Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Reaction Pathway Optimization Using Reinforcement Learning in Steam Methane Reforming and Associated Parallel Reactions

Martín Rodríguez-Fragoso[a], Octavio Elizalde-Solis[a], and Edgar Ramírez-Jiménez[a]*

[a] Department of Chemical Petroleum Engineering, ESIQIE, Instituto Politécnico Nacional, Mexico City, 07738, Mexico
* Corresponding Author: eramirezj@ipn.mx.

ABSTRACT

This study presents the application of a Q-learning algorithm to optimize the selection of chemical reactions for methane reforming processes. Starting with a set of 11 candidate reactions, the algorithm identified three key reactions. These reactions effectively represent the experimental data while aligning with the underlying physics of the process and previously reported findings. The algorithm employed an epsilon-greedy policy to balance exploration and exploitation during the training process. Furthermore, simulations based on the identified reactions revealed trends consistent with experimental data. This work highlights the efficiency and adaptability of Q-learning in modeling complex catalytic systems and provides a framework for further exploration and optimization of methane reforming processes.

Keywords: Methane Reforming, Machine Learning, Optimization, Reaction Engineering, Reinforce Learning

INTRODUCTION

Methane reforming is one of the key technologies for hydrogen production, an essential energy carrier in the transition toward cleaner energy systems. This process involves a series of simultaneous and coupled chemical reactions, among which Methane Steam Reforming (MSR) and the Water Gas Shift (WGS) reaction are particularly noteworthy. However, the complex and highly nonlinear nature of these reactions poses a significant challenge for process modeling and optimization.

Traditionally, kinetic models used to describe such processes are based on a fixed set of reaction rate equations derived from the Arrhenius equation. While this approach has proven effective in many cases, it can be limited when dealing with systems where operating conditions or predominant reaction pathways change significantly, such as in catalytic reforming in refineries, hydrodesulfurization, and fluid catalytic cracking (FCC). In this context, there is a need for tools capable of dynamically identifying the most relevant combinations of reactions to represent experimental data [1].

Reinforcement learning algorithms, such as Q-learning, offer a novel and powerful approach to tackle this problem [2]. These techniques enable the exploration and exploitation of potential solution spaces, iteratively learning optimal decisions based on a defined reward function. For methane reforming, the use of Q-learning facilitates the selection of reactions that best describe experimental data under varying temperature conditions, while simultaneously optimizing the associated kinetic constants.

This work describes the implementation of a Q-learning algorithm to identify the key reactions within an initial set of 11 reactions considered relevant for methane reforming. Based on this selection, the algorithm fits the experimental data using a model grounded in the Arrhenius equation.

This methodology not only contributes to a better understanding of the most significant reaction pathways but also opens new possibilities for optimizing the design and operation of complex catalytic processes.

METHODOLOGY

Q-Learning Framework

To optimize reaction pathways, a framework based on

reinforcement learning is adopted, utilizing a Q-learning process, as illustrated in Fig. 1. A ϵ-greedy policy is defined due to its ability to balance exploration and exploitation. This approach allows the agent to explore new actions with a probability controlled by the epsilon (ϵ) parameter while prioritizing actions with known rewards as learning progresses. This helps avoid suboptimal solutions that could arise from premature exploitation of the environment.

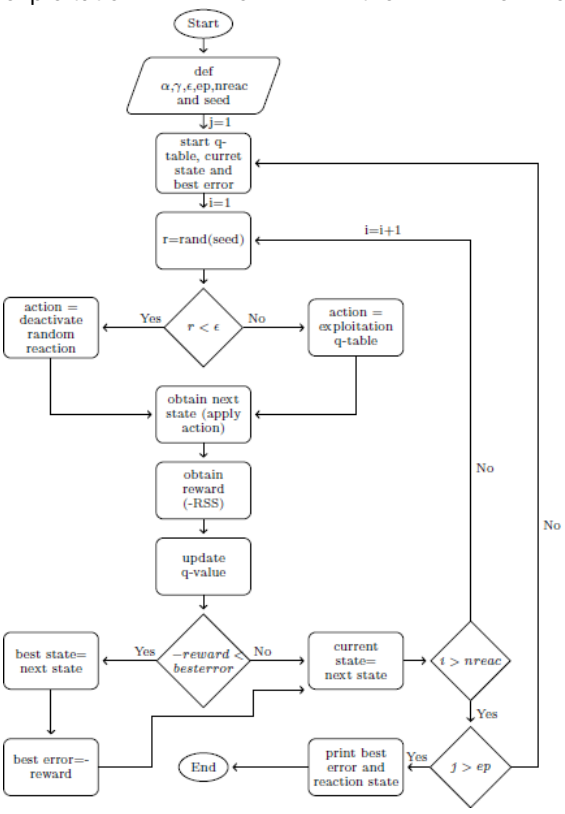

Figure 1. Proposed Methodology to select reactions.

This method relies on the iterative updating of Q values using the equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \quad (1)$$

Based on this equation, the parameters α, γ, and ϵ must be defined considering the following criteria:

The learning rate (α) controls how much the $Q(s,a)$ values are updated in each iteration. A low α ensures that updates are gradual, reducing the risk of abrupt oscillations in $Q(s,a)$. However, this may slow down convergence. This parameter directly affects the adjustment term:

$$\alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \quad (2)$$

The parameter ϵ defines the exploration probability. At the beginning of the learning process, a high ϵ encourages exploration of the state space to avoid suboptimal local solutions. During training, the value of ϵ is dynamically adjusted according to:

$$\epsilon_{i+1} = \begin{cases} \epsilon_i * 0.99, & i < 10 \\ 0.1, & i \geq 10 \end{cases} \quad (3)$$

This adjustment allows for a controlled transition from an intensive exploration phase to more consistent exploitation in the final stages of training.

The discount factor (γ) determines the weight of future rewards. A high γ prioritizes long-term rewards, favoring globally optimal strategies. This parameter affects the term:

$$\alpha\gamma \max_{a'} Q(s',a') \quad (4)$$

The values of the parameters used in this work for the Q-learning algorithm are summarized in Table 1.

Table 1: Q-learning Selected Parameters.

Parameter	Value
Episodes	30.00
α	0.10
ϵ_{in}	0.95
γ	0.90

This parameter configuration is designed to achieve a balance between learning stability, initial exploration, and efficient exploitation in advanced stages.

Experimental Data

The experimental data used in this work were reported by Akers et al. [3] This dataset consists of five runs. The first four runs were conducted at the same temperature, varying the methane-to-water feed ratio. The experiments were performed at different reaction times. The reaction time for methane is directly proportional to the amount of catalyst and inversely proportional to the methane feed rate. This reaction time is expressed as a time factor, defined as the weight of the catalyst divided by the natural gas feed rate, expressed in moles per second.

Table 2: Experimental Sets Conditions

Set	T [K]	Input molar ratio $\left(\frac{H_2O}{CH_4}\right)$	W/F$_{max}$ $\left[\frac{kg \cdot s}{mol}\right]$
1	910	10:1	1.10× 10^{-3}
2	910	5:1	7.00× 10^{-4}
3	910	3.5:1	7.00× 10^{-4}
4	910	2.5:1	7.00× 10^{-4}
5	609	3:1	3.00× 10^{-4}
6	671	3:1	3.00× 10^{-4}
7	723	3:1	3.00× 10^{-4}
8	779	3:1	3.00× 10^{-4}

The fifth run aims to highlight the effect of temperature. For this purpose, the feed ratio and reaction time

were kept constant, while the temperature varied. A summary of the experimental data is presented in Table 2.

Since the catalyst was not used for more than 5 hours, its activity is assumed to remain constant for all runs conducted at the same temperature.

A total of 188 experimental data points is available, corresponding to the partial pressures of the components (CH_4, H_2O, CO_2, H_2, and CO). While this may appear to be a relatively small dataset, the use of first-principles models in combination with Q-learning algorithms is expected to yield optimal results. Moreover, the use of experimental data reduces potential errors associated with simulated data points and ensures that the selection of reaction pathways is not biased by the predefined conditions of simulated datasets.

Kinetics Model

In the methane reforming process, a detailed kinetic model is essential to understand the reaction dynamics. This model considers a network of 11 reactions

The equilibrium constant for each reaction is determined as a function of temperature. Specifically, for each reaction j, the equilibrium constant is expressed as:

$$K_{\mathrm{eq},j} = A_j \cdot \exp\left(-\frac{\Delta H_j^{\circ}}{RT}\right) \tag{5}$$

where A_j is a pre-exponential factor related to the standard entropy change (ΔS_j°), ΔH_j° is the standard enthalpy change, R is the universal gas constant, and T is the temperature in Kelvin. The data for the equilibrium constants used in this model are compiled in Table 3, which includes values reported by various authors [3-8].

The reaction rate for each reversible reaction j is given by:

$$r_j = k_{f,j} \prod_i P_i^{\nu_{i,j}} - k_{r,j} \prod_i P_i^{\nu'_{i,j}} \tag{6}$$

where $k_{f,j}$ and $k_{r,j}$ are the forward and reverse rate coefficients, respectively, and P_i is the partial pressure of component i. The term $\nu_{i,j}$ represents the stoichiometric coefficients of the reactants, while $\nu'_{i,j}$ represents those of the products in reaction j. The forward rate constant $k_{f,j}$ is calculated using the Arrhenius equation:

$$k_{f,j} = A_{f,j} \cdot \exp\left(-\frac{E_{f,j}}{RT}\right) \tag{7}$$

where $A_{f,j}$ is the pre-exponential factor and $E_{f,j}$ is the activation energy for the forward reaction. These parameters $A_{f,j}$ and $E_{f,j}$, which are crucial for determining the forward reaction rates, are calculated through an optimization process, as described in the "Parameter Estimation" section of this work.

The reverse rate constant $k_{r,j}$ is related to the forward rate constant and the equilibrium constant $K_{\mathrm{eq},j}$ by:

$$k_{r,j} = \frac{k_{f,j}}{K_{\mathrm{eq},j}} \tag{8}$$

The rate law thus accounts for both the forward and reverse reactions, and the overall reaction rate depends on the partial pressures of the species involved in each reaction.

To maintain mass balance across the system, the rate of change of the partial pressure of any component i is described by a mass balance equation:

$$\frac{dP_i}{d\left(\frac{W}{F_{\mathrm{CH_4}}}\right)} = \sum_{j=1}^{11} \nu_{i,j} r_j \tag{9}$$

where W is the catalyst mass, $F_{\mathrm{CH_4}}$ is the molar flow rate of methane, and $\nu_{i,j}$ is the stoichiometric coefficient of component i in reaction j. The set of equations for all components is solved simultaneously to predict the evolution of the partial pressures over time and catalyst mass.

Additionally, a Q-learning algorithm is used to optimize the reaction set. This optimization process dynamically selects which reactions from the set of 11 should be included or excluded from the kinetic model. The goal is to identify the reactions that best represent the methane reforming phenomena, based on experimental or simulated data.

Parameter Estimation

Table 3: Reaction set and its Equilibrium Function [8].

I	Reaction	K_{pi}	Dimensions
1	$CH_4 + H_2O \leftrightarrow CO + 3H_2$	$1.1669 \times 10^{13} \exp(-26380/T)$	$(atm)^2$
2	$CO + H_2O \leftrightarrow CO_2 + H_2$	$1.767 \times 10^{-2} \exp(4400/T)$	$(atm)^0$
3	$CH_4 + 2H_2O \leftrightarrow CO_2 + 4H_2$	$2.0620 \times 10^{11} \exp(-22430/T)$	$(atm)^2$
4	$CH_4 + CO_2 \leftrightarrow 2CO + 2H_2$	$6.6038 \times 10^{14} \exp(-31230/T)$	$(atm)^2$
5	$CH_4 + 3CO_2 \leftrightarrow 4CO + 2H_2O$	$2.1136 \times 10^{18} \exp(-40030/T)$	$(atm)^2$
6	$CH_4 \leftrightarrow C + 2H_2$	$4.1066 \times 10^{5} \exp(-10614/T)$	atm
7	$2CO \leftrightarrow C + CO_2$	$5.8201 \times 10^{-10} \exp(20634/T)$	$(atm)^{-1}$
8	$CO + H_2 \leftrightarrow C + H_2O$	$3.215 \times 10^{-8} \exp(16318/T)$	$(atm)^{-1}$
9	$CO_2 + 2H_2 \leftrightarrow C + 2H_2O$	$1.7762 \times 10^{-6} \exp(12002/T)$	$(atm)^{-1}$
10	$CH_4 + 2CO \leftrightarrow 3C + 2H_2O$	$4.2455 \times 10^{-10} \exp(22022/T)$	$(atm)^{-1}$
11	$CH_4 + CO_2 \leftrightarrow 2C + 2H_2O$	$0.730 \exp(1388/T)$	$(atm)^0$

The parameter estimation process involves solving the system of differential equations that describe the methane reforming reactions and adjusting the model parameters to best fit the experimental data. In particular, the activation energy (E_a) and the pre-exponential factor (A) of the forward reaction rate constants were estimated through optimization. To solve this system, it was used the `odeint` function from the `scipy.integrate` package in `Python`, with the `LSODA` integrator.

The objective function used to guide the optimization process is the residual sum of squares (RSS) of the partial pressures. The RSS is calculated as:

$$\mathrm{RSS} = \sum_{j=1}^{M}\sum_{i=1}^{N}\left(P_{i,j}^{\mathrm{exp}} - P_{i,j}^{\mathrm{calc}}\right)^2 \quad (10)$$

where $P_{i,j}^{\mathrm{exp}}$ and $P_{i,j}^{\mathrm{calc}}$ are the experimental and calculated partial pressures of component i in condition j, respectively, and N is the number of components, while M is the number of conditions. This function serves as the reward function for the Q-learning process, where the goal is to minimize the discrepancy between experimental and model-predicted partial pressures.

The optimization process is carried out using the `minimize` function from the `scipy.optimize` library, with the `L-BFGS-B` algorithm [9-10].

The initial values for the parameters $\boldsymbol{A}$ and $\boldsymbol{E_a}$ are defined based on the literature. Moreover, the values of $\boldsymbol{A}$ and $\boldsymbol{E_a}$ are constrained to be positive, as negative values are physically meaningless in the context of reaction kinetics.

In addition to the parameter optimization, a mechanism is incorporated to optimize a set of parameters defined by a vector that activates or deactivates the reactions to be considered in the kinetic model, it ensures that the optimizer adjusts only the parameters corresponding to the active reactions. This is advantageous because it reduces the dimensionality of the parameter space, allowing the optimizer to focus on the relevant reactions.

This approach is particularly useful as the Q-learning algorithm subsequently refines the reaction set, selecting the most appropriate reactions based on experimental data.

RESULTS & DISCUSSION

Evaluation of Reaction Set Combinations

All possible reaction subsets were evaluated, covering every combination from single reactions to the full set of 11 reactions, leading to a total of 2048 cases.

The evaluation of all these combinations provides a comprehensive analysis of the different reaction sets, enabling the identification of the most suitable configuration to describe the methane reforming process. By evaluating every possible combination, it is possible to directly assess the performance of each reaction set based on how well it matches the experimental data. However, this exhaustive approach comes with significant computational costs, especially as the number of reactions increases.

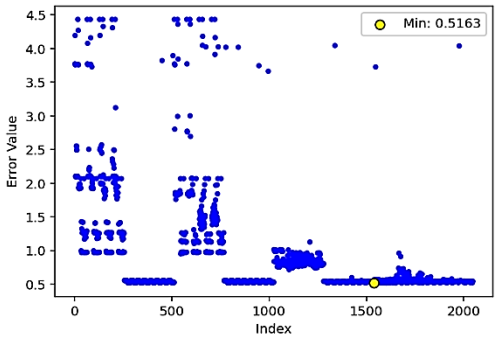

Figure 2. Combinatorial Errors from Reaction Set.

The Fig. 2 shows the evaluation of 2048 reaction set combinations, with each combination indexed on the x-axis and the corresponding residual sum of squares (RSS) on the y-axis. The RSS value indicates the accuracy of the model, with lower values representing a better fit to the experimental data.

The yellow-highlighted combination, with an RSS of 0.5163, is the optimal set, providing the best fit to the data. This combination represents the most accurate model of the methane reforming process among all evaluated sets.

The variation in RSS values illustrates the differing performance of the reaction sets, with most combinations yielding higher RSS values, indicating poorer fits. This emphasizes the importance of selecting the correct reaction set to accurately model the system.

Q-Learning Optimization Results

Fig.3 illustrates the evolution of the best error achieved during the Q-learning training as a function of the number of iterations (11 iterations per episode). It is observed that the best error is reached after approximately 40 iterations, after which no further improvements are identified.

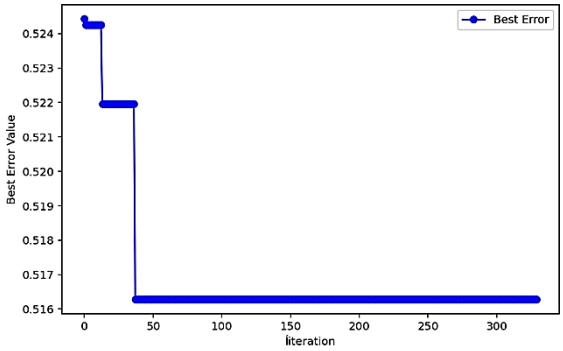

Figure 3. Best Error in training Q-learning.

This behavior suggests that the algorithm converges to the optimal set of reactions during the early stages of training, coinciding with episodes where the epsilon parameter (exploration) is high. As training progresses and epsilon decreases, the algorithm prioritizes exploiting the learned solutions, explaining the absence of better error values beyond this point.

The early convergence indicates that the method is efficient in rapidly identifying the most representative reaction combinations, offering significant advantages in terms of computational time and resource requirements

The Q-learning algorithm identified three key reactions from the initial set of 11 that best describe the methane reforming process reported in table 5.

Table 4: Active Reaction Selected by Q-learning.

Reaction	Active Value
$CH_4 + H_2O \leftrightarrow CO + 3H_2$	1
$CO + H_2O \leftrightarrow CO_2 + H_2$	1
$CH_4 + CO_2 \leftrightarrow 2C + 2H_2O$	1

The first selected reaction is the Methane Steam Reforming (MSR) reaction, which serves as the primary pathway for hydrogen production by breaking down methane in the presence of steam. It is an endothermic reaction that sets the foundation for subsequent reaction pathways, the second is known as the Water Gas Shift (WGS) reaction, this exothermic reaction increases hydrogen yield while reducing carbon monoxide levels. It plays a crucial role in balancing the overall hydrogen and carbon species in the system, the third selected reaction leads to the formation of carbon deposits while consuming hydrogen and carbon dioxide. Its activation highlights the algorithm's ability to capture critical secondary effects that may influence the overall process, especially in systems prone to carbon formation.

The selection of these reactions demonstrates the algorithm's effectiveness in capturing the fundamental and secondary processes necessary for accurately modeling the methane reforming system. These reactions collectively represent the dominant pathways for hydrogen generation and associated carbon transformations under the studied conditions.

The Table 5 presents kinetic constants for the forward and reverse reactions of steam methane reforming (SMR), the water-gas shift reaction (WGS), and the reaction of carbonization. The analysis reveals trends consistent with thermodynamic expectations.

For SMR, the equilibrium shifts towards hydrogen production as temperature increases, aligning with the reaction's endothermic nature. For WGS, the equilibrium shifts towards reactants with increasing temperature. This behavior is expected for an exothermic reaction, were higher temperatures disfavor product formation, as predicted by Le Chatelier's principle.

For the carbonization reaction, the equilibrium similarly shifts to the left at higher temperatures. The reaction's exothermic nature favors the formation of carbon and water at lower temperatures.

Simulation

The Fig. 4 is a parity plot comparing the calculated partial pressures to experimental values for all species. The results reveal a consistent underestimation of water conversion at higher temperatures and elevated methane feed ratios. This behavior is likely due to limitations in the kinetic model, which does not account for adsorption effects on the catalytic surface. Despite this underestimation, the general production trends for all species remain consistent with experimental data, suggesting the model captures the qualitative behavior.

The Fig. 5 presents a simulation of partial pressures as a function of W/F, which is analogous to reaction time. The model demonstrates a tendency to underestimate water conversion throughout the reaction path, consistent with the observations in the parity plot. In contrast, hydrogen production is overestimated, particularly at intermediate and high W/F values. This discrepancy

Table 5: Kinetic Coefficients from Selected Reactions.

T(K)	$\left[\frac{\text{mol}}{\text{atm}\cdot\text{kg}\cdot\text{s}}\right]$	SMR	WGS	Carbonization
609.26	$\mathbf{k_f}$	1.209×10^{3}	4.007×10^{4}	5.767×10^{1}
	$\mathbf{k_r}$	6.601×10^{8} *	1.656×10^{3}	8.095×10^{0}
671.48	$\mathbf{k_f}$	1.429×10^{3}	4.007×10^{4}	5.767×10^{1}
	$\mathbf{k_r}$	1.412×10^{7} *	3.234×10^{3}	9.997×10^{0}
723.15	$\mathbf{k_f}$	1.607×10^{3}	4.007×10^{4}	5.767×10^{1}
	$\mathbf{k_r}$	9.586×10^{5} *	5.166×10^{3}	1.159×10^{1}
779.26	$\mathbf{k_f}$	1.792×10^{3}	4.007×10^{4}	5.767×10^{1}
	$\mathbf{k_r}$	7.734×10^{4} *	8.006×10^{3}	1.331×10^{1}
910.92	$\mathbf{k_f}$	2.198×10^{3}	4.007×10^{4}	5.767×10^{1}
	$\mathbf{k_r}$	7.112×10^{2} *	1.811×10^{4}	1.721×10^{1}

$*\rightarrow \left[\frac{\text{mol}}{\text{atm}^3\cdot\text{kg}\cdot\text{s}}\right]$

may arise from the same model limitations, including the omission of adsorption dynamics and potential inaccuracies in reaction rate parameters.

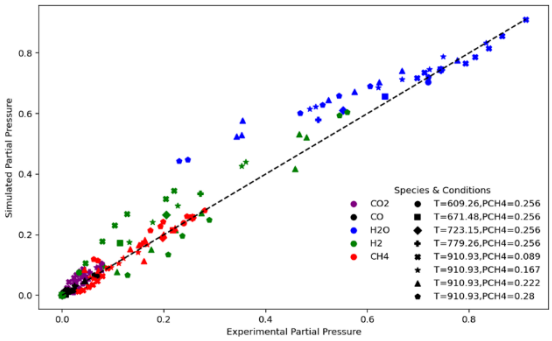

Figure 4. Parity Plot of Simulated Partial Pressures.

Nevertheless, the shapes of the curves and their evolution with W/F align with experimental trends, indicating the model adequately predicts the system's dynamic behavior. These results suggest that while the model requires refinement to improve quantitative accuracy, it reliably captures the qualitative aspects of methane reforming and hydrogen production under varying conditions.

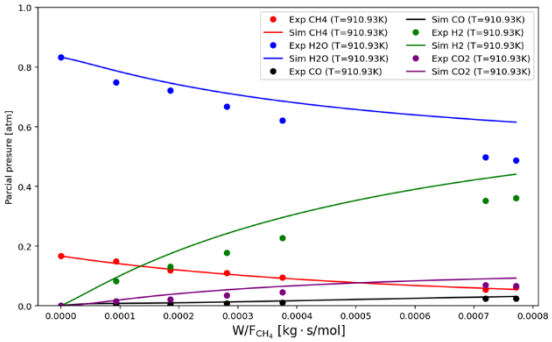

Figure 5. Simulation process with Active Reactions.

CONCLUSIONS

The implementation of the Q-learning algorithm successfully identified the most relevant reactions to describe the methane reforming process. The selected reactions not only align with the physical principles of the process but also agree with previously reported studies, demonstrating the consistency and reliability of the approach. While the model accurately follows experimental trends, it exhibits a tendency to underestimate water conversion and overestimate hydrogen production at high temperatures and higher methane feed ratios. This limitation suggests that exploring alternative models could further enhance representativeness and predictive accuracy.

ACKNOWLEDGEMENTS

The authors thank the National Council for Humanities, Science and Technology (CONAHCyT) of Mexico for the scholarship granted to Martín Rodríguez-Fragoso.

REFERENCES

1. Chen K, Zhao Y, Feng D, Sun S. The intrinsic kinetics of methane steam reforming over a nickel-based catalyst in a micro fluidized bed reaction system. Int J Hydrogen Energy 45: 1615-1628 (2020) https://doi.org/10.1016/j.ijhydene.2019.11.080
2. Sutton RS., Barto. AG. Reinforcement Learning: An Introduction. The MIT Press (2014)
3. Akers WW, Camp DP. Kinetics of the methane-steam reaction. J AIChE 1:471-475 (1955) https://doi.org/10.1002/aic.690010415
4. Allen DW , Gerhard ER, Likins MR. Kinetics of the Methane-Steam Reaction. Ind Eng Chem Process Des Dev 14:256-259 (1975) https://doi.org/10.1021/i260055a010
5. Ross JR, Steel MC. Mechanism of the Steam Reforming of Methane over a Coprecipitated Nickel-Alumina Catalyst. J Chem Soc 69: 10-21(1973) https://doi.org/10.1039/F19736900010
6. Dirksen HA, Riesz CH. Equilibrium in the Steam Reforming of Natural Gas. Ind Eng Chem 45:1562-1565 (1953) https://doi.org/10.1021/ie50523a053
7. Jianguo X, Gilbert FF. Methane Steam Reforming, Methanation and Water-Gas Shift: 1. Intrinsic Kinetics. J AIChe 1: 88-96(1989) https://doi.org/10.1002/aic.690350109
8. Hou K, Hughes R. The kinetics of methane steam reforming over a Ni/α-Al_2O catalyst. J Chem Eng 82: 311-328 (2001) https://doi.org/10.1016/s1385-8947(00)00367-3
9. Zhu C, Byrd RH, Lu P, Nocedal J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans Math Softw 23:550-560 (1997) https://doi.org/10.1145/279232.279236
10. Byrd RH, Lu P, Nocedal J, Zhu C. A Limited Memory Algorithm for Bound Constrained Optimization. J SIAM Sci Comput 16:1190-1208 (1995) https://doi.org/10.1137/0916069