Neurocomputing 380 (2020) 190–200

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Optimizing zinc electrowinning processes with current switching via Deep Deterministic Policy Gradient learning☆

Xiongtao Shi[a], Yonggang Li[a,*], Bei Sun[a,*], Honglei Xu[b], Chunhua Yang[a], Hongqiu Zhu[a]

[a] School of Automation, Central South University, 410083, China
[b] School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Perth, WA 6845, Australia

ARTICLE INFO

Article history:
Received 21 May 2019
Revised 14 October 2019
Accepted 9 November 2019
Available online 13 November 2019

Communicated by Prof. Hamid Reza Karimi

Keywords:
Current switching
Deep Deterministic Policy Gradient
Zinc electrowinning process

ABSTRACT

This paper proposes a model-free Deep Deterministic Policy Gradient (DDPG) learning controller for zinc electrowinning processes (ZEP) to save energy consumption during the current switching periods. To overcome the problems such as inaccurate modeling and various time delays, the proposed DDPG controller utilizes various control periods and parameters for different working conditions. Strategies such as action boundary setting, reward function definition, state normalization are applied to ensure its learning performance. Simulations and experiments show that the DDPG learning controller can significantly decrease energy consumption during the ZEP current switching periods. The optimal control policy will be learnt for different working conditions with only one group hyperparameters. Furthermore, the smoother control actions of the DDPG controller will improve the stability and reduce more energy consumption by comparing with traditional proportional-integral (PI) controller, model predictive control (MPC) and artificial experiences. The artificial intelligence-based optimal control framework brings both energy saving and intelligence to zinc manufacturing plants.

1. Introduction

A zinc electrowinning process (ZEP) is an essential step in zinc production. It extracts pure valuable zinc metal from zinc ions via direct current. As is well known, more than 80% electrical energy consumption in the zinc hydrometallurgy process is used at the ZEP stage [1]. Due to the policy of power-time-sharing, the power price changes dramatically at different periods of a day. Thus, the optimal current density, defined as the amperage per square meter of the current of electrodes, should adapt with the change of power price [2]. More specifically, the ZEP current density should be set to a large value during the low power price period and set to a small value during the high power price period. The adjustment of the zinc ion concentrations (CZN) and acid ion concentrations (CH) is required to achieve optimal values by regulating the leaching solution flow rate. However, due to intrinsic complexity of ZEP, large time delays caused by three huge volume reactors and the time delay varies with current density value, it is too difficult to calculate the optimal leaching flow rate which drives the CZN and CH in the electrolysis cell to reach an optimal value under the required time. It is noticed that the optimal values of CZN and CH in the electrolysis cell will change several times in one day, which will lead to a considerable energy waste during power price transition periods. So from the economic perspective, it is necessary to design an optimal controller to save the enormous energy waste.

To overcome the challenges in the optimal control of ZEP, people have proposed many effective methods. Deng et al. [3] proposed a spatiotemporal distribution model of the ZEP and used orthogonal approximation to estimate the model parameters. Barton and Scott [4] established a mathematical pilot-plant model to scale up the model to the full-size ZEP. Wang et al. [5] developed nonlinear relationships to estimate the current density, cell voltage and energy consumption of the ZEP based on the tools of fuzzy neural networks. Behnajady and Moghaddam [6] investigated a new method to figure out the accurate parameter of the ZEP. Yang et al. [7] developed an optimal power-dispatching system using an artificial neural network to minimize the power consumption when the power price is varying. Barton and Scott [8] established a mechanism model to reveal the nonlinear relationship among CZN, CH and energy consumption. Mahon et al. [9] developed an optimal method to solve the problem that how to control the current efficiency, energy consumption and zinc production rate to the optimal value. Yang et al. [10] designed an optimal controller for the

☆ This work was supported by the National Natural Science Foundation of China (Grant Nos. 61673400, 61890930-2 and 61621062) and the Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2019zzts570).

* Corresponding authors.
E-mail addresses: liyonggang@csu.edu.cn (Y. Li), sunbei@csu.edu.cn (B. Sun).

https://doi.org/10.1016/j.neucom.2019.11.022
0925-2312/

ZEP via a state transition algorithm. More generally, researchers have also put great efforts on mathematical modeling [11–16], optimization methods [17] and control strategy [18].

However, the methods mentioned above are the model-based ones, and the control periods and parameters are constant under different working conditions. Even though mathematical model structures could be obtained by applying physical and chemical theories, accurate model parameters can't be obtained due to many unexpected conditions, such as insufficient data samples [19]. Moreover, different control periods and parameters should be used for distinct working conditions with various time delays. When the system model changes, the model-based controller needs to adjust the corresponding parameters to achieve the optimal performance.

Nowadays deep learning is applied in many aspects ranging from machine vision [20], natural language processing [21] to control systems [22]. And the stability of control system with neural networks is proved by Lyapunov method in [22,23]. In [24], the comparison between RL and model predictive control (MPC) is provided. The results show that RL may certainly be competitive with MPC even in contexts where a good deterministic system model is available. Meanwhile the deep reinforcement learning (DRL) algorithms such as AlphaGo [25] and AlphaGo Zero [26] learn in a new environment automatically. The DRL is validated by the robots systems [27,28] and industry system [29]. The DDPG is one of the DRL that has been applied widely in wireless communications [30], manipulator control [31] and robotic control [32]. In hydrometallurgy industry, the digitalizing zinc production system will provide a large number of data, which can be described elaborately. Hence, model-free learning algorithms are highly demanded, which can automatically learn the optimal control strategy.

With above analysis, the DDPG is applied in industrial system successfully. Compared with model-based methods [10,24], the proposed DDPG controller can overcome the disadvantages of inaccurate modeling due to many unexpected conditions, such as insufficient data samples [19] and poor adaptability in the problem of system model changes as time go on. Compared with PI controller and MPC controller, the DDPG controller can give the smoother control actions which will reduce energy consumption and improve the stability of industrial system.

Therefore, The main contributions are listed as follows:

1. The model-free Deep Deterministic Policy Gradient (DDPG) algorithm is applied in zinc electrowinning processes (ZEP) to overcome the problems of inaccurate modeling and to adapt the problem of system model changes as time go on.
2. We use only one group hyperparameters to learn different policies for different working conditions in ZEP.
3. Strategies such as action boundary setting, reward function definition, state normalization are applied to ensure its learning performance. The algorithm is hard to work well without these strategies.
4. Compared with proportional-integral (PI) controller and MPC controller, the DDPG algorithm can provide a smoother control action, which could reduce more energy consumption and improve the stability of practical industrial system. Moreover the DDPG controller can adapt the changes of system parameters due to its learning characteristic.

The rest of this paper is organized as follows. Section 2 introduces the ZEP in detail. Section 3 develops a ZEP simulation model composed of a mass balance model and an energy consumption model. Section 4 solves the ZEP optimal control problem with the DDPG controller during current switching periods. Several adaptions are brought into the DDPG controller, which yields huge applicability to implement the optimal control in the ZEP. Section 5 provides the training and testing results, which shows that the DDPG controller could achieve a better performance for the ZEP simulation compared with a PI controller, MPC controller and artificial experiences. Finally, some conclusions are provided in Section 6.

2. Problem analysis

There are six parts in Fig. 1 which consists of leaching solution tanks part, mixing cell part, spent electrolysis cell part, electrolysis cell part, spent electrolyte tanks and cooling towers. Before the ZEP, all impurity ion concentrations are reduced below an allowable range in the solution purification process. First, the leaching solution will be stored in the leaching solution tanks about 16–24 h to deposit little impurities to the bottom of tanks. Second, the leaching solution and the spent solution are thoroughly mixed in the mixing cell which prepare for electrolyzing. Third, electrolysis reaction will occur at electrolysis cell, which will get higher acid ion concentrations and lower zinc ion concentrations in the spent electrolyte tanks than the mixed solution. Fourth, most of the spent electrolyzers are pumped back to re-mix with the leaching solution; the rest solution in the spent acid tanks is to prepare for other zinc product processes. Finally, several cooling towers are constructed as the operators to control the temperature of the electrolyzers to a suitable range. When the temperature of electrolyzers are higher than the optimal value, the spent electrolyzer will be cooled via pumping the spent electrolyzer through the cooling towers.

In the ZEP, the primary cathode half reactions for the deposition of zinc [33] are

$$Zn^{2+} + 2e^{-} \mapsto Zn \tag{1}$$

and

$$2H^{+} + 2e^{-} \mapsto H_2 \tag{2}$$

More than 90% of the power is used in zinc production by Eq. (1). The primary anodic half-reactions are

$$2H_2O \mapsto 4H^{+} + 4e^{-} + O_2 \tag{3}$$

and

$$Mn^{2+} + 2H_2O \mapsto 4H^{+} + MnO_2 + 2e^{+} \tag{4}$$

Over 99% of the direct current is used in Eq. (3). We assume that only exists reaction of Eq. (3) in the anodic.

We could get the reaction rate of zinc production ($r_{ZN^{2+}}$) and hydrogen ($r_{H^{+}}$) by the following equation [34]:

$$\begin{aligned} r_{Zn^{2+}} &= \frac{DS\varepsilon}{2F} \\ r_{H^{+}} &= \frac{DS(1-\varepsilon)}{F} \end{aligned} \tag{5}$$

where S is the electrode area, D is the current density, F is the constant of Faraday, and ε is the current efficiency which will be estimated by the energy consumption model of artificial neural networks developed in Section 3.

To reduce the energy consumption, we should adjust the CZN and CH to the optimal values related to the current density which is connected with the power price. When the current density is the constant value, the optimal CZN and CH are also constant values. The optimal values can be finally reached by simple feedback control methods. However, the practical condition is that the current density changes periodically every day, which yields many challenges when we control the CZN and the CH to the optimal values under the required time. Furthermore, the model-based methods' performance is not perfect due to accurate model parameters can't be obtained, which often occurs in practice.

Therefore we will develop the ZEP simulation model constituted by the mass balance model and the energy consumption

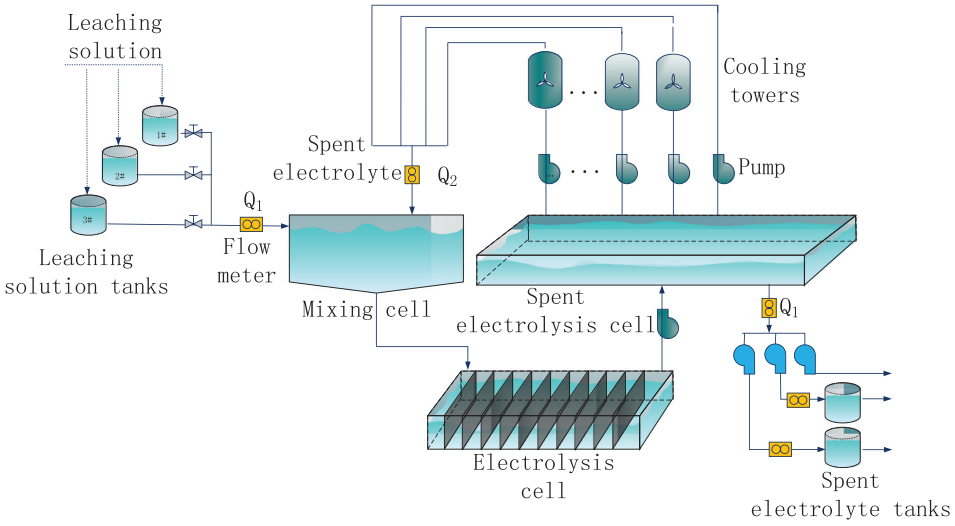


Fig. 1. Flow chart of a zinc electrowinning process.

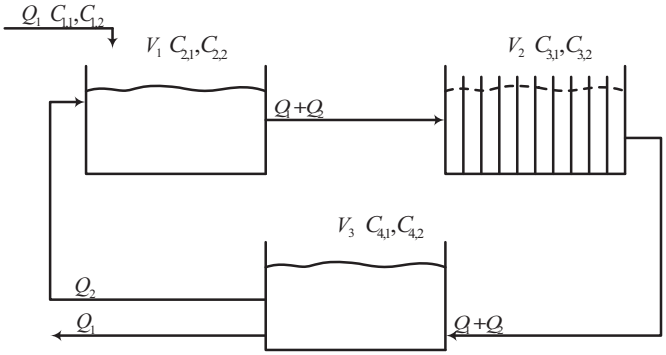


Fig. 2. Simple flow chart of zinc electrowinning process.

model. Then we use different control periods and parameters of the model-free DDPG learning controller with some adaptations to automatically learn the optimal control strategy by self-exploration in the ZEP simulation environment.

3. Mathematical modeling

In this section, we develop the ZEP simulation environment composed of the mass balance model and the energy consumption model to simulate the practical ZEP.

3.1. Mass balance

We assume that the mass balance in Fig. 2 among three major cells (mixing cell, electrolysis cell and spent electrolysis cell) satisfies the assumptions as follows [10]:

Assumption 1. The temperature of electrolyzer is appropriately controlled within the suitable range.

Assumption 2. Impurity ions, such as Ni^{2+}, Cu^{2+}, Co^{2+} et al., are not considered [19].

Assumption 3. The liquid is uniformly mixed in three prime cells [35].

Assumption 4. The subtle loss of the solution such as the evaporation is not considered.

We apply the mass balance theory based on the above four assumptions. The derivative of the concentrations of CZN and CH in three main cells can be obtained by the following equation [10]:

$$\begin{aligned}
\frac{dC_{2,1}}{dt} &= \frac{Q_1C_{1,1} + Q_2C_{4,1} - (Q_1+Q_2)C_{2,1}}{V_1} \\
\frac{dC_{2,2}}{dt} &= \frac{Q_1C_{1,2} + Q_2C_{4,2} - (Q_1+Q_2)C_{2,2}}{V_1} \\
\frac{dC_{3,1}}{dt} &= \frac{(Q_1+Q_2)C_{2,1} - (Q_1+Q_2)C_{3,1} - r_{Zn^{2+}}}{V_2} \\
\frac{dC_{3,2}}{dt} &= \frac{(Q_1+Q_2)C_{2,2} - (Q_1+Q_2)C_{3,2} + r_{H^+}}{V_2} \\
\frac{dC_{4,1}}{dt} &= \frac{(Q_1+Q_2)C_{3,1} - (Q_1+Q_2)C_{4,1}}{V_3} \\
\frac{dC_{4,2}}{dt} &= \frac{(Q_1+Q_2)C_{3,2} - (Q_1+Q_2)C_{4,2}}{V_3}
\end{aligned} \tag{6}$$

where $C_{i,j}$ $(i = 1, 2, 3, 4; j = 1, 2)$ are the CZN and CH of the leaching solution, the mixed solution, the electrolyzers and the spent electrolyzer, respectively. $V_i (i = 1, 2, 3)$ are the volume of the three cells. Q_1, Q_2 are the leaching solution flow rate and the spent electrolyte solution flow rate, respectively. $r_{Zn^{2+}}$ is the reaction rate of zinc ion, and r_{H^+} is the reaction rate of acid ion. It is noticed that $C_{1,2}$ is approximately zero due to the solution purification process.

In the practical ZEP, the flow rate of the leaching solution Q_1 and the spent electrolyte solution Q_2 are constrained by:

$$Q_{i,\min} \le Q_i \le Q_{i,\max} \tag{7}$$

where $Q_{i,\min}$ and $Q_{i,\max}$ $(i = 1, 2)$ are the minimal and maximal values of the leaching solution flow rate and the spent electrolyte solution flow rate, respectively.

To ensure the stability of zinc production, the CZN and CH in the electrolytic cell have the following constraints:

$$C_{i,j,\min} \le C_{i,j} \le C_{i,j,\max} \tag{8}$$

where $C_{i,j,\min}$ and $C_{i,j,\max}$ are the minimal and maximal values of the concentrations in four solutions. The values of $Q_{i,\min}$, $Q_{i,\max}$, $C_{i,j,\min}$, $C_{i,j,\max}$ and V_i can be obtained from the practical ZEP.

3.2. Energy consumption

To implement the optimal control of the ZEP, we need to know the energy consumption under current states which has relations with the CZN, CH and current density. An artificial neural network

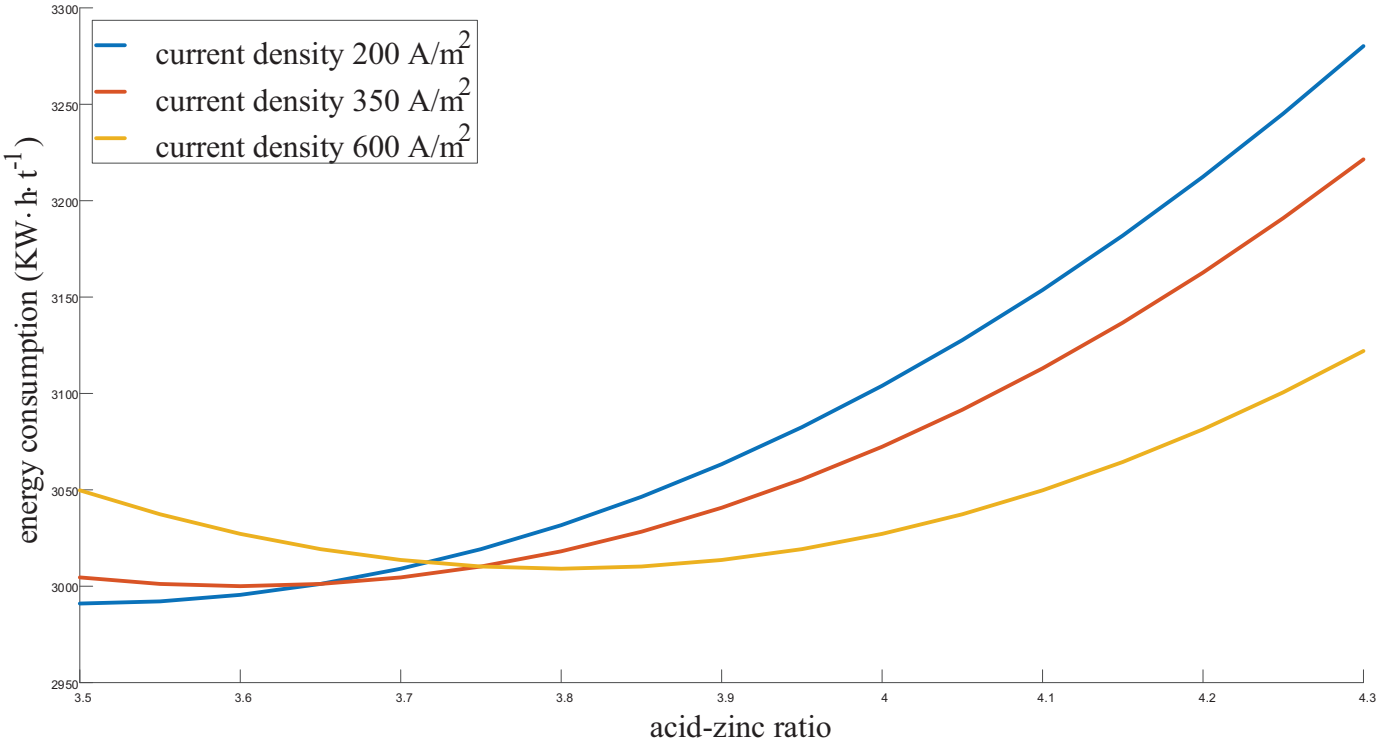


Fig. 3. The nonlinear relationship between the energy consumption with the acid-zinc ratio.

energy consumption model is put forward to estimate the energy consumption. The energy consumption, defined as the power required per unit weight of zinc production, can be calculated using the following equation [36]:

$$W = \frac{81,960V_c}{\varepsilon} \tag{9}$$

where W is the energy consumption, V_c is the cell voltage, and ε is the current efficiency. We use two artificial neural networks to calculate V_c and ε, respectively, and then apply Eq. (9) to calculate the energy consumption. To improve the accuracy of the energy consumption model we choose several different network parameters to estimate the energy consumption under the different current density levels. More specifically, the current density level will be set to three different constant values in one day. Then the energy consumption model also has three group parameters. We load different parameters to estimate the energy consumption under different current density level. The artificial neural network structures of V_c and ε are $2 \times 100 \times 100 \times 1$ and $2 \times 100 \times 100 \times 1$. The network input states are the CZN and CH. We use the Relu activation function to improve the training speed. Then we optimize the parameters of the energy consumption model by using the back-propagation algorithm and feeding in the data of the practical zinc production process.

As shown in Fig. 3, we see that the different current density corresponds to different optimal values of CZN and CH. The optimal acid-zinc ratios under 600 A/m^2, 350 A/m^2, 200 A/m^2 are 3.8, 3.6, 3.5, respectively. So the aim of the optimal controller is to control the CZN and CH in the electrolysis cell to the optimal values as quickly as possible when the current density changes.

3.3. Zinc electrowinning process simulation setup

The ZEP simulation environment is based on the mass balance model and the energy consumption model. We set three volumes V_1, V_2, V_3 to constant values which are obtained from the practical ZEP. Then, we randomly initialize $C_{i,j}$ $(i = 1, 2, 3, 4;\ j = 1, 2)$ which is constrained in a suitable range of the practical ZEP also. We use the mass balance model to update the ZEP's states. We adjust the leaching solution flow rate Q_1 to influence the environment's states. If we need states feedback, we read the mass balance model's states and use the energy consumption model to estimate the energy consumption as the control feedback. The running time of the simulation is up to 6 h. If the states are out of the correct range for the incorrect control input, then we reset the simulation environment and run from the beginning.

4. Optimal control with deep deterministic policy gradient

In this section, we implement the optimal control of the ZEP during current switching with the help of the DDPG controller. We get the different system delay times by the simulation, then we train the DDPG controller under four different working conditions with only one group hyperparameters. The DDPG controller can minimize energy consumption by self-exploration in the ZEP simulation environment automatically.

4.1. Time-delay characteristics of the ZEP system

From Fig. 4, we have four different working conditions, and the current density switches from 600 A/m^2 to 350 A/m^2, from 350 A/m^2 to 600 A/m^2, from 350 A/m^2 to 200 A/m^2, from 200 A/m^2 to 600 A/m^2, respectively. To get the system time delays under four working conditions, we use the leaching flow rate of 100 m^3/h, 0 m^3/h, 100m^3/h, 0m^3/h, which is the fastest action to reach the optimal value of the acid-zinc ratio. The simulation starts from a stable working condition, and use the fastest action which drives the acid-zinc ratio to the optimal value. We stop the simulation once the acid-zinc ratio reaches the optimal value.

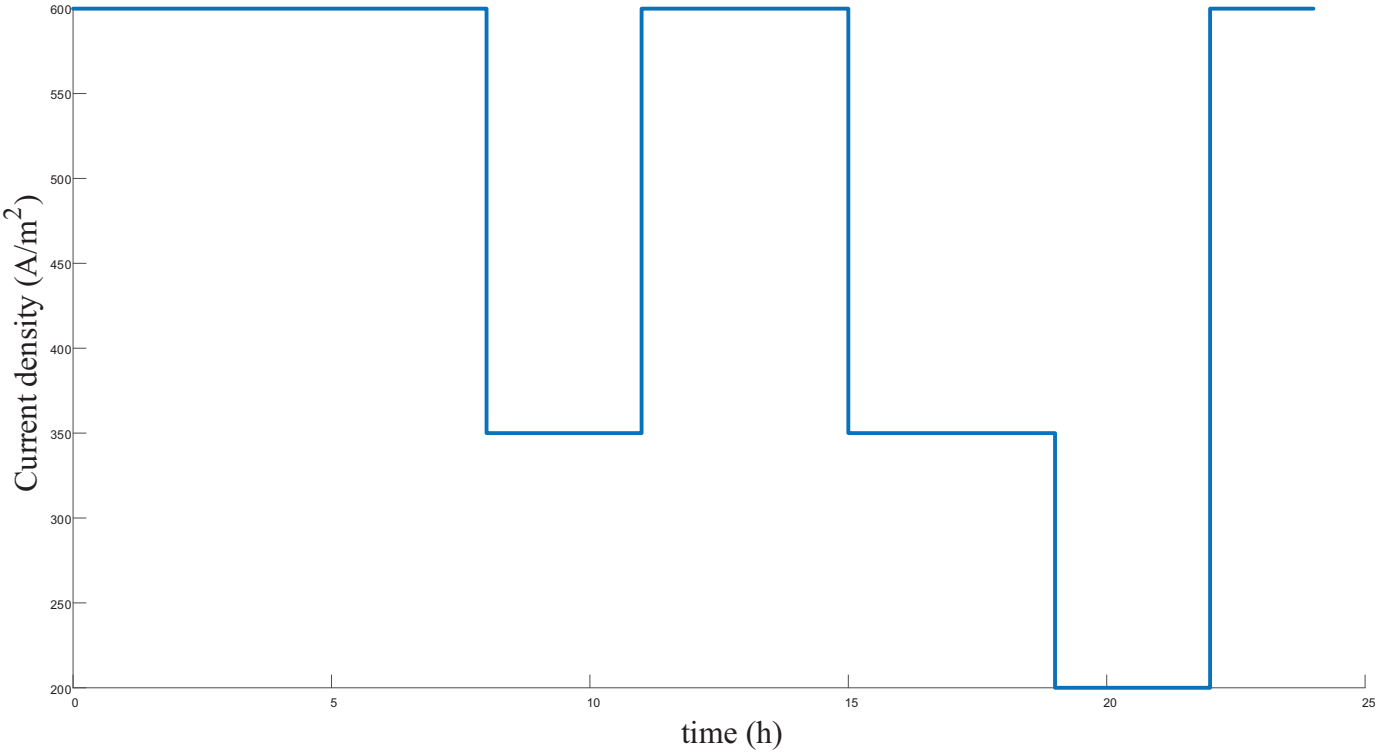


Fig. 4. Current density in one day.

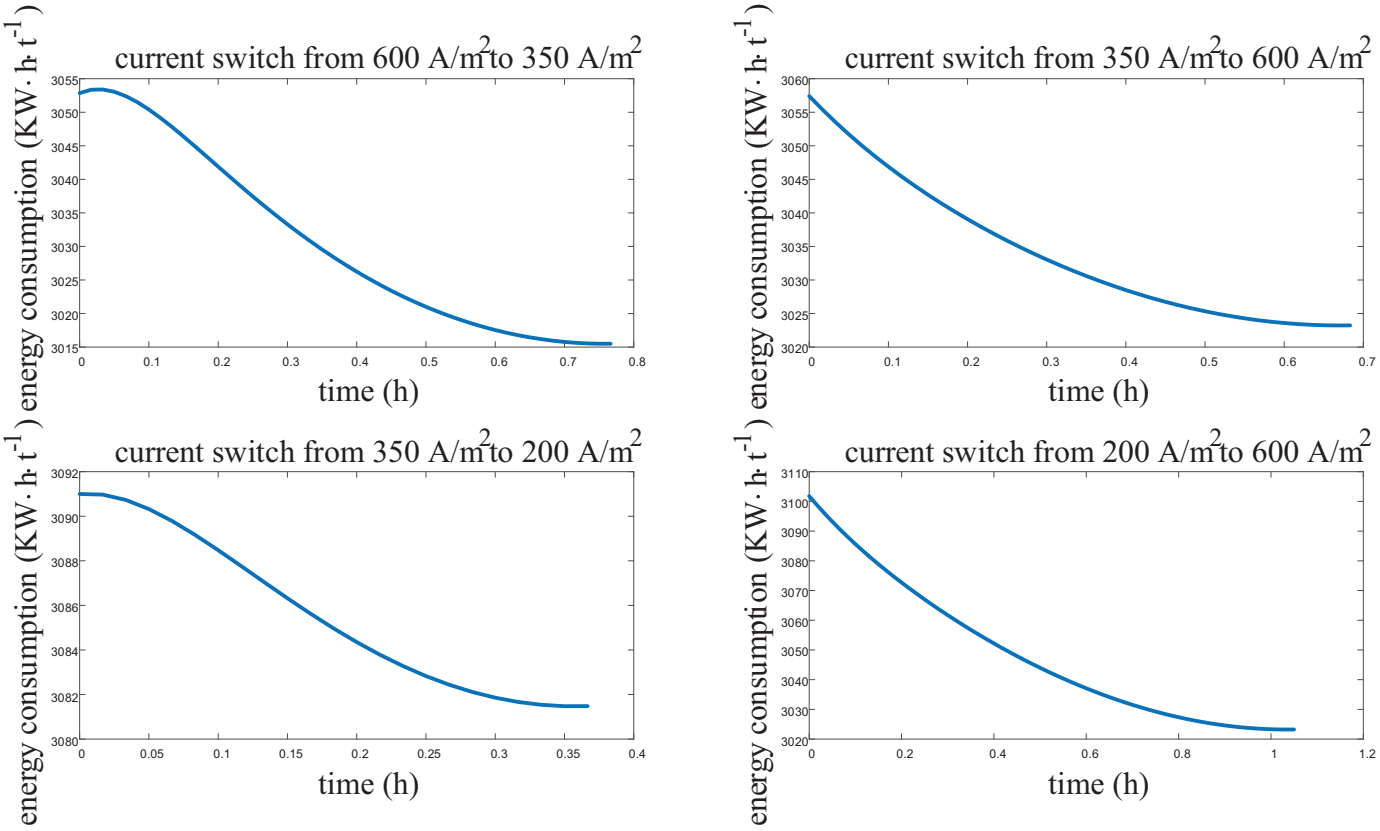


Fig. 5. Energy consumption under different working condition.

From Fig. 5, the system delay times under the above four current density switching cases are about 0.77 h, 0.68 h, 0.37 h, 1.1 h, respectively. We choose a frequency 10 times than the system delay time to control. That means the control periods are 0.077 h, 0.068 h, 0.037 h, 0.11 h, respectively. It is better than the case which uses a constant control period.

4.2. Deep reinforcement learning

Reinforcement learning is an algorithm that learns how to map a situation to an action directly by self-exploration in the environment. The algorithm does not give a right action under current states, but to try different actions in the environment and use a numerical reward signal to learn the action which maximizes the long term discounted reward in Eq. (10) where the discount factor $0 < \gamma < 1$. The time delay exists in the most environment which means that actions may not only affect the immediate reward but also affect the next situation and all subsequent rewards. The try and error search and delayed reward are the notable characteristics of reinforcement learning.

$$r_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots \tag{10}$$

Deep reinforcement learning is a combination of deep learning and reinforcement learning. The artificial neural network can express the strong-nonlinear relationship between states and actions directly. Deep reinforcement learning can control the action in the continuous action-space.

Actor-critic algorithm is a widely known algorithm based on the policy gradient theorem which includes two parts: an actor part is to learn the relationship between states and actions, and a critic part is to lean the value function which evaluates the current states from the long run. However, the standard actor-critic algorithm needs to train for a long time to converge. More seriously, this algorithm cannot converge to the optimal result sometimes. To ensure the convergence of the learning algorithm and improve the learning speed, Lilllicrap et al. [37] developed a Deep Deterministic Policy Gradient (DDPG) algorithm, borrowing an essential concept from Deep Q-Network (DQN) [38] that learning in mini-batch. DQN uses a large replay buffer in the form $[s_t, a_t, r_t, s_{t+1}]$ to store historical samples. During the training, the mini-batch tuple data are grabbed randomly from the replay buffer to update network weights. Furthermore, a new target actor, and a critic network are created as a copy of the origin actor, and critic network at initial phase. The weights of the target network are updated by slowly tracking the learnt networks to improve the training stability.

4.3. Zinc electrowinning process optimal control framework

The simulation starts at random initial states and then receives the controller's action to update the states and estimate the numerical reward signal which represents the current states. Thereby, the simulation model gives the needed training data periodically. The simulation model includes two parts: the mass balance model which can update system states by iteratively calculating the mass balance equation; the energy consumption model which can estimate the energy consumption under the control feedback which can compute the reward signal used in the DDPG controller. The DDPG controller tries to directly map from the states to the optimal action and to use the reward signal by self-exploration in the ZEP simulation environment. The framework is shown in Fig. 6. A neural network with four fully connected layers is used in both actor network and critic network, which have 400 and 200 nodes in first hidden layer and second hidden layer, respectively. The network structure is similar with [37]. The structure of target actor network and target critic network is absolutely same with actor network and critic network, respectively.

In order to represent the ZEP environment we use the sensors related with the ZEP environment. Let $s = [s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}] = [C_{1,1}, C_{1,2}, C_{2,1}, C_{2,2}, C_{3,1}, C_{3,2}, C_{4,1}, C_{4,2}, V_C, \varepsilon]$ represent the states. This ten states can describe the ZEP environment fully. However, different states have different ranges, which makes it hard for the DDPG controller to learn the optimal result. Therefore before states input to the DDPG controller, we add a normalization layer, where states will scale to the range of $[-1, 1]$ to improve the training speed.

The DDPG controller's output action is the flow rate of the leaching solution whose value is set from 0 to 100 m^2/h. To reduce the exploration range, ensure the convergence and improve the speed of training, we correct set action range parameter. The control periods could not be so frequent that the DDPG controller cannot learn. In the practical ZEP, there exists the large time delay, which varies with the working condition. Thus the control periods should be 0.077 h, 0.068 h, 0.037 h, 0.11 h corresponding to four working conditions.

The agent receives a numerical reward signal r_t from the ZEP simulation environment at each control cycle. The reward signal is different between the optimal value with the real-time measurement. The reward function is defined as

$$r = \frac{\alpha - W}{\beta} \tag{11}$$

where α, β are the constant values, W is the energy consumption. Different forms of the reward functions will lead to different learning results. For the DDPG algorithm, if the reward function mean value is approximately zero, it is easier to learn the optimal strategy. So we adjust α, β parameters in Eq. (11) to drive the mean value of the reward function to zero. For the ZEP simulation environment, we let $\alpha = -3300$, $\beta = 10,000$ due to the fact that the energy consumption is about 3300 ($KW\,h\,t^{-1}$).

Due to the time-delay various with working conditions, we train different network parameters under different working conditions based on only one group hyperparameters. The ZEP simulation starts at a random initial states in the suitable range. Then we apply the mass balance model to update the states every one second. The DDPG controller gives its control output every 0.077h, 0.068h, 0.037h, 0.11h under four different working conditions.

5. Result analysis

In this section, we show the training and testing results of the DDPG controller. To further illustrate the superiority of the DDPG controller, we compare the DDPG controller with PI controller, MPC controller and artificial experiences, respectively. The PI controller is commonly used in the factory. The MPC is a model-based controller, which can now be found in a wide variety of application areas including chemicals, food processing, automotive, and aerospace applications. So we compare DDPG algorithm with above two controllers. The energy consumption trajectories of the DDPG controller is better than artificial experiences. Furthermore, the DDPG controller can provide a smoother control action than PI controller and MPC controller, which improves the system stability and can better adapt to the changes of system parameters.

5.1. Training results

The optimal control framework of the ZEP is constructed in the Python platform based on the tools of artificial neural network package Tensorflow. We only need about 500 training iterations to learn the optimal control strategy as shown in Fig. 7. For 500 iterations of training, it takes approximately 1 h to complete in Intel

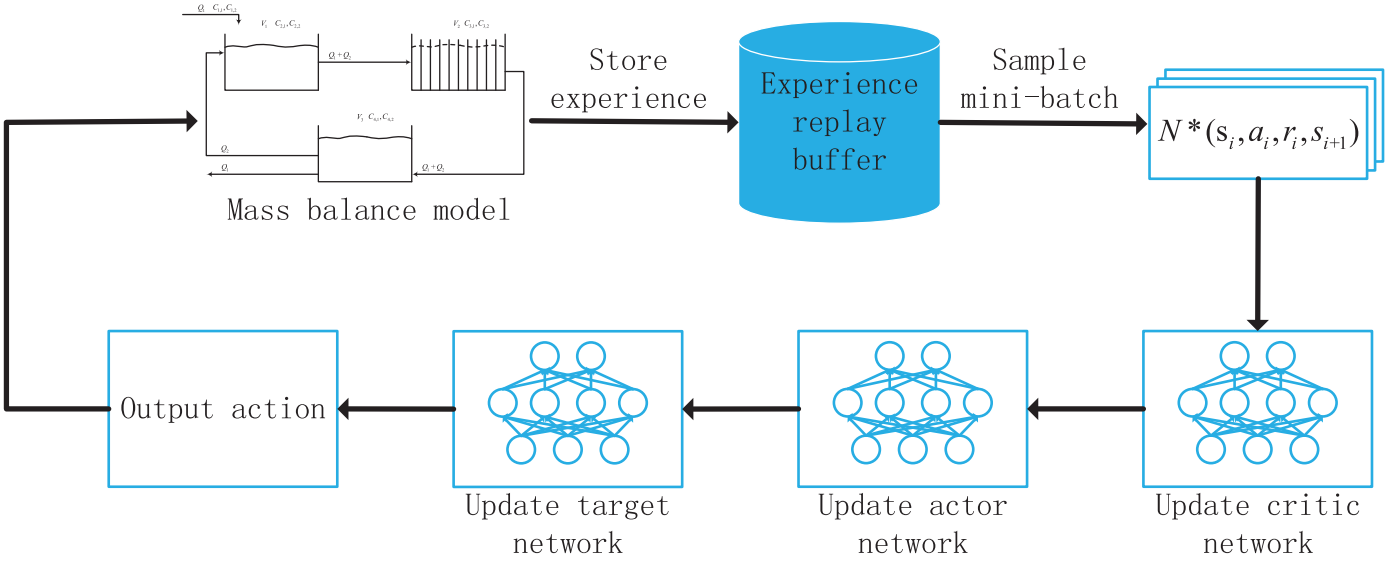


Fig. 6. The DRL optimal control framework.

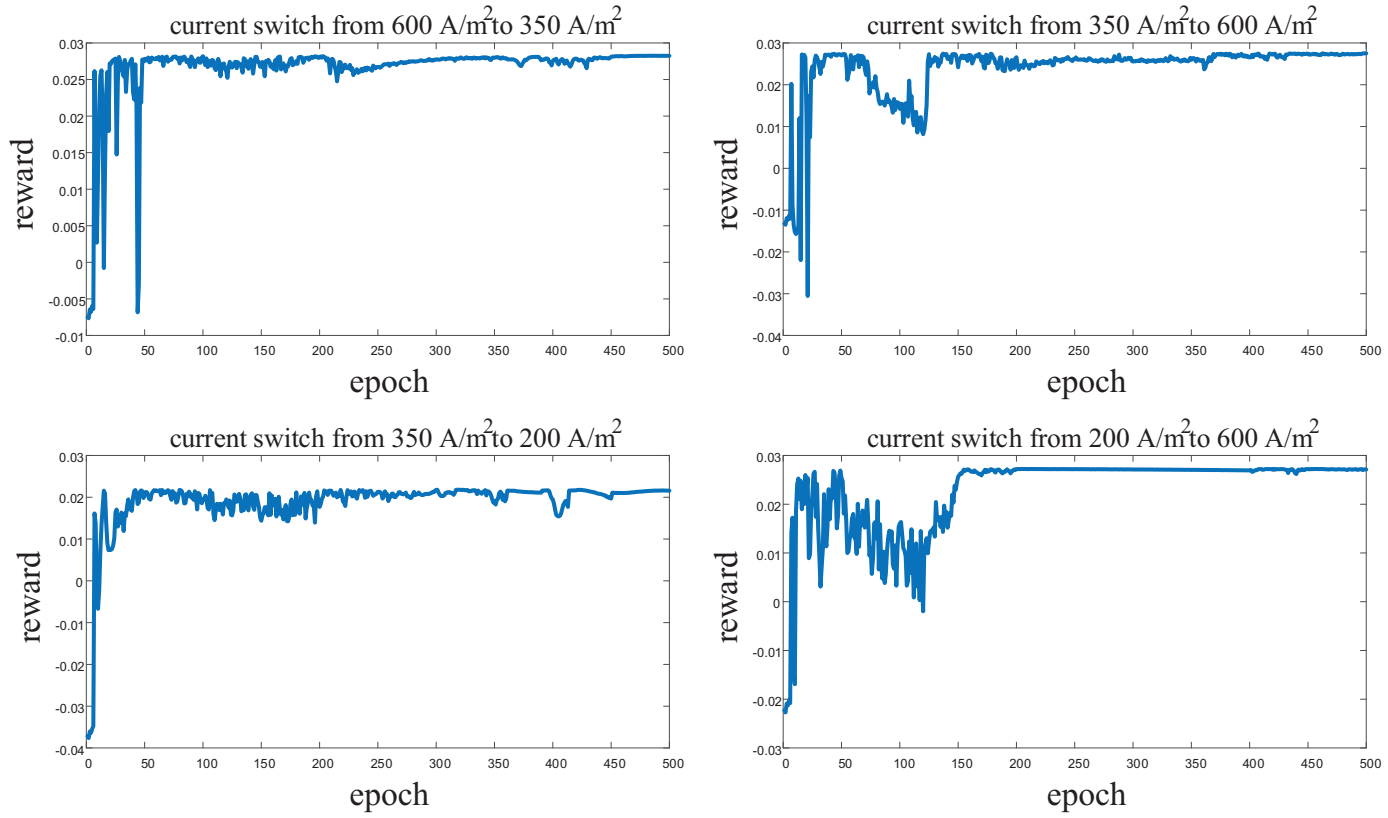


Fig. 7. Reward signal in the training.

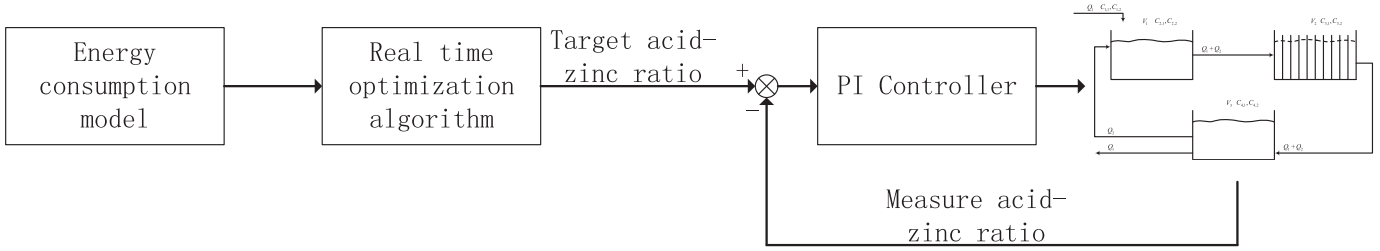


Fig. 8. The ZEP control framework with PI controller.

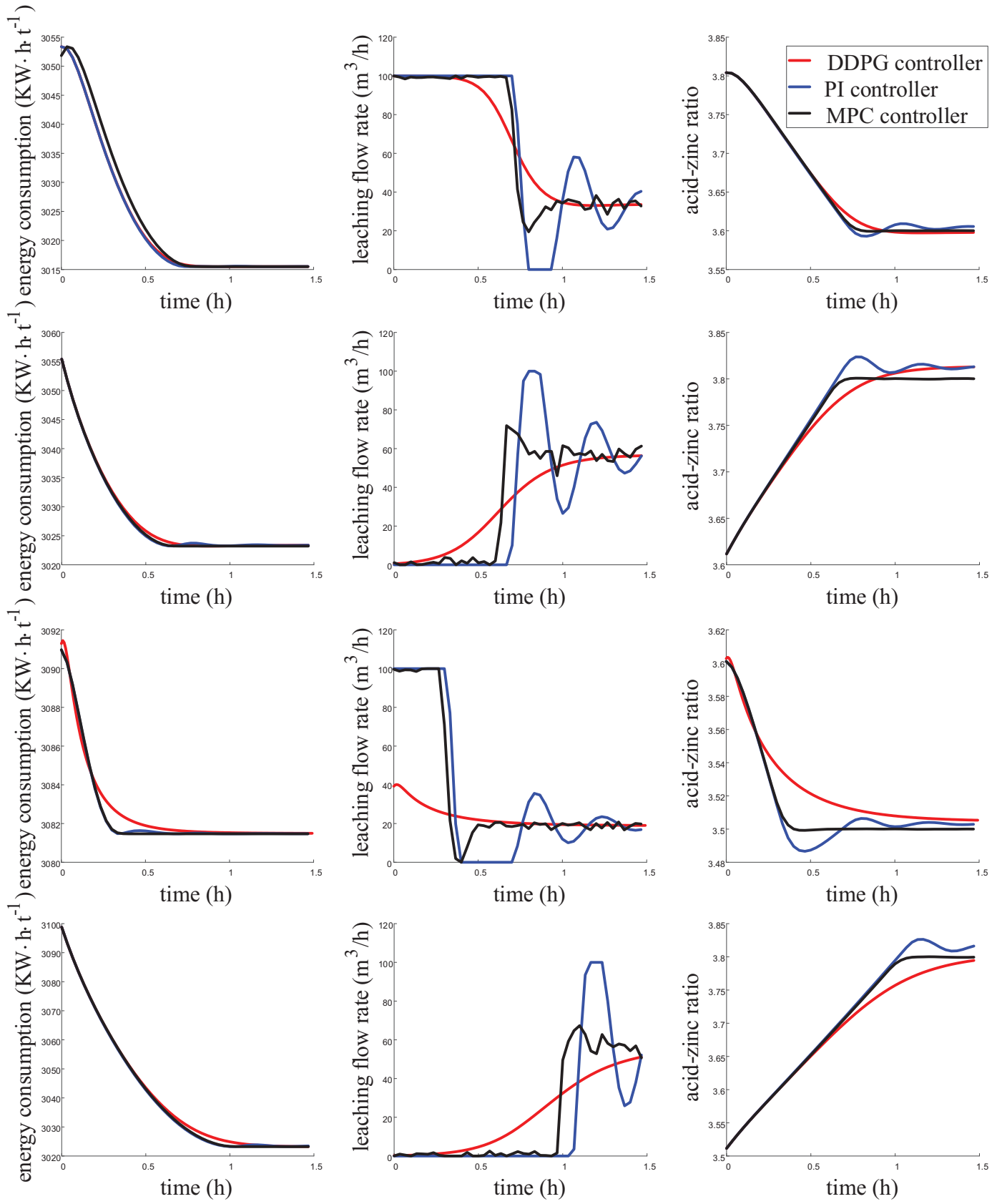


Fig. 9. The energy consumption, leaching flow rate, acid-zinc ratio with the DDPG controller and PI controller under the current density switching from 600A/m^2 to 350 A/m^2, 350 A/m^2 to 600 A/m^2, 350 A/m^2 to 200 A/m^2, 200A/m^2 to 600 A/m^2, respectively.

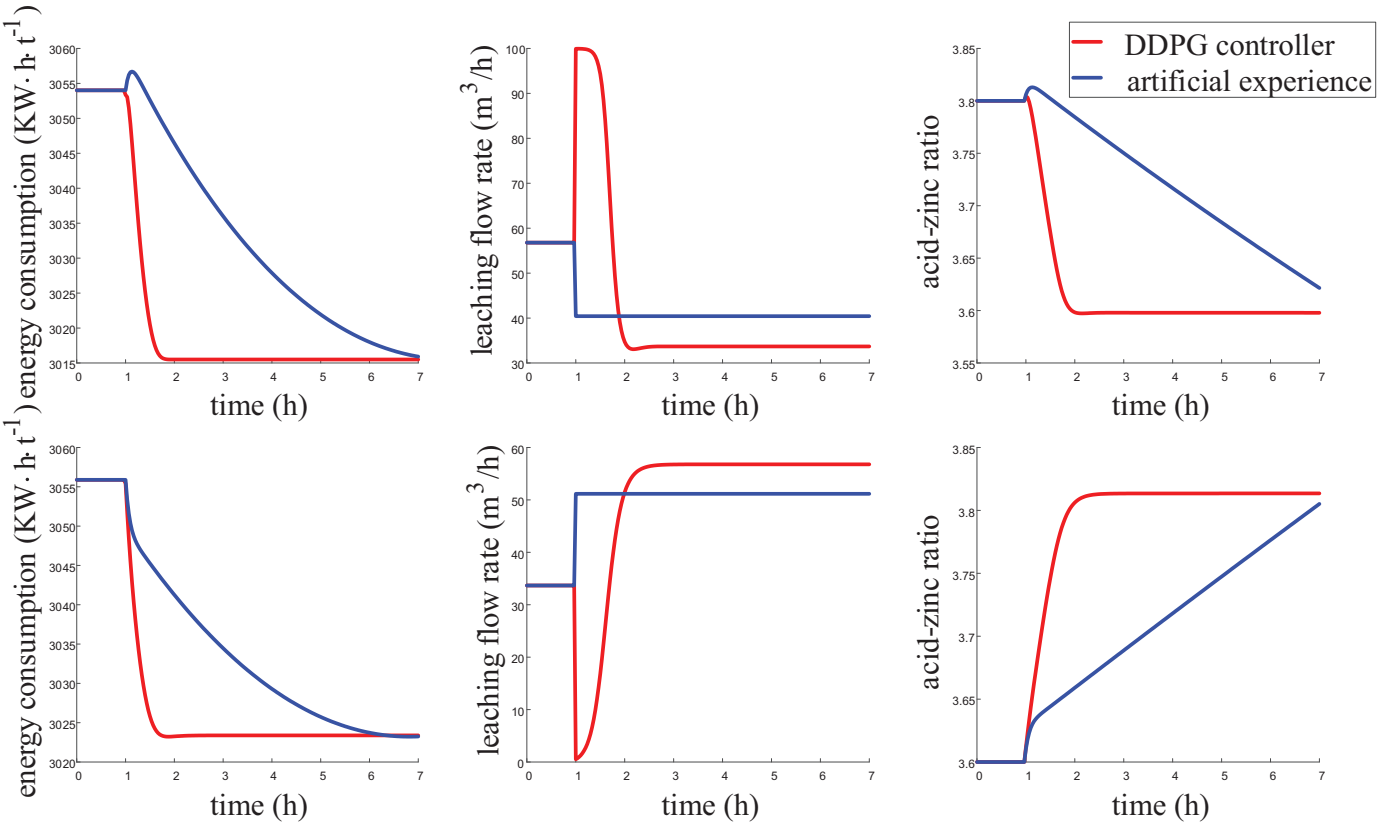


Fig. 10. The energy consumption, leaching flow rate, acid-zinc ratio with the DDPG controller and artificial experience under the current density switching from 600 A/m^2 to 350 A/m^2, 350 A/m^2 to 600 A/m^2, respectively.

Core i5-8400 CPU. So it need consume about 0.1 min to finish one iteration. The minimal control period is 0.077 h about 4.62 min. Therefore, we have enough time to complete the iteration. So we can do simulation in real time. Thus it is very quick to apply the DDPG controller to learn the optimal control strategy of the ZEP compared with [29] which need iterations about 1200–1400.

5.2. Testing results

To further illustrate the superiority of the DDPG controller, we compare the DDPG controller with the PI controller and MPC controller under four working conditions of the ZEP simulation environment. Specially, we use method in [39] to get the optimal parameters of PI controller with the help of genetic optimization algorithm. The implementation of the MPC controller is the same as [40].

We establish the ZEP control framework with the PI controller which includes four parts. In part 1, we develop the same energy consumption model with the framework of the DDPG controller. In part 2, we use the real-time optimization algorithm to search the optimal CZN and CH values under different current density cases which are set as the target values of the PI controller. In part 3, we use the difference between the target value and measure value as the PI controller input, the leaching flow as the PI controller output. In part 4, the ZEP simulation environment accepts the PI controller's action and update states. The framework structure is shown in Fig. 8. And the implementation of MPC controller is similar with PI controller. It is noted that, we add some noise in the predict model of MPC controller. It is reasonable that some error between predict model and practice system due to the inaccurate modeling.

We use the trained DDPG controller to control the leaching solution flow rate. From the testing results in Fig. 9 and 10, we know that the DDPG controller can control the ZEP to achieve the minimal energy consumption under the required time. The DDPG controller is better than the PI controller and MPC controller since it can provide smoother control actions, which improves the system stability. Furthermore, the performance of adjusting time and the mean value of the energy consumption under the DDPG controller is better than those under artificial experiences.

It is noticed in Fig. 10 that when the current density switches from 600A/m^2 to 350A/m^2, the acid-zinc ratio and leaching flow rate should reduce as well. However, in order to speed up the decline rate of acid-zinc ratio, the DDPG controller uses a high leaching flow rate at the initial phase of current switching, and a low leaching flow rate when the acid-zinc ratio enters steady state. A similar work is done when the current density switches from 350A/m^2 to 600A/m^2.

The trained DDPG controller has a strong capability to learn the optimal control strategy from the ZEP simulation environment by controlling the leaching solution flow rate to get the minimal energy consumption. In addition, the DDPG controller is a learning controller that is superior to the policy search algorithm in which it does not require a predefined controller structure, which limits the performance of the agent and increase the costs and human errors. Furthermore, the DDPG controller can gradually lean new optimal control strategy when the system model changes.

6. Conclusion

This paper presents the ZEP simulation environment and develops an improved DDPG controller for the ZEP. We formulate

the ZEP control problem to minimize the energy consumption by maximizing the long term discounted reward. By comparing, it demonstrates that the DDPG controller has a better performance than the PI controller and MPC controller. Furthermore, the energy consumption of the proposed controller is lower than that of the artificial experiences. It should be noted that the performance of the DDPG controller is not affected by the accuracy and assumptions of the ZEP simulation model.

In summary, the DDPG controller can finish complex tasks in the ZEP during current switching periods via self-learning without any prior knowledge. It is completely different from pre-existing control laws and easily applied to a new system which indicates that the DRL-based controller is a strong candidate for smart digitalized systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Zhang, C. Yang, H. Zhu, Y. Li, W. Gui, Kinetic modeling and parameter estimation for competing reactions in copper removal process from zinc sulfate solution, Ind. Eng. Chem. Res. 52 (48) (2013) 17074–17086.

[2] J. Bastian, J. Zhu, V. Banunarayanan, R. Mukerji, Forecasting energy prices in a competitive market, IEEE Comput. Appl. Power 12 (3) (1999) 40–45.

[3] S. Deng, C. Yang, Y. Li, H. Zhu, T. Wu, Spatiotemporal distribution model for zinc electrowinning process and its parameter estimation, J. Central South Univ. 24 (9) (2017) 1968–1976.

[4] G. Barton, A. Scott, Scale-up effects in modelling a full-size zinc electrowinning cell, J. Appl. Electrochem. 22 (8) (1992) 687–692.

[5] Y. Wang, W. Gui, C. Yang, T. Huang, Intelligent modeling and optimization on time-sharing power dispatching system for electrolytic zinc process, Trans. Nonferrous Metals Soc. China 10 (4) (2000) 561–565.

[6] B. Behnajady, J. Moghaddam, Statistical evaluation and optimization of zinc electrolyte hot purification process by Taguchi method, J. Central South Univ. 22 (6) (2015) 2066–2072.

[7] C. Yang, G. Deconinck, W. Gui, Y. Li, An optimal power-dispatching system using neural networks for the electrochemical process of zinc depending on varying prices of electricity, IEEE Trans. Neural Netw. 13 (1) (2002) 229–236.

[8] G. Barton, A. Scott, Industrial applications of a mathematical model for the zinc electrowinning process, J. Appl. Electrochem. 24 (5) (1994) 377–383.

[9] M. Mahon, S. Peng, A. Alfantazi, Application and optimisation studies of a zinc electrowinning process simulation, Can. J. Chem. Eng. 92 (4) (2014) 633–642.

[10] C. Yang, S. Deng, Y. Li, H. Zhu, F. Li, Optimal control for zinc electrowinning process with current switching, IEEE Access 5 (99) (2017) 24688–24697.

[11] H. Xu, X. Wang, Optimization and Control Methods in Industrial Engineering and Construction, Springer, 2014.

[12] J. Ye, H. Xu, E. Feng, Z. Xiu, Optimization of a fed-batch bioreactor for 1, 3-propanediol production using hybrid nonlinear optimal control, J. Process. Control 24 (10) (2014) 1556–1569.

[13] J. Yuan, J. Xie, H. Xu, E. Feng, Z. Xiu, Optimization for nonlinear uncertain switched stochastic systems with initial state difference in batch culture process, Complexity 2019 (2019) 1–15 Article ID 4979580.

[14] D. Yang, G. Zong, H. Karimi, h_∞ Refined anti-disturbance control of switched lpv systems with application to aero-engine, IEEE Trans. Ind. Electron. (2019), doi:10.1109/TIE.2019.2912780.

[15] H. Ren, G. Zong, T. Li, Event-triggered finite-time control for networked switched linear systems with asynchronous switching, IEEE Trans. Syst. Man Cybern.: Syst. 48 (11) (2018) 1874–1884.

[16] H. Ren, G. Zong, H. Karimi, Asynchronous finite-time filtering of networked switched systems and its application: an event-driven method, IEEE Trans. Circuits Syst. I Regul. Pap. 66 (1) (2018) 391–402.

[17] H. Xu, S. Wang, S. Wu, Optimization Methods, Theory and Applications, Springer, 2015.

[18] Y. Wu, R. Lu, Output synchronization and l_2-gain analysis for network systems, IEEE Trans. Syst. Man Cybern.: Syst. 48 (12) (2017) 2105–2114.

[19] B. Sun, M. He, Y. Wang, W. Gui, C. Yang, Q. Zhu, A data-driven optimal control approach for solution purification process, J. Process. Control 68 (2018) 171–185.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.

[22] Y. Wang, H. Shen, D. Duan, On stabilization of quantized sampled-data neural-network-based control systems, IEEE Trans. Cybern. 47 (10) (2016) 3124–3135.

[23] Q. Zhou, P. Shi, H. Liu, S. Xu, Neural-network-based decentralized adaptive output-feedback control for large-scale stochastic nonlinear systems, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 42 (6) (2012) 1608–1619.

[24] D. Ernst, M. Glavic, F. Capitanescu, L. Wehenkel, Reinforcement learning versus model predictive control: a comparison on a power system problem, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 39 (2) (2008) 517–529.

[25] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of go with deep neural networks and tree search, Nature 529 (7587) (2016) 484.

[26] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al., Mastering the game of go without human knowledge, Nature 550 (7676) (2017) 354.

[27] S. Li, L. Ding, H. Gao, C. Chen, Z. Liu, Z. Deng, Adaptive neural network tracking control-based reinforcement learning for wheeled mobile robots with skidding and slipping, Neurocomputing 283 (2018) 20–30.

[28] F. Li, Q. Jiang, S. Zhang, M. Wei, R. Song, Robot skill acquisition in assembly process using deep reinforcement learning, Neurocomputing 345 (2019) 92–102.

[29] Y. Ma, W. Zhu, M.G. Benton, J. Romagnoli, Continuous control of a polymerization system with deep reinforcement learning, J. Process. Control 75 (2019) 40–47.

[30] C. Qiu, Y. Hu, Y. Chen, B. Zeng, Deep deterministic policy gradient (ddpg) based energy harvesting wireless communications, IEEE Internet Things J. 6 (5) (2019) 8577–8588.

[31] L. Liu, E. Chen, Z. Gao, Y. Wang, Research on motion planning of seven degree of freedom manipulator based on ddpg, in: Proceedings of International Workshop of Advanced Manufacturing and Automation, Springer, 2018, pp. 356–367.

[32] S. Gu, E. Holly, T. Lillicrap, S. Levine, Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates, in: Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017, pp. 3389–3396.

[33] B. Zhang, C. Yang, Y. Li, X. Wang, H. Zhu, W. Gui, Additive requirement ratio prediction using trend distribution features for hydrometallurgical purification processes, Control Eng. Pract. 46 (2016) 10–25.

[34] A. Saba, A. Elsherief, Continuous electrowinning of zinc, Hydrometallurgy 54 (2–3) (2000) 91–106.

[35] H. Hoang, F. Couenne, C. Jallut, Y. Le Gorrec, The port hamiltonian approach to modeling and control of continuous stirred tank reactors, J. Process. Control 21 (10) (2011) 1449–1458.

[36] G. Barton, A. Scott, A validated mathematical model for a zinc electrowinning cell, J. Appl. Electrochem. 22 (2) (1992) 104–115.

[37] T.P. Lillicrap, J.J. Hunt, A. Pritzel, et al. Continuous control with deep reinforcement learning: U.S. Patent Application 15/217,758[P]. 2017-1-26.

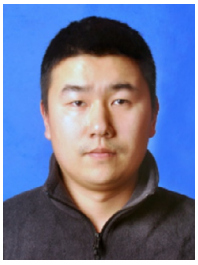
[38] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529.

[39] P. Wang, D. Kwok, Optimal design of pid process controllers based on genetic algorithms, Control Eng. Pract. 2 (4) (1994) 641–648.

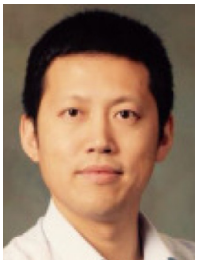
[40] S. Qin, T. Badgwell, A survey of industrial model predictive control technology, Control Eng. Pract. 11 (7) (2003) 733–764.

Xiongtao Shi was born in Handan, China, in 1994. He is currently pursuing the M.S. degree with the School of Automation, Central South University, China. His current research interests include reinforcement learning, multi-agent systems and networked control systems.

Yonggang Li (M'09) received the M.S. degree in control science and engineering and the Ph.D. degree in control science and engineering from Central South University, Changsha, China, in 2000 and 2004, respectively. From 2011 to 2012, he was a Visiting Scholar with the Curtin University, Perth, Australia. Since 2013, he has been a Full Professor with the School of Information Science and Engineering, Central South University. His current research interests include modeling and optimal control of complex industrial process, process control, intelligent control system, and knowledge driven automation.



Bei Sun (M'17) received his Ph.D. in Control Science and Engineering from Central South University, China in 2015, and was with the Department of Electrical and Computer Engineering, Polytechnic School of Engineering, New York University, United States from 2012 to 2014. He is currently an associate professor of Central South University. His research interests include datadriven modeling, optimization and control of complex industrial processes.



Honglei Xu received the B.Eng., M.Eng., and Ph.D. degr ees from the Huazhong University of Science and Technology, Wuhan, China, and Curtin University, Perth, WA, Australia, in 1997, 2002, 2005, and 2009, respectively. He is now a senior lecturer in the Department of Mathematics and Statistics, Curtin University. He held professorship positions in various universities including Central South University and Huazhong University of Science and Technology. His current research interests include hybrid systems and optimization with applications. He acts as an Assistant Editor of the Journal of Dynamics of Continuous, Discrete and Impulsive Systems, Series B: Application and Algorithms and an editorial board member or Guest Editor of other six international journals. He acted as stream chair or program member at numerous major international conferences.

Chunhua Yang (M'09) received the M.Eng. degree in automatic control engineering and the Ph.D. degree in control science and engineering from Central South University, Changsha, China, in 1988 and 2002, respectively. She is currently a Professor with Central South University. From 1999 to 2001, she was a Visiting Professor with the Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium. Her research interests include modeling and optimal control of complex industrial processes, intelligent control systems, and fault-tolerant control of real-time systems.

Hongqiu Zhu obtained his MSc and PHD degree in control theory and application from Central South University, China, in 2002 and 2010 respectively. He worked in Central South University from 2006. His research work includes information processing and optimal control of complex industrial process.