International Journal of Applied Mathematics in Control Engineering 6 (2023) 82-92

Contents lists available at YXpublications

International Journal of Applied Mathematics in Control Engineering

Journal homepage: http://www.ijamce.com

Research on Scheduling Optimization Problem of Molten Steel Hit Rate Based on Markov

Jiayu Peng[a,*], Tingwei Pan[a], Jiaxin Guo[b], Xinqi Hu[c], Congxin Li[a] , Jingjing Lou[a], Liangliang Sun[d]

[a] School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, Liaoning 110168, China
[c] School of Transportation Engineering, Dalian Jiaotong University, Dalian 116045, China
[c] School of Electrical and Contral Engineering, Shenyang Jianzhu University, Shenyang, Liaoning 110168, China
[d] School of Control Engineering, Northeastern University at Qinghuangdao, Qinghuangdao, Hebei 066004, China

ARTICLE INFO

Article history:
Received 5 March 2023
Accepted 13 April 2023
Available online 20 April 2023

Keywords：
Steel refining process
Steel hit rate
Markov chain
Q learning

ABSTRACT

The steel refining process is characterized by multi-stage, multi-equipment and multi-constraints, and there is the problem of uncertainty in the steel hit rate. The steel in the refining process in units of charge is often reworked due to low steel hit rate, so the mathematical model is difficult to describe accurately. Based on this feature, a Markov chain-based stochastic evolution mathematical model of the steel refining process was developed. In addition, based on the fact that an increase in the number of refineries in the steel refining process will lead to a huge increase in computational difficulty, and considering the requirements of the refining process for the efficiency of the scheduling optimization problem in an uncertain environment, we designed an algorithm solution framework based on heuristic simulation strategy and improved Q learning and conducted simulation experiments based on the data of a steel company to prove the effectiveness of the proposed scheme.

1. Introduction

The steel industry is the foundation of the national economy. As the market's demand for steel continues to increase and the requirements for steel production efficiency continue to increase, the optimization of steel production technology has become a crucial issue. Earlier studies (Tan Y., et al., 2013) showed as the bottleneck of the entire steel production, the steelmaking-continuous casting process has a huge impact on the entire steel production. The refining link is the intermediate link between steelmaking and continuous casting. Due to the complex production equipment in the refining process, it is difficult to ensure the stability of the molten steel composition treatment in the production unit. The molten steel hit rate may not meet the standard (that is, from certain refining, the composition of molten steel based on heat from the process did not meet the established requirements), which led to the occurrence of re-smelting, which seriously affected the normal production of subsequent heats, affected the efficiency of steel production, increased energy consumption, and increased the use of personnel costs. Therefore, it is important to propose a study on the uncertain scheduling method for steel hit rate in the refining production process to ensure the smooth running of steel production.

* Corresponding author.
E-mail addresses: pjyjy1996@163.com (J. Peng)
doi:

2. Mathematical model

2.1 Description of Refining production scheduling problem

The refining process is a multi-equipment and multi-stage production process based on heat. Each process in each heat only occupies one refining equipment for production at the same time, and one production equipment can only process one at the same time. The molten steel of the heat is produced. There are multiple refining processes in the refining process, with multiple refining process paths and multiple possible refining production states, that is, the refining production state, initial processing time, The completion time and the quality of the completion (whether it needs to be re-smelted) are uncertain. The uncertainty of the refining process is only affected by the previous refining process produced with the same refining equipment. According to the characteristics of the refining process and the research on steelmaking-continuous casting production, mathematical equations are used to describe the two performance indicators of the refining production scheduling problem: (1) In the refining process, each heat is cast on the continuous casting machine. The deviation between the ideal pouring time and the actual pouring time is similar. The deviation between the ideal pouring time and the actual pouring time is divided into two situations:

① The actual pouring time of heat i on the corresponding continuous casting machine is ahead of the ideal pouring time T_i of the continuous casting machine. ②The actual pouring time of heat i on the corresponding continuous casting machine lags behind the ideal pouring time T_i of the continuous casting machine, as shown in equation (1). (2) The sum of the waiting time of the heat in each process is the smallest, because this article only studies the refining process, so the refining production stage process number $j'=2,3,4$. Equation (2) is listed, which means that the sum of waiting time for heat i in the refining production stage is the smallest.

$$Min\ f_1 = \sum_{i=1}^{D} \sum_{l=1,2,3} |T_i - M_{li}| \tag{1}$$

In the equation, M_{li} represents the starting time of the lth pouring and the l-th heat on the corresponding continuous casting machine, $i \in \{1, \dots, D\}$ is a positive integer; T_i represents the ideal pouring time of the heat i on the corresponding continuous casting machine.

$$Min\ f_2 = \sum_{i=1}^{D} \sum_{j'=2,3,4} C2_i (F_{ij'+1} - F_{ij'} - W_{ij'}) \tag{2}$$

In the equation, $C2_i$ represents the penalty coefficient of the unit waiting time of heat i, $i \in \{1, \dots, D\}$ is a positive integer; $F_{ij'}$ represents the j'th refining process of the i heat Production start time;

Subject to:

$$F_{ij+1} \geq F_{ij}. \tag{3}$$

$$\sum_{j=1}^{j'} x_{ijo} \leq 1,\ i \in [1, D], j' = [2,3,4],\ x_{ijo} = 0 \text{ or } 1 \tag{4}$$

$$t_{ij+1} - t_{ij} \geq 0 \tag{5}$$

Equation (3) means that the next procedure can only start production after the previous procedure of the same heat is completed; equation (4) means that each procedure of each heat can only use one piece of equipment for production at the same time; equation (5) means the same on the refining equipment, the next refining process can only be produced after the previous refining process is completed.

2.2 Description of Uncertain Scheduling Problem of Molten Steel Hit Rate

Because the refining process is a multi-stage, multi-equipment production stage, there are multiple possible production completion methods and multiple refining process paths for each refining process. According to the production experience and historical production data of the steel plant, the approximate completion of each refining process can be obtained, and the completion method of each refining production is described by the refining production state, the initial processing time of the refining process and the quality of molten steel transfer. In the refining production situation of a certain heat refining process described in Figure 1, the three possible situations of the molten steel hit rate in the first refining process are (A_{11}, B_{11}, E), (A_{12}, B_{12}, G_2), (A_{13}, B_{13}, G_2), A represents the state of refining production, B represents the initial processing time of the refining process, E means the quality is good, and G means the quality is qualified. The relationship between adjacent refining steps is described by the transition probability matrix. Step 2 also has three molten steel hit rate states. Therefore, the size of the transition probability matrix connecting step 1 and step 2 is 3×3, which means that under the premise that a certain refining processing method completes step 1, the refining process will change from a certain molten steel hit rate state in step 1 to a certain molten steel hit rate state in step 2. Because the refining process has the characteristics of multiple equipment and multiple processes, there must be multiple refining process paths for each heat. At the same time, the refining production system is uncertain because of various disturbance factors, which may lead to refining conditions during the refining process. If in the actual refining process, there is a situation of refining and refining, it will further expand the number of refining process paths for each heat. From the above description of the refining process, we can know that in the actual refining process, only one of the many refining process paths will be selected as the production process path of the actual refining production. The situation of refining is indirectly expanded. The selection range of the optimal refining process path. Here, the state of molten steel hit rate presented in the second process is derived from the state of molten steel hit rate presented in the first process and the refining equipment used in the first process. The relationship between them is an implicit functional relationship.

Any heat needs to obtain a feasible scheduling plan through scientific dynamic scheduling methods before refining production, make full use of refining production equipment, and shorten the refining process as much as possible on the premise of ensuring that the refining production tasks are completed on time and the quality of molten steel is qualified. Increase the production time of steel and improve the efficiency of steel production.

Figure 1 is a model diagram of uncertain molten steel hit rate parameters.

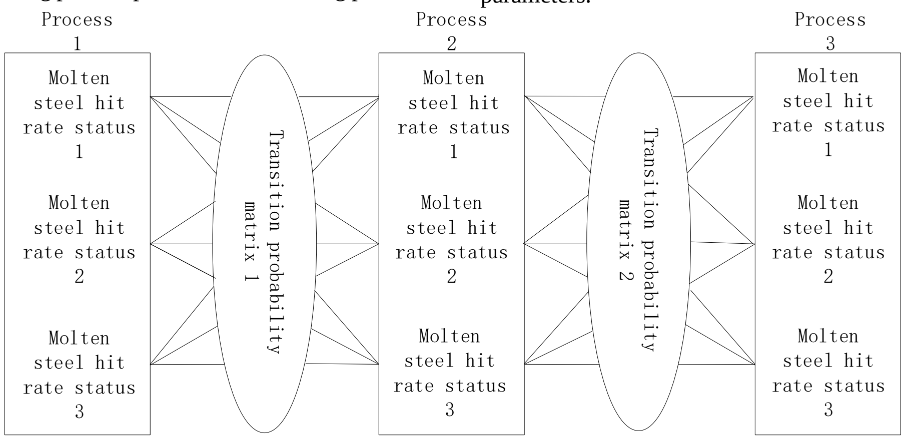

Fig. 1 Parameter model of uncertainty of molten steel hit rate

Suppose a steel refining process has 3 refining processes, as shown in Figure 1, the numbers of the processes are 1,2, and 3. The hit rate status of the molten steel produced by the first refining process is: a, c, e; the hit rate status of the molten steel produced by the second refining process is: b, d, m; the output of the third refining process the state of the hit rate of molten steel is: s_1, s_2, s_3. From Figure 1, we can see that each refining process has three possible molten steel hit rate states, and the state transition relationship of molten steel hit rate between adjacent steel refining steps of the same heat needs to be represented by a state transition probability matrix. For example, both refining stage 1 and refining stage 2 have three possible molten steel hit rate states, and the size of the state transition probability matrix used to describe the relationship between refining step 1(w_1) and refining step 2(w_2) is 3×3.

Representing the probability of turning to a certain state of molten steel hit rate in refining step 2 under the premise of a certain state of molten steel hit rate in refining step 1, it is expressed by a mathematical equation as shown in equation (6):

$$p_{w_1 w_2}^{a} = \mathrm{E}(w_2 | w_1, a). \tag{6}$$

Where, $p_{w_1 w_2}^{a}$ represents the transition probability of performing production action a from the state of the hit rate of molten steel produced in refining process 1(w_1) to refining process 2(w_2), w_1 indicates the state of molten steel hit rate in refining process 1, w_2 represents the state of molten steel hit rate in process 2, a represents the execution of refined production actions, E represents the mathematical expectation of performing refining production action a from the state of the hit rate of molten steel produced in refining process 1(w_1) to refining process 2(w_2).

Using equation (6), the state transition probability of molten steel hit rate between all adjacent processes can be calculated, and the transition probability matrix $\pi_{12(3\times3)}$ can be calculated, as shown in equation (7). The transition probability matrix is related to two adjacent refining processes. If the molten steel hit rate status of a certain refining process and the state transition matrix between two adjacent refining processes can be obtained, the molten steel hit rate status of any refining process can be obtained, and finally Obtain the molten steel hit rate status of the steel refining process in all refining processes.

$$\pi_{12(3\times3)} = \begin{bmatrix} p_{ab} & p_{ad} & p_{ae} \\ p_{cb} & p_{cd} & p_{cm} \\ p_{eb} & p_{ed} & p_{em} \end{bmatrix} \tag{7}$$

Where $\pi_{12(3\times3)}$ represents the transition probability matrix from refining process 1 with a certain molten steel hit rate state of a、c、e to refining step 2 with a certain molten steel hit rate state of b、d、m, The dimension of the matrix is 3×3. The core problem described by using the Markov chain is how to determine the transition probability matrix under the condition of uncertainty in the hit rate of molten steel, and use the Markov chain to derive the refining production state description function required for subsequent model construction, as shown in equation (8) Shown, as well as the refining process execution description function, as shown in equation (9).

$$V_{\pi12}\big(x(k)\big) = \sum_{u(k)\in U} P\big(u(k)\big|x(k)\big) M_{x(k)}^{u(k)} + \gamma \sum_{x(k+1)\in X} P_{x(k)x(k+1)}^{u(k)} v_{\pi}\big(x(k+1)\big) \tag{8}$$

Where, $V_{\pi12}\big(x(k)\big)$ represents the refined production state $x(k)$ the refined production state description function, $x(k)$ represents the state of refined production, $u(k)$ represents the state of execution of refined production, $M_{x(k)}^{u(k)}$ represents the production time of the refined production action u_k in the refined production state x_k, γ means discount factor, $0 \le \gamma \le 1$, $P_{x(k)x(k+1)}^{u(k)}$ represents the probability of transferring from the refined production state x_k to the refined production state $x(k+1)$ from the refined production action u_k, $v_\pi(x(k+1)$ represents the refined production state $x(k+1)$ the refined production state description function.

$$L_{\pi12}(x(k),u(k)) = M_{x(k)}^{u(k)} + \gamma \sum_{x(k+1)\in X} P_{x(k)x(k+1)}^{u(k)} \sum_{u(k+1)\in U} \pi(u(k+1) \,|x(k+1)) L_{\pi}(u(k+1), s(k+1)) \tag{9}$$

Where $M_{x(k)}^{u(k)}$ represents the production time when the refining action u_k is executed in the refining production state x_k, γ means discount factor, $0 \le \gamma \le 1$, $P_{x(k)x(k+1)}^{u(k)}$ represents the probability of transferring from the refined production state x_k to the refined production state $x(k+1)$ from the refined production action u_k, $\pi\big(u(k+1)\big|x(k+1)\big)$ represents the probability of the refined production state $x(k+1)$ performing the refined production action $u(k+1)(\pi\big(u(k+1)\big|x(k+1)\big) = p\{x(k+1)\big), [u(k+1)]\}$. The percentage of the total heats (the hit rate of molten steel) produced by the molten steel produced by a refining process that meets the customer's product requirements may be expressed in mathematical equations as shown in equation (10).

$$P_z(j+1) = \sum_{z=S_1}^{S_3} \pi_{yz} P_y(j) \tag{10}$$

Where $P_z(j+1)$represents the probability that the $j+1$ th process turns to the molten steel hit rate state z ($z \in \{s_1, s_2, s_3\}$), π_{yz} represents the probability matrix of transition from molten steel hit rate state y to molten steel hit rate state z, $P_y(j)$ represents the probability that the j th steel refining process turns to the molten steel hit rate state y ($y \in \{b, d, f\}$), and j is a positive integer.

2.3 Mathematical model construction

Earlier studies (Sun L. L., et al., 2020) showed in the actual refining process, the transfer of the molten steel hit rate state from the current refining stage to the next refining stage is not only related to the previous refining stage, but also related to the molten steel hit rate state in the previous refining stage.

This will cause the built-up molten steel hit rate transition model to be too complex and even difficult to model. Therefore, we need to simplify the state transition model of molten steel hit rate in different refining production stages. The simplification method used in this article is to assume that the state transition of the molten steel hit rate in different refining stages is Markovian, that is, it is assumed that the state transition of the molten steel hit rate depends only on the current state and the state of the next refining stage, which is different from the previous refining stage. Irrelevant.

Markov chain is a model that describes the random process of a series of events. Earlier studies (Nishi T, et al., 2009) showed the probability of each event in this series depends only on the state achieved by the previous event. In the related fields of probability theory and mathematics After an in-depth study by the Russian mathematician Andrey Markov, the Markov process proposed is a random process with no memory and no aftereffect. The future state of the process has nothing to do with its past state. If a research process needs to use the current state to predict the state that may occur in the future, we say that the research process satisfies the Markov property, that is, it takes the current state of the research system as the condition, and its past and future states are independent. Markov chain is also a kind of Markov process. In a random process, each research unit has no influence on its subsequent research units. It is a random process with no memory and no aftereffect. Earlier studies (Nishi T, Inuiguchi M, 2007) showed in the known current state, the future state of the process has nothing to do with its past state. It has discrete state space or discrete index set, but the precise definition of Markov chain is different. Usually, Markov chain is defined as a discrete or continuous time Markov process with countable state space. In today's world, there are many models established by Markov chain for process statistics. Earlier studies (Hoitomt D. J., et al, 1993; Meng J., et al., 1995) showed They are mainly used to study cruise control systems of motor vehicles, customer queues or lines arriving at airports, storage systems such as dams, and population growth of certain animal species. the study.

Definition of parameters and symbols in the mathematical model:

D: The number of refining furnaces that the steel company under takes at the same time, D is a positive integer.

E: The number of types of refining equipment, E is a positive integer.

J: the total number of processes, J is a positive integer.

i: serial number, $i \in \{1, \ldots, D\}$, i is a positive integer, D is the total number of furnaces.

j: serial number, a set containing all processes, $1 \le j \le J$, j is a positive integer.

l: serial number, which is a set containing pouring times, l is a positive integer.

L_{li}: serial number, which is the serial number of the l th heat of the i th pour.

B_j: indicates the number of equipment in the j th refining process, $B_j \ge 1$, B_j is a positive integer.

H_k: the quantity of each type of refining production equipment, k is the type of refining production equipment, and k is a positive integer.

s_i: current processing status of heat i $i \in \{1, \ldots, D\}$), $s_i$is a positive integer.

q_i: The number of the completion method of the refining process closest to the current moment in heat i, $q_i \in \{1, \ldots, C_{in}\}$, q_i is a positive integer.

C_{in}: indicates the number of possible ways to complete the n th refining process of the i th heat, C_{in} is an integer.

R_k: The number of idle k th refining equipment in the current refining production state, R_k is an integer.

t: current moment.

t_{ij}: the start processing time of heat i in the jth refining process.

P_{ij}: the production time of heat i in the j th refining process.

U: Define the execution status of a certain steel refining process in the heat.

β_i: the execution status of a certain steel refining process of the i th heat, β_i=0 means that a certain refining process of the i th heat is not executed, and β_i=1 means that the i th heat is executed A refining process.

$x(k)$: refined production status.

$u(k)$: Refining production actions taken.

$x'(k+1)$: The current refining production state is $x(k)$, and the execution of refining production $u(k)$ is transferred to a temporary refining production state.

T_i: the ideal pouring time of the i th heat.

T_φ: the processing time of a certain heat in the refining stage.

T_ω: the waiting time of a certain heat in the refining stage.

Where we define the state of the refining process scheduling system as follows:

$$X = [s_1, s_2, \ldots, s_D, q_1, q_2, \ldots, q_D, R_1, R_2, \ldots, R_E, t]^T \tag{11}$$

Where s_i is a positive integer, representing the current refining production status of the i th heat, $i \in \{1, \ldots, D\}$, q_i is an integer, which means the number of ways to complete the production task in a refining process of the i th heat, R_j represents the number of refining and processing equipment of type j that is not occupied at the current moment, and $R_j \in \{0, \ldots, H_j\}$ is a positive integer, ($j \in \{1, \ldots, E\}$, j is a positive integer).

We define the execution status of the production process in the refining process as:

$$U = [\beta_1, \beta_2, \ldots, \beta_D]^T \tag{12}$$

Where U represents the production execution status of each heat in the refining process, β_i represents the execution status of a certain production process in the ith heat, $\beta_i = 1$ or 0, ($i \in \{1, \ldots, D\}, i$ is a positive integer), $\beta_i = 1$ means that a certain process of the i th heat is executed, $\beta_i = 0$ means that a certain refining process of the i th heat will not be executed.

$$x(0) = [1, \ldots, 1, 0, \ldots, 0, R_1, \ldots, R_E, 0]^T \tag{13}$$

Where $x(0)$ represents the initial state of the refined production scheduling system, 1 indicates the status of the production scheduling system in which each heat is ready to execute the first refining process, 0 means the completion method of the first refining process for each heat, $R_j \in \{0, \ldots, H_j\}, j \in \{1, \ldots, E\}, j$ is a positive integer, R'_j is a positive integer), Indicates the number of type j refining and processing equipment that is not occupied at the current moment.

System status — Way of completion — Equipment usage — time

$$x(0) = [1, ..., 1, 0, ..., 0, R_1, ..., R_E, 0]^T \quad (14)$$

Establish the following mathematical objective function, $Q\big(x(k), u(k)\big)$ represents the start processing time when the refining production action $u(k)$ is executed in the refining production state $x(k)$, where $Q(x(k+1), u(k+1))$ represents the start processing time of the set of all refining production actions that can be selected in the refining production state $x(k+1)$, $g(x(k), u(k), x(k+1))$ is the steel refining production action $u(k)$ from the refining production state $x(k)$ to the steel refining production state $x(k+1)$ time:

$$\text{Min}\big(Qx(k), u(k)\big) = \alpha \text{Min}_{u(k+1)\in U^{x(k+1)}} \text{E}\big[Q\big(x(k+1), u(k+1)\big)\big](1-\gamma)Q\big(x(k), u(k)\big) + \gamma\{[g(x(k), u(k), x(k+1)\} + \big|Q\big(x(k), u(k)\big) - T_i\big| \quad (15)$$

Where $\text{Min}Q\big(x(k), u(k)\big)$ represents the minimum processing time for the refining production system to execute the refining production action $u(k)$ in the refining production state $x(k)$, $Q\big(x(k), u(k)\big)$ represents the processing time for the current refining production system to execute the refining production action $u(k)$ in the refining production state $x(k)$, $\text{g}(x(k), u(k), x(k+1))$ means from the refined production state $x(k)$ to execute the refined production action $u(k)$ to the refined production state $x(k+1)$ time, $\text{Min}_{u(k+1)\in U^{x(k+1)}} \text{E}\big[Q\big(x(k+1), u(k+1)\big)\big]$ means that it is in refining production state $x(k+1)$ the minimum value of the expected value of the mathematics of the processing time $u(k+1)$ the execution of the refined production action, γ represents the discount factor ($\gamma \in \{0, ..., 1\}$), α represents the learning coefficient ($\alpha \in \{0, ..., 1\}$), $\big|Q\big(x(k), u(k)\big) - T_i\big|$ |Indicates that the refining production system executes the refining production action in the refining production state $x(k)\, u(k)$ the difference between the ideal pouring time and the actual pouring time, T_i represents the ideal pouring time of heat i ($k \in \{1, ..., E\}$, $i \in \{1, ..., D\}$, k and i are both positive integers), the ideal pouring time of heat i needs Satisfy:

$$T_i > T_\varphi + T_\omega \quad (16)$$

T_φ represents the processing time of the heat on the refining equipment, T_ω represents the sum of the waiting time for processing on the refining equipment for the adjacent processes of the heat. Further explain the relationship between pouring times and heats, and use calculation examples to briefly explain. A steel production enterprise undertakes the refining production tasks of 5 heats at the same time, and divides them into 2 pouring cycles to complete the production tasks of the subsequent continuous casting stage. An example table of the relationship between the number of heats and the number of castings in the production process of steelmaking-refining-continuous casting is shown in Table 1.

It is stipulated that the ideal pouring time of L_{11} is 8:00 and the processing time is 10 minutes; the ideal pouring time of L_{13} is 8:10 and the processing time is 10 minutes; the ideal pouring time of L_{15} is 8:20 and the processing time is 10 minutes; L_{21} The ideal pouring time is 9:00, and the processing time is 20 minutes; the ideal pouring time of L_{22} is 9:40, and the processing time is 20 minutes. According to the relationship between the heat and the pouring time, the ideal pouring time of each heat and the processing time on the continuous casting machine, draw a schematic diagram of the ideal pouring time of each heat, as shown in Figure 1 and Figure 2.

Table 1 Example table for the relationship between cast and pouring times in the steelmaking-refining-continuous casting production process

L_{li}	1	2	3	4	5
l	1	2	1	2	1
i	1	2	3	4	5

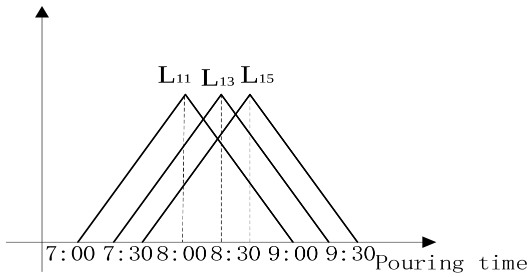

Fig 1 Schematic diagram of ideal opening time of each charge in pouring time 1

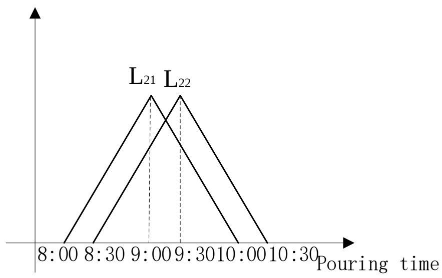

Fig 2 Schematic diagram of ideal opening time of each charge in pouring time 2

Through the equation (15), the performance indicators "the sum of the waiting time for each process in the refining process" and "the difference between the ideal pouring time and the actual pouring time of each heat in the refining process" are transformed into the optimization goal. Next, another performance index "processing adjacent heats on the same refining equipment without operating conflicts at the same time" is transformed into a constraint by equation (17).

$$Q\big(x(k+1),u(k+1)\big) > Q\big(x(k),u(k)\big) + g\big(x(k),u(k),x(k+1)\big) k \in \{1,\dots,E\}; i \in \{1,\dots,D\} \quad (17)$$

Where: $g\big(x(k),u(k),x(k+1)\big)$ is the refining production action $u(k)$ taken from the refined state $x(k)$ to the refined state $x(k+1)$ Production time.

3. Algorithm design

3.1 Traditional Q learning algorithm

The traditional Q learning algorithm is an off-policy reinforcement learning method. It is a reinforcement method proposed by Watkins to solve Markov decision problems with incomplete information. Calculate through the basic equation of the Q learning algorithm, and the obtained calculation result is greater than 0, then the Q value table is updated. When the Q value table is updated, the optimal value of the next state will be calculated, and the optimal value of the next state calculated according to this method will make corresponding actions. Earlier studies (Chen H., et al., 1998) showed it can be found that the action does not depend on the current solution strategy. At present, as the research on traditional Q learning algorithms has become more mature, traditional Q learning algorithms have been widely used in real life. For example: learning optimal operating procedures in factories, learning chess skills, controlling mobile robots, etc. The traditional Q learning algorithm is the expectation that $Q(s,a)$ can obtain the optimal benefit by taking action $a(a \in A)$ in the s state ($s \in S$) at a certain time. Earlier studies (Luh P. B., et al., 1998) showed the exploration environment gives the corresponding reward feedback according to the action taken by the agent, and uses the basic equation of the Q learning algorithm to obtain the corresponding reward (r) through calculation. The core idea of the Q learning algorithm is first to make the agent explore the environment at every moment The state pairs composed of state (s) and action (a) are concentrated to form a state-action pair set, and a Q value table is constructed, Used to store the Q value obtained by iterative calculation through the basic equation of the Q learning method; Secondly, a new Q value table is reconstructed according to the obtained updated Q value; finally, the re-constructed Q value table is used to guide the agent's subsequent exploration route and obtain the optimal agent's exploration strategy. The traditional Q learning algorithm is mainly used in the production system in the Markov environment. It uses the action sequence experienced by the agent and selects the optimal action of the next state according to the Q value calculated by the basic equation of Q learning. The traditional Q A key assumption of the learning algorithm is to regard the interaction between the agent and the environment as a Markov decision process. The current state is only related to the next state and has nothing to do with other states, which simplifies the state transition model and improves computational efficiency. There are many applications in the scheduling of production lines and planning the optimal walking path of the agent. There are two reasons for the use:

(1) Firstly, the traditional Q learning algorithm "does not have a model", which can directly estimate and predict the Q value of any action in each production state; Secondly, refer to the Q value obtained by iterative calculation using the basic equation of the Q learning algorithm. Finally, Earlier studies (Watkins C. J. C. H., Dayan P., 1992) showed the online decision method is used to decide the action that can obtain the highest Q value in the current state of the agent, and then update the Q value table. If the state-action pair can be accessed for an unlimited number of times during the running of the algorithm, then we can get the optimal function value of convergence. The traditional Q learning algorithm can directly predict and estimate the Q value of each action in the next state of the agent through its own basic algorithm equation, which improves the efficiency of the solution.

(2) The traditional Q learning algorithm can make scheduling decisions and estimate the production time for any process in any production stage of the production process. According to the actual production situation, the corresponding scheduling strategy is adopted to select the next optimal production action. The traditional Q learning algorithm uses the random scheduling method to select the next optimal production action to ensure that the optimal action in the next production state can be selected to achieve "overall optimal" instead of "local optimal", ensuring the scheduling plan The optimization effect of.

The traditional Q learning algorithm first needs to initialize the learning matrix Q; secondly, according to the scheduling environment, calculate the reward value of all the actions that the agent may take, and establish the reward matrix R; again, select an action according to the scheduling environment of the agent (a) As the initial action, and use the basic equation of the Q learning algorithm:

$$\mathrm{Q}(s,a) = (1-\alpha)Q(s,a) + \alpha\{R(s,a) + Gamma * \mathrm{Max}[Q(next\ s, next\ a) - Q(s,a)]\} \quad (18)$$

In the equation: α is the learning coefficient.

Calculate the Q value return corresponding to the optimal execution action of the next state, and update the learning matrix Q; finally, judge whether the learning matrix Q converges, and the algorithm ends when convergence, otherwise it returns to use the Q learning basic equation to continue the iterative calculation, Until the learning matrix Q converges.

3.2 Improved Q learning algorithm

Based on the traditional Q learning algorithm, this paper proposes a solution strategy based on the improved Q learning algorithm to solve the steel refining production scheduling optimization problem with uncertain molten steel hit rate. Through earlier studies (Jaakkola T., et al., 1994) showed the research on the basic principles of the traditional Q learning algorithm, the refining production scheduling optimization Comprehensive consideration of the various performance indicators of the problem can be obtained. In the initial stage of the traditional Q learning algorithm, the agent cannot accurately select the optimal action of the next state. Earlier studies (Zitzler E., 1999) showed pareto that can transform multiple performance indicators into optimization goals and constraints is introduced. (Pareto) solution set, and use the Pareto optimization solution idea to introduce the action selection probability P, use the basic equation of the action selection probability to calculate the action selection probability of all possible completion methods in the next system state, and select the action selection probability The largest completion method is used as the production method of the next system state.

Improving on the traditional Q learning algorithm and utilizing it for iterative calculation will save processing time in the refining production stage, improve the efficiency of refining production scheduling, and better cope with unexpected situations in the actual refining production scheduling process where the steel composition cannot meet production requirements and needs to be reheated. In response to the scheduling mathematical model built in this article, A method for solving the uncertain scheduling problem of molten steel hit rate in refining process using improved Q learning method has been proposed.

The procedure of the improved Q learning algorithm are as follows:

Step 1: Define the state space and action set of the improved Q learning algorithm;

$$R_{n\times n}=\begin{bmatrix} R_{[(1,1),1]} & \cdots & R_{[(1,n),n]} \\ \vdots & \ddots & \vdots \\ -1 & \cdots & R_{[(n,n),n]} \end{bmatrix}. \tag{19}$$

Step 2: Initialize the Q matrix, that is, set the Q matrix to a zero matrix, and set the learning coefficient α;

Step3: Select a state action pair $\left(x(k),u(k)\right)$ from the learning matrix Q as the initial state of the agent;

Step 4: Use the basic equation of the Q learning algorithm to calculate the Q values of all possible state action pairs that the agent transfers from the current state to the next state;

Step 5: When the agent chooses the optimal action in the next state during the running of the algorithm, use the action transition probability calculation equation: $P\left(\frac{a_i}{S_t}\right)=Q(S_t,a_i)/\sum_j Q\left(S_t,a_j\right)$, respectively calculate the selection probability of each state action pair that the agent may choose, compare the calculated action selection probabilities, and select the state action pair with the highest probability of performing action selection $\left(x(k+1),u(k+1)\right)$ continue to calculate;

Step 6: Use the basic equation of the Q learning algorithm to calculate the Q value of the selected state action pair $(x(k+1),u(k+1))$;

Step 7: If $Q_k^{(r+1)}>T^*$, where T^* is the R value of the state reward matrix, return to step 4 for calculation. If $Q_k^{(r+1)}<T^*$, update the learning matrix Q;

$$Q=\begin{bmatrix} Q_{[(1,1),1]} & \cdots & Q_{[(1,m),m]} \\ \vdots & \ddots & \vdots \\ Q_{[(n,1),1]} & \cdots & Q_{[(n,m),m]} \end{bmatrix} \tag{20}$$

Step 8: Judge whether the Q matrix converges, and the algorithm ends if convergence, otherwise, return to step 4 to continue the iterative calculation;

Step 9: Online decision-making, refer to the convergence learning matrix Q obtained by the improved Q learning algorithm for online decision-making, using the equation $u^*(k)=\arg\max_{u(k)\in U^{x(k)}} Q(x(k),u(k)$ compile production and processing strategies that meet the production requirements of subsequent processes.

4. Experimental verification

The improved Q learning algorithm solution proposed in this paper is used to solve and verify the refining production scheduling problem under the uncertain environment of molten steel hit rate. A convergent refining process learning matrix Q is obtained the time required for all processes in each heat to be processed in each refining production equipment is shown in Table 2.

Analysis of Table 2 shows that, considering that one of the two heuristic simulation strategies in the solution strategy is "let the heat with the ideal pouring time earlier enter the refining equipment for production", the start processing time of the three heats is set and the ideal pouring time, through the setting of the start time of the refining process of the three heats and the ideal pouring time on the continuous casting machine for each heat. Analyzing Table 3, we can see that the implementation of the refining process of each heat is permuted and combined, and with reference to the transition probability between adjacent refining processes, the implementation of the refining process with the transition probability of different heats between adjacent processes is 0 Different simulation strategies can get the refinement production process removal with the same production end time. Through calculation, it can be concluded that the number of refining process paths of the three heats has 16, 16, and 16 refining process paths respectively (a heat starts from the first refining process to the processing tasks of all refining processes Completion, including refining and refining, the implementation methods of the refining processes passed through are arranged and combined, which is called a refining process path), so the number of possible realization trajectories for the steel production enterprise to complete 3 heat production tasks is 16×16 ×16=4096 pieces. Table 3 is the ideal pouring schedule of each heat, and Table 4 is the realization method of each heat production task and the probability transition matrix table.

Table 2 Schedule of each charge processing in refining production equipment

Processing time / Process / Furnace sequence	KIP equipment Refine	LF equipment Refine
Heat 1	20	30
Heat 2	35	25
Heat 3	30	35
	RH equipment refining	
Heat 1	20	
Heat 2	20	
Heat 3	20	

Table 3 Ideal start time of casting for each charge

Furnace number	sequence	Refining production start time	Ideal pouring time
Heat 1		10：00	11：47
Heat 2		10：00	12：00
Heat 3		10：26	12：36

Table 4 Implementation mode and probability transfer matrix of production tasks for each charge

Heat 1	RH refining	KIP refining	Remelting 1	LF refining
	Coefficient matrix 1	Coefficient matrix 2	Coefficient matrix 3	Coefficient matrix 4
	$\begin{bmatrix}(1,1) & 800 & E \\ (1,1) & 760 & G\end{bmatrix}$	$\begin{bmatrix}(1,2) & 1450 & E \\ (1,2) & 2410 & G_1 \\ (1,2) & 2590 & F\end{bmatrix}$	$\begin{bmatrix}(1,3) & 3000 & E \\ (1,3) & 3100 & G_1 \\ (1,3) & 3200 & G_2\end{bmatrix}$	$\begin{bmatrix}(1,4) & 4070 & G_1 \\ (1,4) & 4010 & G_2\end{bmatrix}$
	Transition probability Matrix 1	Transition probability Matrix 2	Transition probability Matrix 3	Transition probability Matrix 4
	$\begin{bmatrix}0.32 \\ 0.68\end{bmatrix}$	$\begin{bmatrix}0.39 & 0.40 \\ 0.39 & 0.11 \\ 0.22 & 0.49\end{bmatrix}$	$\begin{bmatrix}0.41 & 0.2 & 0.17 \\ 0.3 & 0.4 & 0.40 \\ 0.29 & 0.4 & 0.43\end{bmatrix}$	$\begin{bmatrix}0.2 & 0.19 & 0.3 \\ 0.8 & 0.81 & 0.7\end{bmatrix}$
Heat 2	RH refining	KIP refining	Remelting 2	LF refining
	Coefficient matrix 1	Coefficient matrix 2	Coefficient matrix 3	Coefficient matrix 4
	$\begin{bmatrix}(2,1) & 1600 & E \\ (2,1) & 1700 & G\end{bmatrix}$	$\begin{bmatrix}(2,2) & 2030 & E \\ (2,2) & 2041 & G \\ (2,2) & 2081 & F\end{bmatrix}$	$\begin{bmatrix}(2,3) & 3073 & E \\ (2,3) & 3094 & G_1 \\ (2,3) & 3104 & G_2\end{bmatrix}$	$\begin{bmatrix}(2,4) & 4000 & G_1 \\ (2,4) & 3940 & G_2\end{bmatrix}$
	Transition probability Matrix 1	Transition probability Matrix 2	Transition probability Matrix 3	Transition probability Matrix 4
	$\begin{bmatrix}0.2 \\ 0.8\end{bmatrix}$	$\begin{bmatrix}0.62 & 0.31 \\ 0.34 & 0.45 \\ 0.04 & 0.24\end{bmatrix}$	$\begin{bmatrix}0.52 & 0.21 & 0.07 \\ 0.44 & 0.55 & 0.50 \\ 0.04 & 0.24 & 0.43\end{bmatrix}$	$\begin{bmatrix}0.34 & 0.45 & 0.57 \\ 0.66 & 0.55 & 0.43\end{bmatrix}$
Heat 3	RH refining	KIP refining	Remelting 3	LF refining
	Coefficient matrix 1	Coefficient matrix 2	Coefficient matrix 3	Coefficient matrix 4
	$\begin{bmatrix}(3,1) & 5228 & E \\ (3,1) & 5310 & G_1\end{bmatrix}$	$\begin{bmatrix}(3,2) & 6033 & E \\ (3,2) & 6091 & G \\ (3,3) & 7004 & F\end{bmatrix}$	$\begin{bmatrix}(3,3) & 7090 & E \\ (3,3) & 7145 & G_1 \\ (3,3) & 7245 & G_2\end{bmatrix}$	$\begin{bmatrix}(3,4) & 8004 & G_1 \\ 3,4) & 8540 & G_2\end{bmatrix}$
	Transition probability Matrix 1	Transition probability Matrix 2	Transition probability Matrix 3	Transition probability Matrix 4
	$\begin{bmatrix}0.43 \\ 0.57\end{bmatrix}$	$\begin{bmatrix}0.39 & 0.40 & 0.17 \\ 0.31 & 0.31 & 0.34 \\ 0.3 & 0.29 & 0.49\end{bmatrix}$	$\begin{bmatrix}0.62 & 0.31 & 0.07 \\ 0.34 & 0.45 & 0.60 \\ 0.04 & 0.24 & 0.33\end{bmatrix}$	$\begin{bmatrix}0.34 & 0.45 & 0.67 \\ 0.66 & 0.55 & 0.33\end{bmatrix}$

Using heuristic strategy simulation to study two heuristic simulation strategies suitable for this subject, namely the refining process, the heat in the refining process has a small sum of waiting time in each process, the production process first enters the refining production equipment for production and pouring The earlier furnace processes are first entered into the refining production equipment for production, and each simulation is performed 1000 times, and 2000 refining process paths can be obtained. Figure 4 and Figure 5 show the results of different heuristic simulation strategies.

In Figure 4, simulation strategy 1, that is, the refining process enters the refining equipment for production with a small sum of waiting time for each process in the refining process, and the simulation results are concentrated around the average production time of the refining process path. In the simulation process, the average production time of the refined production process path is 7.1 hours ahead of the simulation strategy 2 in Fig. 5.

According to the simulation results of the two heuristic strategies, the Gantt chart with the shortest production time of the refining process path corresponding to the simulation results is drawn, as shown in Figure 6 and Figure 7. By comparing the processing completion time of the refining production stage and the deviation between the ideal pouring time and the actual pouring time by comparing Figure 6 and Figure 7, it is known that the heuristic simulation strategy 1 completes the refining production task before the heuristic simulation strategy 2 and makes full use of the refining production equipment. So simulation strategy 1 is better than simulation strategy 2.

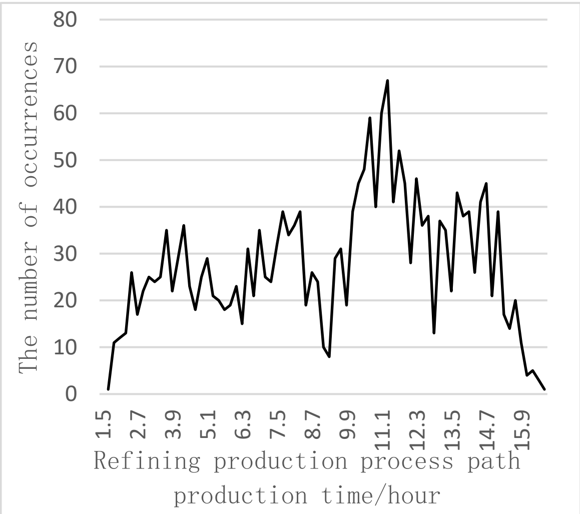

Fig. 4 Simulation results of heuristic simulation strategy 1

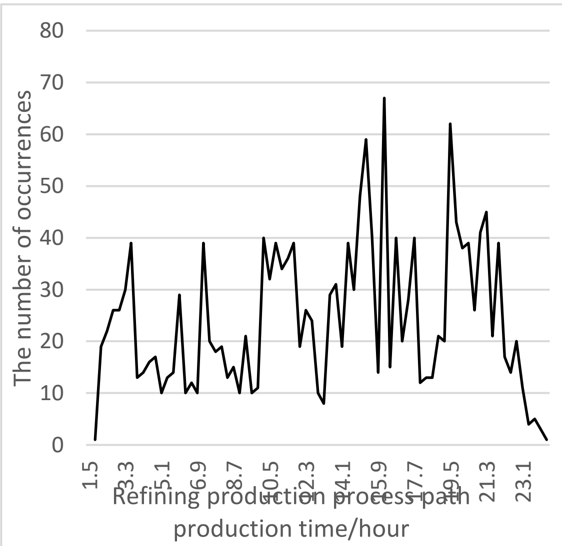

Fig. 5 Simulation results of heuristic simulation strategy 2

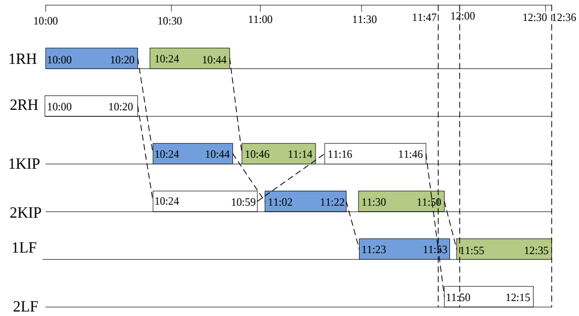

Fig 6 Gantt chart of refining process path based on simulation strategy 1

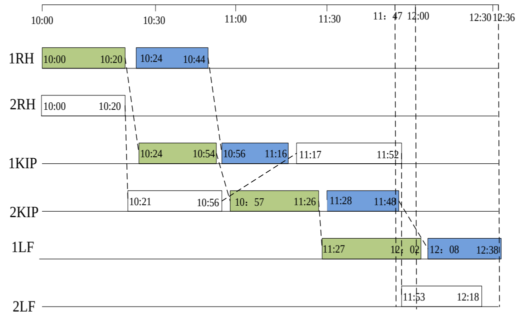

Fig 7 Gantt chart of refining process path based on simulation strategy 2

The simulation result of each heuristic strategy is a set of state action pairs for each heat in the refining stage. Combining the sets of state-action pairs of refining process obtained by two heuristic simulation strategies, 1776 non-redundant state-action pairs and the next state subset of each refining process state-action pair are obtained. By improving the Q learning algorithm iterative calculation 19 times, the convergent learning matrix Q of refined production process is obtained. Finally, the convergent refining process learning matrix Q is used to make online decisions on the actual refining process.

$$u^*(k) = \arg \max_{u(k)\in U^{x(k)}} Q(x(k), u(k)) \tag{21}$$

Dispatch the uncertainty scheduling optimization problem of the molten steel hit rate in the refining process of the same process path, and obtain the scheduling results of each heat suitable for the refining process of the same process path, or when the molten steel composition is different in the actual refining process. When meeting the production requirements and requiring refining, the entire refining process needs to be rescheduled. Using the improved Q learning algorithm iteratively obtained the convergent refinement production process learning matrix Q for decision-making, the average decision-making time is only 0.13s.

5. Summary

This essay discusses the significance and importance of the molten steel impact rate uncertainty problem, which frequently arises throughout the actual refining process. The molten steel hit rate of the refining process is enhanced by the optimization of uncertainty parameter technique, modeling method, and optimization solution method. The uncertain scheduling optimization problem's state of study is introduced. Analyze the scheduling optimization problem of the molten steel hit rate with uncertainty in the refining process in detail, specify the decision variables and model parameters from the viewpoint of the refining process, state the problem hypothesis, and use the Markov chain to analyze the molten steel hit rate uncertainty. The difference between the ideal pouring time and the actual pouring time of each heat in the refining process, as well as the total waiting time of the heat in each process, are the smallest for three steel refining production performance indicators, including adjacent processes in the same refining equipment. Heat conflicts won't be taken into account if the minimal value is met and nearby heats are processed simultaneously on the same refining equipment, according to a mathematical model. The Pareto solution set optimization solution concept and the introduction of action selection probability improve the traditional Q learning algorithm. The molten steel hit rate uncertainty in the refining process is used to build the matrix, and the Q learning is used to solve the molten steel hit rate uncertainty in the scheduling problem iteratively. The approach of enhancing the Q learning algorithm to resolve the molten steel hit rate uncertainty scheduling problem in the steelmaking production process is examined in this research. On the basis of data verification experiments, the effectiveness of the proposed algorithm and its high utilization value in actual steel production are further verified.

Acknowledgements

We would like to acknowledge financial support provided by National Natural Science Foundation of China (61873174), Liaoning Provincial Natural Science Foundation (2020-KF-11-07), Ministry of Education Cooperative Education Project (201902233028), and China Postdoctoral Science Foundation Funded Project (2017M611261).

References

Chen, H., Chu, C., & Proth, J. M. (1998). An improvement of the Lagrangean relaxation approach for job shop scheduling: a dynamic programming method. IEEE transactions on robotics and automation, Vol. 14 No. 5, pp.786-795.

Hoitomt, D. J., Luh, P. B., & Pattipati, K. R. (1993). A practical approach to job-shop scheduling problems. IEEE transactions on Robotics and Automation, Vol.9 No. 1, pp.1-13.

Jaakkola, T., Jordan, M. I., & Singh, S. P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. Neural computation, Vol.6 No.6, pp.1185-1201.

Jinyan, M., Chai, S. Y., & Youyi, W. (1995, May). FMS jobshop scheduling using Lagrangian relaxation method. In Proceedings of 1995 IEEE International Conference on Robotics and Automation Vol. 1, pp. 490-495.

Luh, P. B., Gou, L., Zhang, Y., Nagahora, T., Tsuji, M., Yoneda, K., ... & Kano, T. (1998). Job shop scheduling with group-dependent setups, finite buffers, and longtime horizon. Annals of Operations Research, Vol.76, pp. 233-259.

Nishi, T., Hiranaka, Y., & Inuiguchi, M. (2007, September). A successive Lagrangian relaxation method for solving flowshop scheduling problems with total weighted tardiness. In 2007 IEEE International Conference on Automation Science and Engineering, pp. 875-880.

Nishi, T., Isoya, Y., & Inuiguchi, M. (2009, October). An integrated column generation and Lagrangian relaxation for flowshop scheduling problems. In 2009 IEEE International Conference on Systems, Man and Cybernetics, pp. 299-304.

Sun L. L., Lu T. Y., Sha S. Y., et al., (2020), 'Research on Scheduling Method for Uncertainty of Hit Rate of Molten Steel Based on Q Learning'. The Second International Symposium on Simulation and Process Modelling. Shenyang Jianshu university, Shenyang, China.

Tan, Y. Y., Huang, Y. L., & Liu, S. X. (2013). Two-stage mathematical programming approach for steelmaking process scheduling under variable electricity price. Journal of Iron and Steel Research, International, Vol.20 No.7, pp,1-8.

Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine learning, Vol.8 No.3-4, pp.279-292.

Zitzler, E. (1999). Evolutionary algorithms for multiobjective optimization: Methods and applications. Ithaca: Shaker, Vol. 63.

Jiayu Peng is currently pursuing her PhD study at the School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, China. She obtained the BS degree and the MS degree in Automation and Control Engineering from Shenyang Jianzhu University in 2018 and 2021. Her main research interests are in decision making and optimization of complex industrial production scheduling.

Tingwei Pan is currently pursuing his baccalaureate study at the School of Civil Engineering, Shenyang Jianzhu University, Shenyang, China. His main research interest is structural mechanics.

Jiaxin Guo is currently studying for a bachelor's degree in transportation at Dalian Jiaotong University, Dalian, China. His main research interest is transportation.

Xinqi Hu is currently pursuing his baccalaureate study at the School of Electrical and Contral Engineering, Shenyang Jianzhu University, Shenyang, China. His main research interest is Electrical Engineering and Automation.

Congxin Li is currently pursuing her PhD study at the School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, China. He obtained his BS degree from North University of China, China in 2017 and his MS degree from Shenyang Jianzhu University, China in 2020. His main research interests are in the areas of include industrial scheduling decision making and batch planning.

Jinging Lou received the M. S. degree in Control Theory and Control Engineering from Shenyang Jianzhu University, Shenyang, China, in 2006. She is currently pursuing the Ph. D. degree in mechanical engineering at Shenyang Jianzhu University, Shenyang, China.She is currently a Professor with the School of Mechanical Information, Yiwu Industrial & Commercial College. Her research interests include computer graphic, computer aided design, geometric design, control engineering and computational geometry.

Liangliang Sun obtained the B.S. degree in Information and communication engineering from Beijing Jiaotong University, Beijing, China, in 2003, and the M.S. and Ph.D. degree in control theory and control engineering form Northeastern University in 2007 and 2015, respectively. He is a professor with the School of Control Engineering, Northeastern University at Qinghuangdao. His research interests are intelligent manufacturing systems – planning, scheduling, and coordination of design, manufacturing, and service activities.