Systems & Control Transactions

Research Article
Foundations of Computer Aided Process Design (FOCAPD 2024)
Breckenridge, Colorado, USA. July 14-18, 2024
Peer Reviewed Conference Proceeding

PSE PRESS

Learn-To-Design: Reinforcement Learning-Assisted Chemical Process Optimization

Eslam G. Al-Sakkari[a,b], Ahmed Ragab[a,b]*, Mohamed Ali[c], Hanane Dagdougui[a], Daria C. Boffito[d], and Mouloud Amazouz[b]

[a] Polytechnique Montréal, Department of Mathematics and Industrial Engineering, Montréal, Québec, H3T 1J4, Canada
[b] CanmetENERGY Varennes, Varennes, Québec, J3X 1P7, Canada
[c] CanmetENERGY Devon, Devon, Alberta, T9G 1A8, Canada
[d] Polytechnique Montréal, Department of Chemical Engineering, Montréal, Québec, H3T 1J4, Canada
* Corresponding Author: ahmed.ragab@nrcan-rncan.gc.ca

ABSTRACT

This paper proposes an AI-assisted approach aimed at accelerating chemical process design through causal incremental reinforcement learning (CIRL) where an intelligent agent is interacting iteratively with a process simulation environment (e.g., Aspen HYSYS, DWSIM, etc.). The proposed approach is based on an incremental learnable optimizer capable of guiding multi-objective optimization towards optimal design variable configurations, depending on several factors including the problem complexity, selected RL algorithm and hyperparameters tuning. One advantage of this approach is that the agent-simulator interaction significantly reduces the vast search space of design variables, leading to an accelerated and optimized design process. This is a generic causal approach that enables the exploration of new process configurations and provides actionable insights to designers to improve not only the process design but also the design process across various applications. The approach was validated on industrial processes including an absorption-based carbon capture, considering the economic and technological uncertainties of different capture processes, such as energy price, production cost, and storage capacity. It achieved a cost reduction of up to 5.5% for the designed capture process, after a few iterations, while also providing the designer with actionable insights. From a broader perspective, the proposed approach paves the way for accelerating the adoption of decarbonization technologies (CCUS value chains, clean fuel production, etc.) at a larger scale, thus catalyzing climate change mitigation.

Keywords: Carbon Capture, Process Design, Optimization, Artificial Intelligence, Reinforcement Learning, Machine Learning, Simulation-based Optimization

INTRODUCTION

Climate change has become a crucial problem, where massive efforts are underway to mitigate its disastrous outcomes [Ref. Paris Agreement]. In this regard, there is a major trend to develop efficient carbon capture technologies, recognized as one of the promising solutions to the challenges posed by climate change. Furthermore, the research is accelerating in this hot field due to the clear and prioritized need for developing cost-efficient and environmentally friendly carbon capture processes. Various methods based on mechanistic models and combined with traditional operation research methods, including linear and non-linear programming, are used to optimize existing carbon capture processes. Nevertheless, despite their robustness, most existing methods for optimizing process design suffer from certain limitations that hinder the development and optimization of carbon capture systems. These limitations include time consumption, high computational costs, and a lack of adaptability and transferability.

Therefore, there is a pressing need for the development of new methodologies to help overcome the abovementioned limitations. Artificial intelligence (AI) and its related machine learning (ML) technologies holds significant promise to address this critical gap. Specifically, reinforcement learning (RL), as a prescriptive machine learning technique, offers an effective approach due to

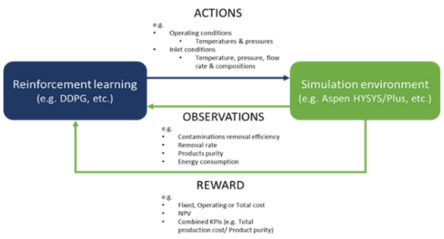

Figure 1. General simulation-based optimization approach.

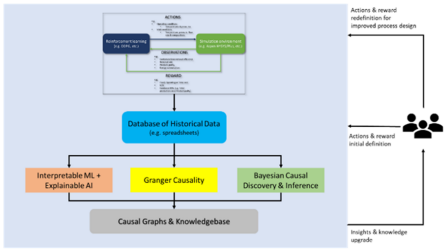

Figure 2. Summary of Causality Analysis Methodology

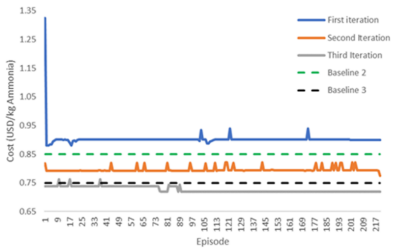

Figure 4. Optimization results (cost reduction along iterations and episodes) for ammonia-water system.

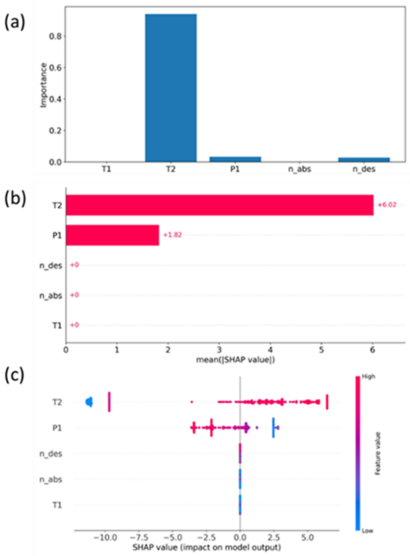

Figure 3. Results of IML+XAI (Dependency graphs) for ammonia-water system.

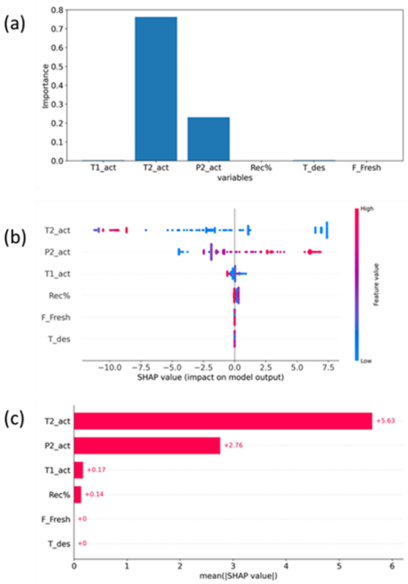

Figure 5. Results of IML+XAI (Dependency graphs) for CO_2-MEA system.

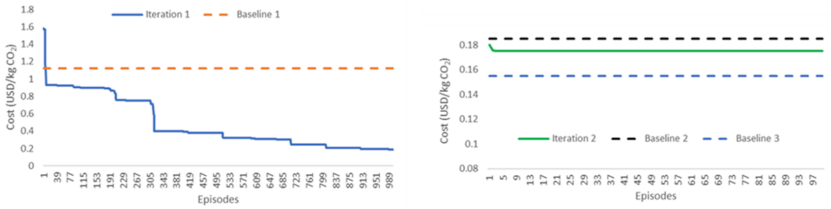

Figure 6. Optimization results for CO_2-MEA system.

being a learnable optimizer. In addition, its application has been successful in various fields, including process control [1], [2] which underscores its potential for accelerating chemical process design. RL encompasses various types, ranging from model-free to model-based algorithms, as well as policy-based and value-based agents. Examples of policy-based RL agents include deep Q-network (DQN) [3] and proximal policy optimization (PPO) [4]. Besides, there are agents that combine both policy-based and value-based calculations, such as deep deterministic policy gradient (DDPG) [5]. These diverse types offer versatile tools for enhancing chemical process design acceleration.

Recently, several research studies have employed RL within the realm of chemical engineering to optimize process sequences and to suggest new process flow diagrams for simple processes [6]-[12]. For instance, in [10], the authors used a hierarchical deep RL (DRL) agent to automate the synthesis of flowsheets for the ethyl tert-butyl ether (ETBE) production process. The study adopted the *SynGameZero* approach outlined in [13]. The Soft Actor Critic (SAC) algorithm was used in [9] to design a distillation train to separate a mixture of benzene, toluene, and *P*-xylene. The SAC agent interacts with a Distillation Gym environment that interfaces with *COCO* open-source simulator. The paper [14] conducted a survey on the RL-assisted flow sheet synthesis, highlighting the interaction with other simulators, i.e. DWSIM, and the concept of transfer learning for process design [8]. Nonetheless, some of these attempts are still at early stages as the agents developed suggest technically infeasible flowsheets in some cases due to the lack of reasoning for the trained agent. Besides, one major limitation of these studies is the absence of causality. Analyzing the historical data gathered during the optimization process can offer new insights to the designers, enabling them to identify the causal variables that impact selected key performance indicators (KPIs), e.g. costs. Extracting knowledge from causality analysis part will lead to more generalizable results and provide actionable insights into hidden information and relationships that can be transferred to enhance other processes.

The novelty of the current work lies in the introduction of causal DRL to accelerate the design of chemical

processes. Two systems were considered for illustration: 1) ammonia-water system (hypothetical physical absorption-desorption process as a simplified example) and 2) mono-ethanolamine (MEA)-based carbon capture process. Introducing the concept of causal reinforcement learning to process design, specifically in applications related to solvent-based capture processes, can enhance the agent's reasoning capabilities, and enrich the knowledge of design experts with new insights. This approach can be generalized to other systems, under the full supervision of human experts, to streamline and accelerate the design process. The specific research objectives of this study are outlined as follows: 1) Develop an incremental DRL (IDRL) simulation-based optimization methodology to accelerate the design of steady state processes; 2) Integrate a newly developed causality analysis methodology, employing various algorithms with the IRDL to capture the underlying cause-effect relationships among different process variables and the specified KPI(s); 3) Deploy the developed IDRL coupled with causality analysis methodology to optimize different processes, including absorption-based carbon capture process utilizing amine solvents.

METHODOLOGY

General DRL Simulation-based Optimization Approach

The general simulation-based optimization approach and its underlying methodology achieving its objectives is depicted in **Figure 1.** A DRL agent is linked with a simulation environment in an interactive and learnable manner that minimizes the search space. Within this approach, the term "rewards" is used to denote the key performance indicators, whereas the term "states" is used to represent the observations defining the current condition of the process. The agent provides ”actions” (representing design variables including equipment sizing and operating conditions) to the simulation environment based on the current state and the observed reward. The agent tries to find the optimal point without sticking in local minima where it may follow the shortest path, depending on factors such as problem complexity, reward formulation, the RL algorithm selected, and available computational resources. Unlike most of the traditional optimizers, DRL-based optimization is driven by the concept of reward and penalty. The agent receives a reward when following the right path of optimizing the objective function, i.e. minimizing the process costs, or minimizing its environmental impact. Conversely, a penalty is incurred if the agent followed the wrong direction, where the actions result in cost or environmental impacts exceeding specified values. To avoid settling in local minima, the agent posses the ability to explore new scenarios, aiming to identify paths leading to the global minimum point. Balancing between exploitation and exploration is crucial to speed up achieving the optimal solution. Additionally, achieving an accelerated optimization requires proper formulation of reward function and optimization of the agent's hyperparameters.

DRL and problem formulation as Markov Decision Process. From a mathematical standpoint, when employing RL, the problem is typically formulated as a *Markov* Decision Process (MDP). This formulation involves a set of observations that characterize the state of the environment, along with a distribution of initial states. Additionally, it entails a set of actions, whether continuous or discrete, that can modify the state of the environment based on rewards received by the agent. An MDP comprises transition dynamics or probabilities, which describe the relation between states and actions. Two other important components are the immediate rewards and the discount factor (which ranges from 0 to 1), a lower discount factor places greater emphasis on the immediate reward. For a more detailed explanation of the mathematical formulation of RL and MDP, refer to this study [3]. It is worth noting that one of the primary differences between dynamic and steady-state problems lies in the definition of discount factor during MDP formulation. The discount factor is an optional element in defining an MDP and can be disregarded (assigned a zero value) when considering the optimization of steady-state processes, where immediate rewards take precedence over long-term ones. Disregarding the discount factor in case of steady-state processes facilitates the design optimization problem definition/formulation. Nevertheless, finding the optimal policy based on the cumulative reward for the RL agent remains relevant. This design problem adheres to the *Markov* Property, resulting in a memoryless agent, where the current state depends only on the previous state and influences the selection of the subsequent actions that modify the successive states. Consequently, the transition dynamics or probabilities over time and the whole states history will not be taken into consideration. The transition probabilities between two successive states based on the proposed or taken actions are primarily deterministic. In chemical process design, these transitions may involve thermodynamic/equilibrium relations, kinetic/rate equations, or heuristic rules.

RL Algorithm selection & development of the simulation environment. Various RL algorithms can be developed, optimized, and then utilized for process design optimization. As previously mentioned, the PPO and SAC algorithms were employed process flowsheet synthesis. In this work, we consider the DDPG algorithm [5], which is one of the promising algorithms. It combines the value-based and policy-based optimization principles.

Upon developing the simulation environment, the potential actions (design variables) under consideration include equipment sizing and process operating

conditions. Additionally, parameters such as recycle rate, purge percentage and flowrates of inlet streams are deemed significant actions to be prescribed or optimized. Equipment sizing varies according to the equipment type. For instance, the number of stages is a crucial variable for separation columns, defining their heights and costs. Continuous stirred reactors (CSTRs) are sized by volume, which is also related to the percentage conversion of limiting reactants. Heat exchange equipment sizing is determined by the heat exchange area, which is also related to other process operating conditions. Process operating conditions encompass different streams/equipment temperatures and pressures. Key observations describing the simulation environment's state include the purities of specified product streams/components, recovery rates of components of interest, and utilities consumption. These states and actions vary from a problem to another depending on the specific requirements of the problem and the specifications set by the process expert.

Reward definition (formulation). The reward function can take various forms, depending on the specified objective defined by the process expert. For instance, it can be formulated as the relative production cost, i.e. production cost per kg of product or material. Alternatively, it could represent the total capital investment (TCI), net present value (NPV) or any other form of economic KPIs. Furthermore, it might be a combined value of both techno-economic KPIs such as relative production cost per material/product recovery or purity. One objective function that can serve as a highly relevant reward function is the total environmental impact or the relative emissions per kg of the product. This function holds significant importance in current and future research studies for process design optimization, enabling the assessment of processes' environmental footprints and promoting the development of eco-friendly processes.

Incremental learning. It is worth noting that our approach for the DRL-based optimization is built upon the concept of incremental learning. Specified costs, environmental impacts and actions ranges change from one case to another until the maximum allowable performance is achieved, i.e. lowest possible relative production cost and minimum acceptable environmental impact. More specifically, the actions ranges are narrowed or redefined after analyzing the historical data collected from previous RL-assisted optimization cases. In addition, reward function and its value are redefined based on this analysis. For example, the specified cost value may be reduced to explore the agent's ability to surpass it and reach the minimum allowable cost.

Towards causality analysis for causal and rational DRL

Following the incremental learning-based optimization, historical data representing the simulation runs resulting from the agent-simulation interactions are stored in spreadsheets for the causal analysis, where each run represents a data point (observation). This is to gain more actionable insights into the root causes and the importance of design variables impacting the KPIs. Three methods were employed to perform a preliminary causality analysis based on observations. These methods include combined interpretable ML (IML)-Explainable AI (XAI), Granger causality (GC), and Bayesian causal discovery & inference.

Combined IML-XAI. Initially, the decision tree classification algorithm (i.e. IML) is combined with XGBoost and SHAP explanation algorithm (i.e. XAI) to extract the patterns, assess variables importance and analyze the relationships governing the dependence of the specified KPI (reward) and observations (state) on the design variables (actions). The IML extracts the patterns (ranges of design variables) that lead to processes with reasonable feasibility. Additionally, it assesses the variables importance, determining the most significant ones. The XAI gives relative magnitudes and directions of the effect of each input variable on the final KPI (process overall production cost). This will help afterwards in the incremental learning procedure, helping to select/redefine the relevant input variables ranges for guided optimization.

Granger causality (GC). This method is used preliminary to validate the relations between the actions (especially the most significant ones identified by the combined IML-XAI method) and the KPI. Since actions are provided continuously by agent and they depend on the previous set of actions and states, consequently actions, states, and rewards/KPIs are treated as a form of time series data, leading to the application of GC. As mentioned earlier, in case of optimizing steady-state processes, the current state (representing the current process configuration) is solely influenced by the preceding one, rendering longer-term relationships negligible. The dependence between each successive pair of states (S_t & S_{t+1}) and their associated set of actions based on the reward provided to the agent, justifies the utilization of approximated time series assumption to apply GC.

Bayesian causal discovery & inference. This method is used to generate causal graphs to discover the root causes that significantly affect changes of the KPI. It also reveals the interdependence among actions, discovering the root cause design variable that affects both other design variables and KPI. **Figure 2** summarizes the proposed causality analysis methodology. Further details regarding the combined IML-XAI method can be found in our recent study [15]. The insights gained from this causality analysis will upgrade/enrich the experts' knowledge, where they can redefine the baselines/rewards and the range of actions to retrain the agent, thereby fostering further improvement in design optimization.

This methodology can be applied in flowsheet synthesis to enhance the DRL agent's reasoning capability by providing it with the extracted knowledge and causal graphs. This empowers the agent to operate as a rational causal DRL agent that avoids the synthesis of technically infeasible flowsheets/processes, thus accelerating the synthesis and design process. In addition to the other design variables, equipment and their sequencing will be added to the list of actions. Consequently, the causal graphs will account for these two crucial categorical variables and quantify their impact on the defined KPI/reward. Furthermore, integrating domain knowledge from process designers and operators as constraints into the causal DRL agent's structure will enhance its rationality. These domain knowledge-informed constraints ensure the avoidance of invalid streams/equipment selection, placement, and sequencing, e.g. separation train sequencing [16] and chemical reactors sequencing [17]. This will further shorten the time needed for the agents to converge on the optimal new process, thus ensuring the desired ML-assisted design acceleration.

SIMULATION CASE STUDIES

Aspen HYSYS was selected as the simulation software for constructing the process flow diagrams and conducting the process design. Two simulation case studies were developed to validate the proposed CIRL approach. The simplified process flow diagrams of the two case studies are presented in the supplementary materials and their details are summarized in the following subsections.

Case study 1: Ammonia water system

An ammonia-water absorption-desorption system was developed with a stream of air contaminated with ammonia is introduced to the absorption column, where pure water is utilized as a physical solvent to separate ammonia from air. The captured ammonia is then removed from the water solvent using steam in a re-boiled stripping column. Heat exchange between the stream entering the stripping column and its bottoms product stream enables heat integration, minimizing the reboiler heat load. A condensation and a flash separation are applied for the top product stream of stripping column to remove any water associated with the separated ammonia, thus enhancing its purity and recovering any escaped solvent. Recycle was not considered in this simplified case study. The NRTL simulation fluid package was used [18]. The actions considered include number of stages in both absorption and stripping columns, temperatures of fresh solvent and inlet contaminated gaseous stream, and the pressure of the inlet contaminated gaseous stream.

Case study 2: Mono-ethanolamine (MEA)-based carbon capture system

This case considers chemical absorption (CO_2-MEA system), where flue gas enters the absorption column after solids removal, pressurization, and cooling to the desired operating temperature. The absorption-desorption setup is like the case of ammonia (case study 1). However, in this case, lean MEA is recirculated, after carbon dioxide stripping, to the absorption column and mixed with a makeup stream to enhance the process feasibility. A purge stream is used to avoid excessive accumulation of carbon dioxide in the system and the recycled lean MEA stream. Effect of temperature and pressure of inlet flue gas stream and temperature of makeup solvent were considered as actions. Effect of recycle percentage and temperature of stream entering the stripping column were also considered for optimization. The *Acid Gas - Chemical Solvents* Aspen built-in fluid package was used to perform the simulation calculations. The stream of flue gases comes from a power plant, where all conditions defined according to that study [19]. A baseline/reference for the carbon dioxide recovery cost of 1.12 USD/kg was adopted from a published study [20] that ultimately achieved the lowest cost after several iterations.

Reward definition

Important observations that describe the state of simulation are purity of purified gas stream, recovery of contaminants (for ammonia and carbon dioxide), purity of the product stream, and utilities consumption. The reward function is defined as the relative removal cost, i.e. removal cost per kg of captured material. The procedure for calculating the removal cost and other related costs follows the guidelines outlined in this textbook [21] and these two studies [22], [23]. Utilities and MEA prices were sourced from a study published in [24]. The reward was determined as follows: if the agent progresses in the specified direction (new cost <= specified cost), it receives a value equivalent to the total production cost (unity factor); otherwise, it receives a zero value. Furthermore, if the new cost in current step is lower than the cost of previous step, the agent receives the full total production cost value; otherwise, it receives zero.

RESULTS AND DISCUSSION

Optimized DRL agent

The agent developed and adopted in this study is based on DDPG, a policy-based model-free reinforcement learning algorithm. It comprises two neural networks, one for the actor and one for the critic. For a deeper understanding of the mathematical underpinnings of the DDPG, interested readers can refer to [1], [25], [26]. The structure and crucial hyperparameters of

the optimized DDPG agent are summarized in **Table 1**.

Table 1: DDPG optimized hyperparameters.

Hyperparameter	Value
Action noise standard deviation	0.2
Actor learning rate	0.001
Critic learning rate	0.002
Discount factor (gamma)	0
Tau parameter used to update target networks	0.005
Total number of episodes	100-1000
Activation function for the critic neural network dense layers	ReLU
Activation function for the actor neural network inputs dense layers	ReLU
Activation function for the actor neural network outputs dense layer	Tanh

The *Tanh* activation function was selected to standardize the range of actions between -1 to 1, which are then converted back to their real values before being used in the simulation.

Process design optimization

Ammonia-water system

For the ammonia-water absorption system, the initial range for both inlet solvent and flue gas streams temperatures was set between 20 to 100 °C, with pressure ranging from 600 to 950 kPa. It is important to mention that the absorption column pressure was set equal to the inlet pressure of flue gas. This simplification facilitates the manipulation of the absorber pressure in the simulation environment by the RL agent. It is important to note that the inlet flue gas pressure must be higher than that of the absorber to ensure propre flow inside the column. Hence, a pressure drop value will be added to specify the pressure of inlet flue gas to the designer to incorporate it in real life operations. Additionally, the range of number of stages for both absorption and desorption columns was from 3 to 15. The initial desired cost (baseline) had the value of 2 USD/kg of ammonia based on a simulation of base case. As a result, after 220 episodes, a cost of almost 0.880 USD/kg of ammonia was obtained initially, which is significantly lower than the specified baseline. Accordingly, following the proposed incremental learning method, this value was given as a new baseline to the agent for further improvement of the process economic optimization, where the ranges of variables were the same. Surprisingly, the agent could reach another lower production/recovery cost of 0.775 USD/kg of ammonia. The preliminary causality analysis methodology was applied to extract some other knowledge about the most effective variables and the patterns to achieve lower process costs. Thus, the historical data from these two first optimization iterations was collected, preprocessed, and analyzed. The preprocessing was done through dividing the observations to high cost (above 0.9 USD/kg of ammonia) observations and labeled as Class (0). The other portion was labeled as Class (1), which represented the low cost (below 0.9 USD/kg of ammonia) observations. **Figure 3** represents the results of employing IML+XAI to get the dependency graphs.

As shown, the most effective two variables are T2 (temperature of flue gases entering the absorption column) and the pressure of the solvent stream. These results were validated through performing the GC analysis to test the dependency of the continuous cost change corresponding to the continuous actions given by the agent, in which the new actions depend on the previous actions, state and reward. *p*-value of the dependence between T2 and cost was < 0.001 which indicates high GC. Besides, according the XAI dependency graphs, higher values of T2 lead to lower costs (Class (1)). The variable P1 possesses clearly lower significance, where, after data preprocessing/analysis, the lower values in the range of 830-900 kPa (around 850 kPa) lead to lower costs (Class (1)). These results of the current case mean that the operating costs dominate the removal cost. In addition, upon analyzing the available data, to achieve higher purities and recoveries of ammonia, it is preferred to utilize absorption columns with low number of stages and stripping column with a higher number of stages. The causal incremental learning methodology aims at minimizing the search space to accelerate the process design optimization. Hence, after obtaining the dependency graphs, the ranges of the variables were redefined and narrowed to see if further improvement can be obtained. The new ranges are (85-99 °C) for T2, (25-41 °C) for T1, (800-900 kPa) for P1, (3-5) for absorption number of stages and (10-15) for stripping column number of stages. In addition, a new baseline was set to 0.75 USD/kg of ammonia. As a result, a new lower production cost of 0.719 USD/kg of ammonia was obtained at the new conditions of T1 = 40.5 °C, T2 = 99 °C, P1 = 900 kPa, absorption number of stages = 4.8 (almost 5) and stripping column number of stages = 15. At these conditions, the purity of outlet air and ammonia recovery were 97.2% and 96.3%, respectively. **Figure 4** summarises the results of the proposed causal incremental learning methodology for process design optimization. The Bayesian causal graph showed that the temperature of the flue gas stream is one of the root causes of changing the KPI (removal cost) and can even affect other design variables (absorption column stages (n_abs) & stripping column stages (n_des)) and state. This graph supports the results of the first two methods and is provided in the supplementary materials.

MEA-based Carbon Capture system

In the case of MEA-based carbon capture system, the first ranges, according to literature [19], were inlet solvent and flue gas temperatures (T1 and T2) of 30-90 °C and inlet flue gas pressure (P1) of 150-500 kPa. Besides, recycle percentage ranged initially between 90% and 98%, while the range of temperatures of rich MEA stream entering the stripping column was 77-85 °C. As in the case of ammonia-water system, the same assumption of considering the absorption column pressure as the inlet flue gas pressure is still applicable. After using the first baseline of 1.12 USD/kg CO_2, this initial optimization led to a lower production cost of almost 0.190 USD/kg CO_2. The collected data was then preprocessed and labeled to the IML-XAI based low level causality analysis of the historical simulation data collected from this iteration show that the most significant variables are flue gas inlet temperature and its pressure. The GC analysis confirmed this where the *p*-value in the case of cost dependence on T2 did not exceed 0.0001. It is worth mentioning that when the test was done in the reverse direction the *p*-value exceeded 0.9; this means that T2 causes the change in the cost but not the vice versa. In addition, the *p*-values in the case of the other variables T1, Rec%, T before stripping column and the fresh solvent flowrate exceeded 0.1, which indicates low to no significance of these variables compared to T2 in the current study within the specified range. Figure 5 shows the results of the combined IML+XAI approach in the current case of MEA-based carbon capture system. According to these results the ranges of the variables were narrowed/redefined for the sake of exploration for further optimization through retraining the agent considering the extracted causal knowledge, which gives it more reasoning. This reasoning/causality accelerates the retraining step and the subsequent design optimization. In this regard, the ranges of the actions were narrowed/redefined to become inlet solvent and flue gas temperatures (T1 & T2) of 23-35 °C and inlet gas pressure (P1) of 250-400 kPa in the second iteration. In addition, the new range of the recycle percentage was 85%-95% and accordingly the range of temperatures of rich amine stream entering the stripping column changed to 71-81 °C. Furthermore, the baseline was decreased to 0.185 USD/kg CO_2. A third baseline, based on the expertise of carbon capture and process engineering experts, was put at 0.155 USD/kg CO_2 for comparison and more improvement in future research. Figure 6 summarizes the results of the two training iterations with their corresponding baselines. as shown, the minimum relative removal cost obtained was about 0.175 USD/kg CO_2. This cost was obtained after 25 episodes during the second iteration (agent retraining step). The conditions that achieved this value were T1 of 24 °C, T2 of 25 °C, P2 of 270 kPa, recycle percentage of 85 % and temperature of the rich amine entering the desorption column of 71 °C, respectively. Additionally, the carbon capture percentage at these conditions was 92.5% with a carbon dioxide purity of 98%. The causal graph based on Bayesian framework showed that one of clear the root causes of changing the KPI and can even affect other design variables and state is the temperature of the flue gas stream in agreement with what was obtained by the first two methods. This graph is put as supplementary material.

RECOMMENDATIONS FOR FUTURE WORK

In forthcoming research, we aim to incorporate emissions and other environmental impacts as crucial optimization objectives. In particular, the net emissions should be considered as a reward, where the emissions of the capture process itself should be minimized as possible. We advocate for exploring the application of transformers-based RL for flowsheet synthesis and process design acceleration. This will allow adding more reasoning to the causal incremental learning-based RL agent to help in generating/suggesting new technically feasible process flow diagram. This will be done also through defining constraints based on process design experts' domain knowledge and rules as a practical application of the interesting concept of RL with human preferences. This will guarantee the acceleration of process design under the full supervision of human experts to always insure simultaneous technical and economic feasibility of the generated flowsheets. Uncertainty of inlet streams composition and energy/fuel prices will be considered in our future research as well which is one of the advantages of RL to deal with uncertainties.

CONCLUSION

This study introduced and employed a causal Incremental Reinforcement Learning (CIRL) agent as a learnable optimizer with reasoning capabilities to accelerate efficient process design. Two case studies were considered to validate the approach, physical absorption (ammonia-water system) and chemical absorption (CO_2-MEA system). In the case of chemical absorption, the causal DRL agent reached an optimal cost after only 25 episodes. The causality analysis revealed that key variables, such as inlet flue gas temperature and absorption pressure, significantly influence the CO_2-MEA system's production cost. For instance, it showed that the relative production cost of the process is inversely proportional to the flue gas temperature fed to the absorber. Feeding the agent with the extracted knowledge and causal information obtained from the causality analysis helped it to achieve more improved process design. These findings hold relevance beyond the studied systems and offer actionable insights for future absorption system designs.

Supplementary Materials are available on LAPSE at https://psecommunity.org/LAPSE:2024.0276

REFERENCES

1. K. Nadim et al., "Learn-to-supervise: Causal reinforcement learning for high-level control in industrial processes," Eng. Appl. Artif. Intell., vol. 126, p. 106853, (2023)
2. P. Daoutidis et al., "Machine learning in process systems engineering: Challenges and opportunities," Comput. Chem. Eng., vol. 181, p. 108523, (2024)
3. K. Arulkumaran et al., "Deep reinforcement learning: A brief survey," IEEE Signal Process. Mag., vol. 34, no. 6, pp. 26–38, (2017).
4. C. Yu et al., "The surprising effectiveness of ppo in cooperative multi-agent games," Adv. Neural Inf. Process. Syst., vol. 35, pp. 24611–24624, (2022)
5. T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," arXiv Prepr. arXiv1509.02971, (2015)
6. L. Stops et al., "Flowsheet generation through hierarchical reinforcement learning and graph neural networks," AIChE J., vol. 69, no. 1, p. e17938, (2023)
7. Q. Göttl et al., "Deep reinforcement learning uncovers processes for separating azeotropic mixtures without prior knowledge," arXiv Prepr. arXiv2310.06415, (2023)
8. Q. Gao et al., "Transfer learning for process design with reinforcement learning," arXiv Prepr. arXiv2302.03375, (2023)
9. L. I. Midgley, "Deep reinforcement learning for process synthesis," arXiv Prepr. arXiv2009.13265, (2020)
10. Q. Göttl et al., "Automated flowsheet synthesis using hierarchical reinforcement learning: proof of concept," Chemie Ing. Tech., vol. 93, no. 12, pp. 2010–2018, (2021)
11. S. J. Plathottam et al., "Solvent extraction process design using deep reinforcement learning," J. Adv. Manuf. Process., vol. 3, no. 2, p. e10079, (2021)
12. A. A. Khan and A. A. Lapkin, "Designing the process designer: Hierarchical reinforcement learning for optimisation-based process design," Chem. Eng. Process. Intensif., vol. 180, p. 108885, (2022)
13. Q. Göttl et al., "Automated synthesis of steady-state continuous processes using reinforcement learning," Front. Chem. Sci. Eng., pp. 1–15, (2021)
14. Q. Gao and A. M. Schweidtmann, "Deep reinforcement learning for process design: Review and perspective," arXiv Prepr. arXiv2308.07822, (2023)
15. E. G. Al-Sakkari et al., "Machine learning-assisted selection of adsorption-based carbon dioxide capture materials," J. Environ. Chem. Eng., p. 110732, (2023)
16. E. Marcoulaki et al., "Design of separation trains and reaction-separation networks using stochastic optimization methods," Chem. Eng. Res. Des., vol. 79, no. 1, pp. 25–32, (2001)
17. R. Chebbi, "Optimizing reactors selection and sequencing: minimum cost versus minimum volume," Chinese J. Chem. Eng., vol. 22, no. 6, pp. 651–656, (2014)
18. Simoni, L. D., Lin, Y., Brennecke, J. F., & Stadtherr, M. A. Modeling liquid– liquid equilibrium of ionic liquid systems with NRTL, electrolyte-NRTL, and UNIQUAC. Industrial & engineering chemistry research, 47(1), 256-272. (2008)
19. L. E. Øi, "Aspen HYSYS simulation of CO2 removal by amine absorption from a gas-based power plant," in The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), pp. 73–81 (2007)
20. J. Chen and F. Wang, "Cost reduction of CO2 capture processes using reinforcement learning based iterative design: A pilot-scale absorption--stripping system," Sep. Purif. Technol., vol. 122, pp. 149–158, (2014)
21. M. S. Peters et al., Plant design and economics for chemical engineers, vol. 4. McGraw-hill New York, (1968)
22. E. G. Al-Sakkari et al., "Comparative Technoeconomic Analysis of Using Waste and Virgin Cooking Oils for Biodiesel Production," Front. Energy Res., p. 278, (2020)
23. M. M. Naeem et al., "Single-stage waste oil conversion into biodiesel via sonication over bio-based bifunctional catalyst: optimization, preliminary techno-economic and environmental analysis," Fuel, vol. 341, p. 127587, (2023)
24. N. Wang et al., "MEA-based CO2 capture: a study focuses on MEA concentrations and process parameters," Front. Energy Res., vol. 11, (2023)
25. D. Mehta, "State-of-the-art reinforcement learning algorithms," Int. J. Eng. Res. Technol., vol. 8, pp. 717–722, (2020)
26. Y. Hou et al., "A novel DDPG method with prioritized experience replay," in 2017 IEEE international conference on systems, man, and cybernetics (SMC), pp. 316–321 (2017)