, Willem van Jaarsveld, and Nathalie Vanvuchelen. 2022. Deep Reinforcement Learning for Inventory Control: A Roadmap. European Journal of Operational Research298 (2): 401–412.

Gang, Z., and S. Ruoying. 2006. Policy Transition of Reinforcement Learning for An Agent Based SCM System. In 2006 IEEE International Conference on Industrial Informatics, INDIN'06, Singapore, 793–798.

Giannoccaro, I., and P. Pontrandolfo. 2002. Inventory Management in Supply Chains: A Reinforcement Learning Approach. International Journal of Production Economics 78 (2): 153–161.

Hirano, M., H. Matsushima, K. Izumi, and T. Mukai. 2021. Simulation of Unintentional Collusion Caused by Auto Pricing in Supply Chain Markets. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12568 LNAI, 352–359. Cham: Springer.

Lee, J. M., Lee, J. H., 2004. Approximate dynamic programming strategies and their applicability for process control: A review and future directions. International Journal of Control, Automation, and Systems, 2(3), 263-278.

Sanchez-Lengeling, B., Outeiral, C., Guimaraes, G.L. and Aspuru-Guzik, A., 2017. Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv, 2017.