


Letter

Multimodal Data-Driven Reinforcement Learning for Operational Decision-Making in Industrial Processes

Chenliang Liu , Yalin Wang , *Senior Member, IEEE*,
Chunhua Yang , *Fellow, IEEE*, and Weihua Gui 

Dear Editor,

This letter proposes a multimodal data-driven reinforcement learning-based method for operational decision-making in industrial processes. Due to the frequent fluctuations of feedstock properties and operating conditions in the industrial processes, existing data-driven methods cannot effectively adjust the operational variables. In addition, multimodal data such as images, audio, and sensor data are still not fully used in industrial processes. To overcome the impact of feedstock condition fluctuations and effectively utilize operational conditions based on the multimodal data, a new method named feedstock-guided multimodal actor-critic (FGM-AC) is proposed. This letter incorporates the feedstock properties and multimodal data into the state space to guide the decision-making process based on a reinforcement learning (RL) framework to achieve a comprehensive human perception. The effectiveness of the proposed method is verified via extensive experiments conducted on actual industrial data. The results reinforce its potential to provide accurate and dependable strategies for decision-making.

The process industry plays a crucial role in the economic growth of modern society, encompassing steel, petroleum, chemicals, and other fields [1]. In the production process of the process industry, the optimal decision-making of operating variables is crucial for enhancing product quality and yield. However, the decision-making process is often influenced by the experience levels of on-site workers, which can significantly impact the achievement of overall production goals [2], [3]. Moreover, due to the existence of physical and chemical reactions in the production process, it is difficult to establish complex nonlinear relationship models between operational variables and production metrics via mechanism analysis. Hence, optimizing operational variables remains a complex and daunting problem in process industries.

With the increasing availability of industrial data, data-driven decision-making methods that generate decision values for operating variables have become increasingly prevalent in industrial processes. [4] developed a supervised monitoring strategy to adjust the operational variables of the industrial grinding process based on changes in boundary conditions. However, the multimodal data collected from industrial processes, such as images, audio, and sensor data, may be incomplete due to uncontrollable factors. Latent factor analysis is effective in extracting inherent latent features from incomplete data. For instance, [5] proposes a Kalman-filter-incorporated model for performing representation learning to incomplete temporal data. Reference [6] proposes a highly-efficient model for performing representation learning to incomplete industrial data with temporal dynamics. Reference [7] can extract essential non-linear features from incomplete temporal data with high computational efficiency.

Corresponding author: Yalin Wang.

Citation: C. Liu, Y. Wang, C. Yang, and W. Gui, "Multimodal data-driven reinforcement learning for operational decision-making in industrial processes," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 1, pp. 252–254, Jan. 2024.

The authors are with the School of Automation, Central South University, Changsha 410083, China (e-mail: lcliang@csu.edu.cn; ylwang@csu.edu.cn; ychh@csu.edu.cn; gwh@csu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2023.123741

In recent years, with the development of RL, the application of RL-based methods to industrial decision-making has been widely studied [8], [9]. Hence, a model-free RL algorithm presents a promising solution for industrial processes. RL is an innovative and efficient approach to obtaining optimal decision-making policies in industrial processes by interacting with agents and situations approaching real-world complexity. It is noteworthy that in industrial processes, the optimal strategy of the operational variables is conventionally designed by engineers based on historical data and experience, resembling an expert system grounded on the knowledge of operators. Analogous to expert systems, RL has the potential to continuously enhance operational decision-making policies based on reward data that update the performance metrics function. This attribute renders the application of RL algorithms in industrial processes more reasonable.

The motivation of this letter is to develop an intelligent operational decision-making method that overcomes feedstock fluctuations and utilizes multimodal data in industrial processes. The main contributions of this letter are summarized as follows:

1) The multimodal data of the industrial process is utilized to enhance the adaptability of the operational decision-making strategy by fully simulating the overall perception of the operators at the industrial site.

2) To overcome the frequent fluctuations of feedstocks in the industrial processes, the feedstock conditions are introduced as the state space of the proposed algorithm to enhance its accuracy.

3) The unique reward function and state representation are designed to better handle the complexity and specific characteristics of multimodal data in the industrial process, which enhance the performance of the proposed RL framework.

Problem statement: The flotation process plays a significant role in the mineral processing of the process industry, which entails the separation of minerals from raw ores through physicochemical surface properties. The objective of the flotation process is to concentrate the valuable minerals from the raw ores by attaching the desired mineral particles to air bubbles. These air bubbles then ascend to the surface of the flotation cell and create a froth layer that contains the mineral concentrate. Then, the froth is collected and further processed.

To achieve effective flotation in the industrial process, it is necessary to adjust the operating variables in real time based on the working condition fluctuations. These operating variables include the slurry level, aeration, flotation agent, and agitation rate. In the current industrial process, the values of these operating variables are determined by operators based on their experience, with the aim of achieving the desired concentrate yield and grade within the target range. However, due to frequent changes in feedstock and operating conditions, manual selection of setpoints by operators is prone to errors resulting in significant fluctuations in both concentrate output and concentrate grade. A potential solution to this issue involves circumventing the selection of setpoints based on manual experiential knowledge and instead utilizing alternative intelligent decision-making strategies. Effectively implementing such strategies has the potential to significantly enhance the utilization value of raw ore and the overall efficiency of the mineral processing process.

Proposed operational decision-making method: In the industrial flotation process, the formulation of a rational program and the definition of states, rewards, and actions are fundamental to achieving an optimal global decision-making strategy. Inspired by the above analysis and RL algorithm, a new method called the FGM-AC algorithm is proposed for effective operational decision-making in the flotation process. This operational decision-making strategy aims to obtain relatively optimal decision-making values of the operational variables to ensure that the concentration and grade of flotation froth remain within the desired range. Fig. 1 further provides a visual framework of its application in the flotation process, which mainly includes the

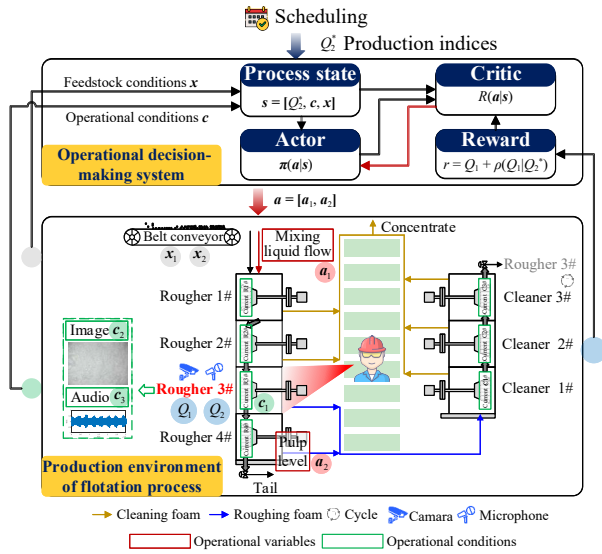


Fig. 1. Operational decision-making framework based on feedstock-guided multimodal actor-critic RL method.

operational decision-making system and the production environment of the flotation process. As shown in Fig. 1, the multimodal data (green circle) derived from the industrial process and feedstock conditions (gray circle) are input into the operational decision-making system. Then, the corresponding product quality and yield (blue circle) are fed back to the system to calculate the resulting rewards. Finally, the decision values of the operational variables are obtained using the proposed method.

Therefore, based on industrial process mechanisms and prior knowledge, the state space of the RL algorithm includes operational conditions, feedstock conditions, and the target grade of flotation froth. It is denoted as $s = [Q_2^*, c, x]$. In particular, industrial cameras and microphones are used to collect flotation froth images and audio from actual industrial sites to assist in operational decision-making.

To achieve this, the reward function of the proposed FGM-AC algorithm is defined as

$$r = Q_1 + \rho(Q_2 | Q_2^*) \quad (1)$$

$$r(Q^*, a, c, x) = f_1(a | c, x) + \rho(f_2(a | c, x) | Q_2^*) \quad (2)$$

where r is a nonpositive scalar function, $f_1(a | c, x)$ represents the concentration of flotation froth, $\rho(f_2(a | c, x) | Q_2^*)$ is a penalty function.

The decision-making framework for operational variables, as presented in (2), can be transferred to (3) in the RL algorithm framework, which is given as

$$J(\pi) = E_{s \sim e, a \sim \pi} [r(s_t, a_t)]. \quad (3)$$

It is worth noting that, unlike other sequential decision processes, the decision-making of operational variables in this context is not sequential since they are often interrelated and influenced by multiple factors. Therefore, the step size T is selected to be one in each episode. The iterative approach is frequently employed to refine the optimal decision-making policy, which can be characterized as a continuous process. Thus, the derivation of this policy is described as

$$\pi_{\text{new}} = \arg \max_{\pi} E_{s \sim e(s), a \sim \pi(a|s)} [r(s, a)] \quad (4)$$

where $\pi(a|s)$ is assumed as a conditional distribution belongs to Gaussian distribution.

Considering the high-dimensional and continuous nature of the state and action spaces involved in the optimal operational decision-making problem, an actor network is employed using a neural network implementation denoted as $\pi_{\theta}(a|s)$, where parameter θ is used to approximate the Gaussian distribution. The actor network utilizes the state as input and produces the action as output.

In addition, the critic network $R_{\varphi}(a|s)$ with parameter φ is used to estimate the reward generated by $\pi_{\theta}(a|s)$. The input of the critic net-

work consists of the state and action. During the training process, the loss function of the critic network is defined as follows:

$$J(\varphi) = \frac{1}{2} [R_{\varphi}(s, a) - r(s, a)]^2 \quad (5)$$

where $r(s, a)$ represents the actual reward of the production data. Then, $R_{\varphi}(s, a)$ can be replaced by $r(s, a)$ when the training accuracy is satisfied. Hence, the policy is updated as

$$\pi_{\theta_{\text{new}}} = \arg \max_{\pi_{\theta}} E_{s \sim e(s), a \sim \pi_{\theta}(a|s)} [R_{\varphi}(s, a)]. \quad (6)$$

Furthermore, integrating experience replays into the FGM-AC algorithm allows for repeated learning from experiential data with benefits such as reduced costs, fewer trials and errors, and faster learning speeds. In the experience replay method, a set of experiences consisting of the state, action, and immediate reward obtained during the interaction between the FGM-AC algorithm and the flotation production process is stored in the experience pool. By minimizing the loss function defined based on the criterion, the decision-making policy can be improved as

$$\pi_{\theta_{\text{new}}} = \arg \max_{\pi_{\theta}} E_{s \sim P, a \sim \pi_{\theta}(a|s)} [R_{\varphi}(s, a)] \quad (7)$$

where P denotes the experience replay pool. It should be noted that a batch gradient descent method is used to train the critic network. Subsequently, the loss function is reformulated as shown below:

$$J(\varphi) = \frac{1}{2} E_{(s, a, r)} [R_{\varphi}(s, a) - r(s, a)]^2. \quad (8)$$

Subsequently, the FGM-AC algorithm is used to obtain the relatively optimal decision-making policy based on the realizations of actor and critic networks based on iteratively updating (7) and (8) in an alternating manner. Finally, the optimal decision-making values of the operational variables are obtained from the actor network, denoted as

$$\tilde{a} = \arg \max_a R_{\varphi}(s, a) \quad (9)$$

where \tilde{a} represents the optimal decision-making values of the operational variables.

Experiments and analysis: The proposed operational decision-making method based on FGM-AC is applied to an actual industrial flotation process. All experimental data sets are collected from the largest potassium chloride flotation plant of a mineral processing enterprise. A total of 223 data sets were collected, including the feed ore conditions, operational conditions, operational variables, and performance metrics. A detailed description of these variables is given in Table 1. The first 180 data sets were used for training, while the remaining 43 data sets were used for validation.

Table 1. Discription of Data Sets in the Industrial Flotation Process

Tag	Description
Feedstock condition	Feedstock flow (x_1), feedstock grade (x_2)
Operational condition	Stirring current (c_1), froth image (c_2), froth audio (c_3)
Operational variable	Mixed mother liquid flow (a_1), roughing flotation pulp level (a_2)
Performance metric	Froth concentration (Q_1), froth grade (Q_2)

In the RL framework of the proposed FGM-AC algorithm, the state vector is composed of the feedstock conditions x , operational conditions c , and target flotation froth grade Q_2^* . The action vector is obtained from the proposed operational decision-making method based on the FGM-AC algorithm. The production goal of the industrial flotation process is to maximize the flotation froth concentration while meeting the flotation froth grade specifications. Hence, the reward function is designed as

$$r = r_1 + r_2 \quad (10)$$

$$r_1 = Q_1, \text{ and } r_2 = \begin{cases} -0.6, & Q_2 < Q_2^* \\ 0, & Q_2 \geq Q_2^* \end{cases} \quad (11)$$

Comparative experiments are designed to assess the effectiveness

of the proposed method. Manual operations collected at industrial sites were used as a baseline for comparison. In addition, operational decision frameworks based on the deep Q-network (DQN) [10] and the standard actor critic (AC) [11] are used as additional comparisons. For unbiased and impartial experimentation, all actor networks use three-layer neural networks comprising 64 hidden-layer neurons and are trained using a learning rate of 0.01.

The experimental results of the flotation froth performance metrics under four comparison methods are presented in Table 2 and Fig. 2. Table 2 gives the minimum, maximum, and average values (in parentheses) of the performance metrics. Fig. 2 intuitively depicts the trajectories of two performance metrics. It can be seen from Table 2 that the proficiency of on-site operators lies primarily in regulating the froth grade, while their control of froth concentration has no significant advantages. However, other methods based on the RL framework, including DQN, AC, and FGM-AC, have significantly improved froth concentration, which indirectly guarantees an increase in yield. Specifically, the proposed FGM-AC-based operational decision-making method increases the froth concentration by 8.51% and the froth grade by 1.43% compared to manual operation. However, improving froth concentration while maintaining froth grade in actual industrial processes is usually difficult. Hence, it also demonstrates its effectiveness in optimizing industrial processes.

Table 2. Comparison Results of Four Methods

Method	Concentrate (%)	Grade (%)
Manual operation	36.36–44.37 (40.64)	27.07–31.81 (29.40)
DQN	40.85–46.49 (43.50)	28.05–30.93 (29.33)
AC	38.96–46.79 (43.64)	26.66–31.51 (29.46)
FGM-AC	40.24–46.88 (44.10)	27.12–31.54 (29.82)

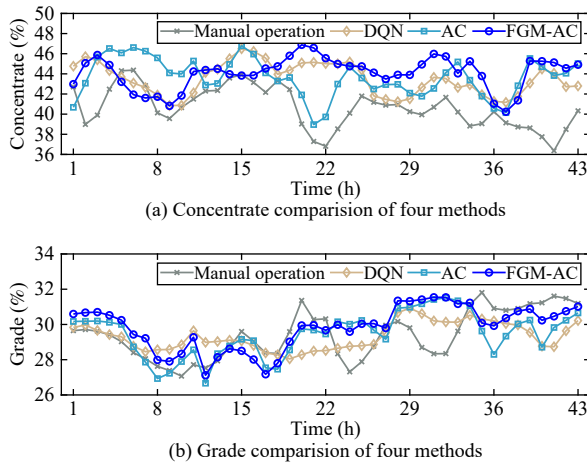


Fig. 2. Comparison results of froth concentrate and grade.

Moreover, the trajectories of two operational variables are shown in Fig. 3. It is evident that the mixed mother liquor flow is set higher in the three operational decision-making methods based on the RL framework. This is done by increasing the mixed mother liquor flow rate to boost the concentration of flotation froth, indirectly leading to an increase in froth production, which is consistent with the knowledge and experience of experts. Furthermore, the flotation pulp flow is maintained relatively low compared to manual operation to prevent the loss of flotation froth.

Conclusion: This letter proposes a multimodal data-driven RL-based decision-making method for operational variables in industrial processes, which aims to mitigate the effect of feedstock conditions and exploit underutilized multimodal data. Specifically, a new FGM-AC algorithm is proposed to convert the operational variable decision-making problem into an RL problem. Compared to the existing algorithms, the proposed FGM-AC algorithm makes full use of the multimodal data of the industrial sites and has a more comprehen-

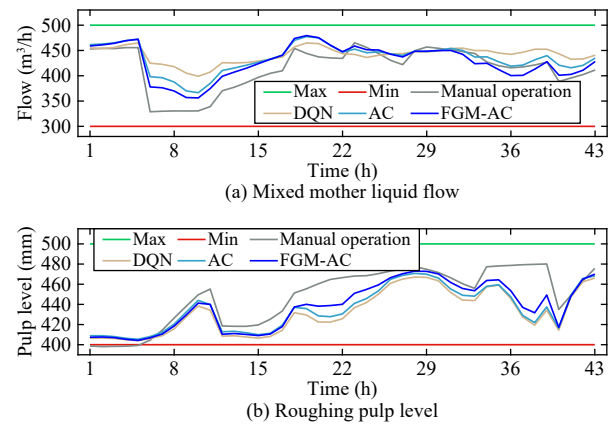


Fig. 3. Optimal operational variables of four methods.

sive perception ability. Finally, the experimental results using actual data of the industrial flotation process demonstrate the favorable potential for guiding the production of industrial processes. The future work will focus on enhancing the security of online RL algorithms in industrial applications and extending our work to other industrial processes where multimodal data are available.

Acknowledgment: This work was supported by the National Key Research and Development Program of China (2020YFB1713800), the National Natural Science Foundation of China (92267205), the Hunan Provincial Innovation Foundation for Postgraduate (CX2022 0267) and the Fundamental Research Funds for the Central Universities of Central South University (2022ZZTS0181).

References

- [1] J. Huang, Z. Li, and Z. Zhou, "A simple framework to generalized zero-shot learning for fault diagnosis of industrial processes," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 6, pp. 1504–1506, 2023.
- [2] L. Hu, K. Chan, X. Yuan, and S. Xiong, "A variational bayesian framework for cluster analysis in a complex network," *IEEE Trans. Knowl. Data Engineering*, vol. 32, no. 11, pp. 2115–2128, 2020.
- [3] J. Wang, Q. Zhang, and D. Zhao, "Highway lane change decision-making via attention-based deep reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 567–569, 2022.
- [4] P. Zhou, T. Chai, and J. Sun, "Intelligence-based supervisory control for optimal operation of a DCS-controlled grinding system," *IEEE Trans. Contr. Syst. T.*, vol. 21, no. 1, pp. 162–175, 2013.
- [5] Y. Yuan, X. Luo, M. Shang, and Z. Wang, "A Kalman-filter-incorporated latent factor analysis model for temporally dynamic sparse data," *IEEE Trans. Cyber.*, vol. 53, no. 9, pp. 5788–5801, 2023.
- [6] X. Luo, H. Wu, Z. Wang, J. Wang, and D. Meng, "A novel approach to large-scale dynamically weighted directed network representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9756–9773, 2022.
- [7] X. Luo, H. Wu, and Z. Li, "Neultf: A novel approach to nonlinear canonical polyadic decomposition on high-dimensional incomplete tensors," *IEEE Trans. Knowl. Data Engineering*, vol. 35, no. 6, pp. 6148–6166, 2023.
- [8] L. Hu, Y. Yang, Z. Tang, Y. He, and X. Luo, "FCAN-MOPSO: An improved fuzzy-based graph clustering algorithm for complex networks with multi-objective particle swarm optimization," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 10, pp. 3470–3484, 2023.
- [9] Y. Wang, S. Li, C. Liu, K. Wang, X. Yuan, C. Yang, and W. Gui, "Multi-scale feature fusion and semi-supervised temporal-spatial learning for performance monitoring in the flotation industrial process," *IEEE Trans. Cyber.*, 2023. DOI: 10.1109/TCYB.2023.3295852.
- [10] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, USA: MIT press, 2018.
- [11] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. 35th Int. Conf. Mach. Lear.*, 2018, pp. 1861–1870.