Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-10 (2022) 359–364

Reinforcement learning and digital twin-based real-time scheduling method in intelligent manufacturing systems

Lixiang Zhang, Yan Yan, Yaoguang Hu, Weibo Ren

**Beijing Institute of Technology, Beijing, China 100081*
(Tel: +86-010-6891-7880; e-mail: hyg@bit.edu.cn).

Abstract: Optimization efficiency and decision-making responsiveness are two conflicting objectives to be considered in intelligent manufacturing. Therefore, we proposed a reinforcement learning and digital twin-based real-time scheduling method, called twins learning, to satisfy multiple objectives simultaneously. First, the interaction of multiple resources is constructed in a virtual twin, including physics, behaviors, and rules to support the decision-making. Then, the real-time scheduling problems are modeled as Markov Decision Process and reinforcement learning algorithms are developed to learn better scheduling policies. The case study indicates the proposed method has excellent adaptability and learning capacity in intelligent manufacturing.

Keywords: Real-time scheduling; reinforcement learning; digital twin; intelligent manufacturing;

1. INTRODUCTION

Customized production becomes an important mode to satisfy individual needs, which brings new challenges to production scheduling, such as dynamic and uncertain orders, limited delivery dates, and complex machining processes. Many researchers focused on the re-scheduling method to respond to dynamic changes. But the responsiveness can't be fully satisfied because of the complexity of heuristic algorithms. In this context, a variety of heuristic rules have been proposed to meet the real-time requirements of production scheduling (Gan & Tao, 2013; Jeong & Randhawa, 2001), but it has poor generalization ability and optimization performance.

In recent years, reinforcement learning has been attempted to solve task allocation, job sequencing, and path planning(Shiue et al., 2018). Xue *et al.*(2018) established the Markov Decision Process(MDP) of the real-time scheduling of multiple AGVs in the flow shop and used Q learning algorithm to minimize the average delay and completion time in a dynamic environment. Hu *et al.*(2020) developed a deep Q network to solve the joint optimization MDP of job assignment and AGV scheduling. Tang *et al.*(2021) proposed a hierarchical reinforcement learning based on Actor-Critic to solve the real-time task allocation problem of multi-robots in unmanned warehouses. Malus *et al.*(2020) used a multi-agent reinforcement learning algorithm to solve the real-time scheduling problem of multiple AGVs and verified the effectiveness by preliminary and accurate simulation, which showed robustness and generalization ability. However, the above research mainly focused on the scheduling optimization of a single resource and ignored the interaction among task order flexibility, equipment processing flexibility, and AGV path flexibility. Therefore, how integrating the flexibility of multiple resources to achieve efficient collaboration and flexible manufacturing is the key to improving the performance, which is also a difficult issue in the optimization of manufacturing system management and control.

Digital twin, as one of the key enabling technologies of intelligent manufacturing, realizes accurate representation of geometry, physics, behaviors, and rules to construct a real-time virtual mapping of the physical world, which provides a new research perspective for production optimization. More and more researchers have focused on the production scheduling problems in the digital twin environment. Fang *et al.*(2019) proposed a digital twin-based rescheduling method and designed a scheduling resource parameter update method and dynamic interactive scheduling strategy, where the prototype system was established by Demo3D to realize real-time interaction of dynamic events. Negri *et al.*(2019) embedded a modular that integrated equipment health predictions into a digital twin and utilized a genetic algorithm to provide scheduling alternatives. Zhang *et al.*(2020) utilized the real-time information from the digital twin and presented a process planning approach based on reinforcement learning to derive near-optimal process plans. The above-mentioned research utilized the real-time perception and simulation capability of the digital twin and achieved significant performance in job shops, assembly shops, and intelligent manufacturing units. However, the self-perception, self-learning, self-decision-making, and self-interaction capabilities of the multi-level decision-making process at the workshop level are insufficient.

As a result, this paper proposes a reinforcement learning and digital twin-based real-time scheduling method, called twins learning(TL), which integrates the physical space, the digital space, and the learning optimization space to satisfy the responsiveness and optimization efficiency simultaneously. The main contributions of this paper are summarized as follows:

- A novel real-time scheduling method is proposed to explore the application of reinforcement learning algorithms in intelligent manufacturing systems.

10.1016/j.ifacol.2022.09.413

- The digital twin modeling method is presented to represent special features for mapping the physical system and the virtual system for scheduling optimization.
- The solutions for real-time scheduling are proposed to analyze scheduling problems, model Markov Decision Process, and design optimization algorithms.
- A case study for AGV scheduling with avoidance is presented to verify the feasibility and effectiveness of the proposed twins learning method.

2. TL FRAMEWORK

2.1 TL definition

The concept of digital twin(DT) was first proposed by Professor Grieves (2014). In the past few years, a variety of the concept about DT was proposed and explained(Reifsnider & Majumdar, 2013). DT is a virtual mapping of the physical system, which fully reflects the characteristics of the structure and behavior of the physical world. It's a new way to realize monitoring, simulation, and optimization. Reinforcement learning(RL) is an important branch of machine learning, which tries to explore the optimal policy for decision-making problems through continuous interaction with the environment. In recent years, deep reinforcement learning provides new insights to solve decision-making optimization problems.

TL refers to the interaction between the digital and the learning optimization space to realize real-time decision-making and optimization simultaneously. First, the digital space is established to represent the real-time status of the physical space. Then, the special features are treated as the input of the agent to make decisions. And the physical space performs the decisions and returns the feedback into the learning optimization space. Finally, a set of historical experiences, state-action-reward-state, are utilized to learn a better policy until obtaining a stable and adaptive agent, which assists the real-time production scheduling. Based on this basis, major functions, such as self-perception, real-time analysis, self-decision-making, and self-execution, are integrated to support the management and control of intelligent manufacturing systems and help enterprises to improve production quality and efficiency.

2.2 TL model

TL model includes five dimensions, as shown in Figure 1, the physical space, the digital space, and learning optimization space, data center, and network, which is defined as TL={PS, DS, LOS, ED, NT}. PS is composed of intelligent equipment such as machines, AGVs, robots, warehouses, and various sensors installed on the equipment, which can perform flexible machining or assembly processes according to individual parameters. DS is the virtual mapping of PS, which communicates with other dimensions by NT and interacts with information systems, such ERP, WMS, and MES. LOS is the major function to improve the production performance and ensure stable operation. The self-learning agent in LOS provides decision-making according to the real-time status and evolves a better policy by interacting with environments. The DC functions include production data storage, real-time

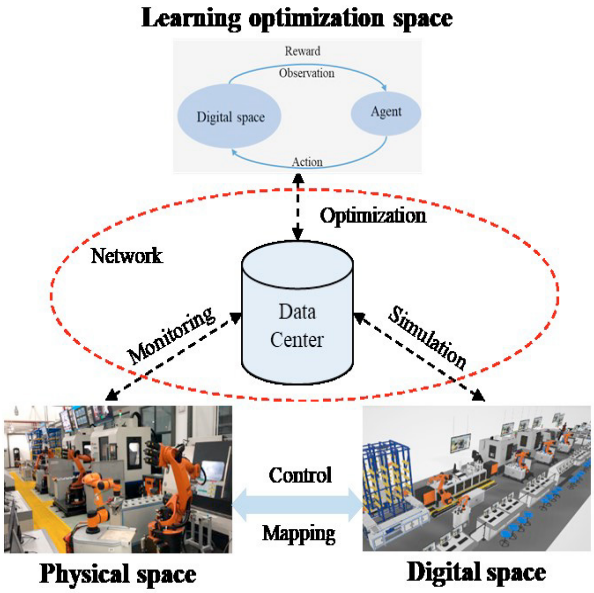

Figure 1. The framework of TL model

data preprocessing, and learning data storage. NT guarantees the interaction among five dimensions with low latency and high security.

The TL model can be constructed as shown in Figure 2.

1) A digital twin of PS is constructed to accurately represent the geometry and physics of equipment.
2) The behaviors and rules between equipment in manufacturing systems are represented to realize the self-perception and self-execution through the real-time interaction between the physical space and digital space.
3) The learning optimization problems, such as task allocation, job sequencing, and AGV routing, are formulated to MDP.
4) The reinforcement learning agent is designed and a suitable algorithm is selected and trained.
5) The agent is integrated with the digital twin to optimize the algorithm by interacting with the real environment.
6) The learning optimization model and agent are deployed to provide decision-making according to the real-time status in the manufacturing system to realize self-perception, self-learning, self-decision-making, self-execution, and self-adaptation in smart manufacturing.

2.3 TL functions

Compared with digital twins, the main functions of TL could include simulation, optimization, and real-time decision-making. TL takes advantage of digital twins to build the virtual mapping of physical entities as well as production processes, which can be treated as a surrogate model to simulate the performance of solutions and optimize production performance. Compared with interacting with the physical system, TL is helpful to reduce the cost of learning activities.

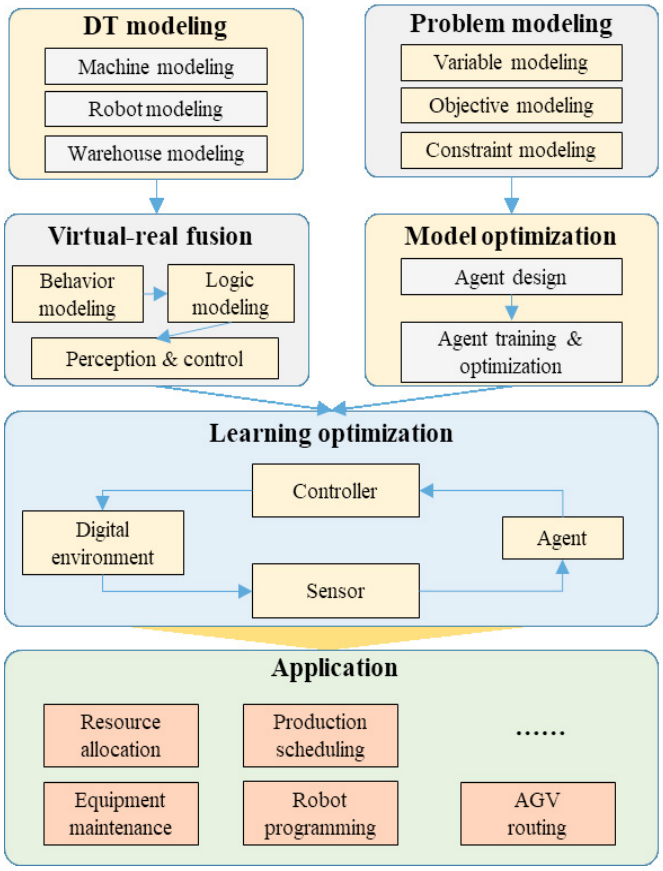

Figure 2. The construction processes of TL model

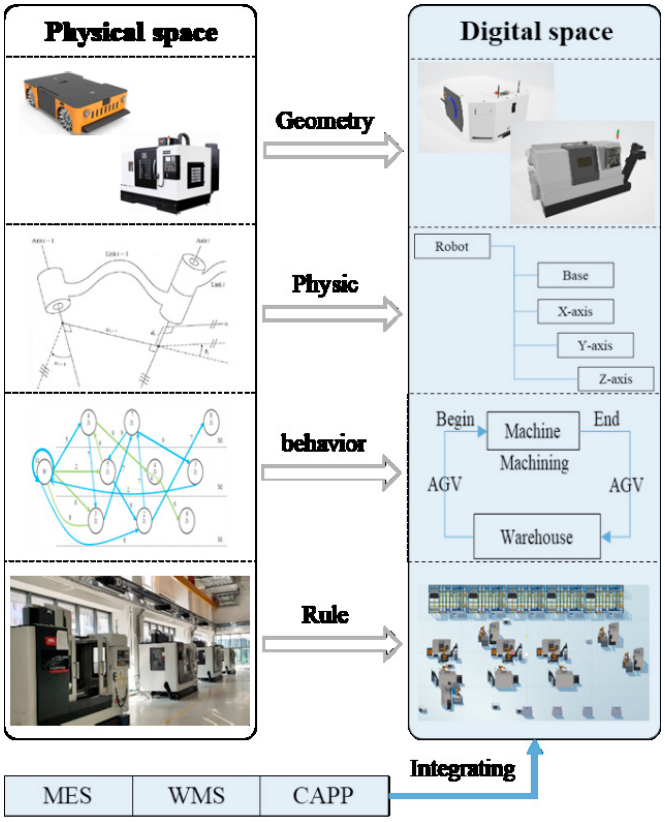

Figure 3. The representation of DT model

In addition, agents are designed to learn implicit knowledge from decision-making experiences and make decisions according to real-time information. After fully interacting with the digital twin, the agent can efficiently provide real-time decision-making for the physical system to control and manage the system operation.

3. DT MODEL

3.1 DT features

Compared with the traditional digital twin for real-time monitoring and analyzing, the digital twin for twins learning have more complex and dynamic features as follows.

a) The self-perception and real-time interaction of multiple manufacturing resources is the primary feature, including the multi-source data collection, the interaction between equipment, and the interaction between the learning optimization model and the manufacturing system, which aims at utilizing the real-time data to update the decision-making model and provide a better policy.
b) DT can dynamically configure and adjust according to the real-time environment, such as the number of equipment, machining or assembly order, and facility layout, which ensures the accurate virtual-real mapping between PS and DS as well as improves the adaptability of the learning optimization model.
c) The light representation and high compatibility are necessary for TL, which can not only satisfy the frequent interaction between DS and LOS and also extend major functions by connecting with other information systems.

3.2 DT representation

To utilize TL to realize the optimization of facility layout, production scheduling, and resource management. The representation processes of the DT model as shown in Figure 3, are composed of the representation of geometry, physics, behaviors, and rules as well as the integration with information systems.

First, relevant software is used to build the geometry model of various equipment according to the real measurement. Next, the physical features are constructed to represent the movement relationship of equipment, such as the parent-child relationship between joints. After the virtual mapping of equipment, the interactive behaviors among equipment are constructed, such as the interaction between materials and AGVs, the interaction between two AGVs for conflict-free routing, as well as the interaction between materials and machines, which integrates various equipment into a manufacturing system. Then, the operation rules are added to the operation and interaction processes to represent the production flexibility and limit illegal actions in the virtual environment. Based on the above DT model, advanced information systems are integrated to perform the learning optimization model to improve productivity.

4. RTS SOLUTION

To improve production efficiency and responsiveness, real-time scheduling methods have attracted more attention in recent years. In this paper, the real-time scheduling based on TL is proposed to utilize the learning capacity of deep reinforcement learning and the mapping capacity of the digital

twin to support efficient learning and decision-making. According to the previous studies, the real-time scheduling problems in the manufacturing system follow the Markov property. The current status just is related to the last step status and the last action. As a result, the current status in the manufacturing system is just considered when a decision needs to be made, which forms a Markov Decision Process. It's described as a state-action-reward-state, which can be optimized by reinforcement learning algorithms to obtain a better policy.

The real-time scheduling solution based on TL can be described as shown in Figure 4. First, the problems are analyzed and categorized into resource allocation, machining scheduling, AGV scheduling, AGV routing, and so on, which is helpful to explore the solution and design the algorithm from historical experience. Then, the problem is modeled as an MDP, including the action space, the state space, and the reward function. On this basis, the output layer and the input layer can be defined. As for the reward function of real-time scheduling problems, the direct reward is hardly obtained at the current step, whereas the indirect reward is usually added to guide the learning processes. After obtaining the MDP, a suitable algorithm can be selected according to the state space and the action space.

Reinforcement learning algorithms can be categorized into policy-based, model-based, and value-based methods. Because of the complexity of real-time scheduling problems, an accurate model of RTS problems is hardly obtained. Thus, value-based and policy-based reinforcement learning algorithms are usually developed against production scheduling problems. Generally, the value-based methods are selected if the action space is discrete and the policy-based methods are selected if the action space is continuous. Last but not least, the designed deep reinforcement learning algorithm is integrated with the corresponding manufacturing information system and connected with the digital space to explore a better policy as well as provide real-time decision-making according to the real-time status of the manufacturing system. After that, the solution processes of actual problems could be summarized as historical experience, which helps to rapidly generate solutions if similar problems are encountered.

5. CASE STUDY

The production logistics system(PLS) ensures the stable and efficient operation of the manufacturing system and improves the flexibility of production processes to respond to customized needs, where a variety of AGVs are widely applied to perform tasks of material handling between equipment and warehouses(Zhang et al., 2019, 2021). However, limited by the requirement of responsiveness and flexibility, scheduling rules are usually used to realize task allocation in real production and the AGV routing is usually pre-defined within a fixed range. It causes that the autonomy and flexibility of AGVs have not been fully utilized yet. Consequently, we proposed a novel method based on twins learning to realize AGV scheduling with avoidance by an assignment agent and an avoidance agent respectively.

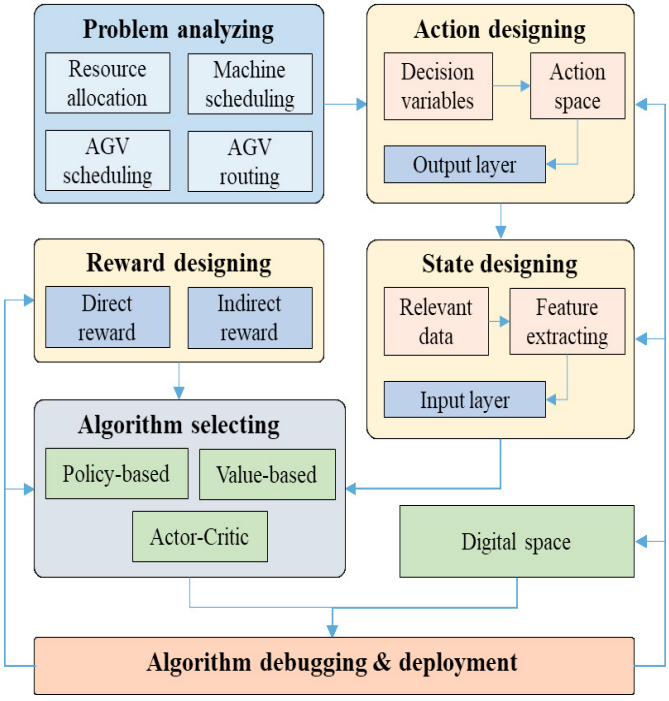

Figure 4. The solution of TL for RTS

5.1 Problem definition

5.1.1 Task assignment MDP

To improve the efficiency and feasibility of material distribution, a task assignment agent is designed to assign a suitable AGV for the material handling task according to the real-time task and AGV information. In this paper, the task assignment problem is formulated as an MDP, which is expressed as $(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\gamma}, \boldsymbol{R}, \boldsymbol{\pi})$. $\boldsymbol{S}$ represents the real-time status information of the production logistics system, including the starting point, target point, weight, and delivery date of the task, as well as the position, working status, and power information of all AGVs. It's represented by $6+4\times q$ dimensions, where q is the number of AGVs in the production logistics system and the position information includes X-axis and Y-axis information. $\boldsymbol{A}$ represents the action of the task assignment agent, including the assigned AGV and the charging strategy for maintaining the continuous operation. The charging strategy includes charging before the task starts and charging after the task is completed. $\boldsymbol{\gamma}$ is a discount factor to balance immediate and future rewards. $\boldsymbol{\pi}$ represents the policy to make decisions. $\boldsymbol{R}$ represents the reward after performing the action, which is combined with energy consumption and tardiness.

5.1.2 AGV routing MDP

Compared with traditional manufacturing systems, there are many dynamic and uncertain changes affecting the facility layout in the intelligent manufacturing system. Therefore, AGVs must autonomously avoid obstacles according to perceived surroundings to perform material handling tasks in time. In this paper, the AGV conflict-free routing is realized by a global path planning algorithm and a local avoidance agent. As for the global path routing algorithm, the A* algorithm is used to find a feasible path according to the current environmental map. The avoidance agent is designed

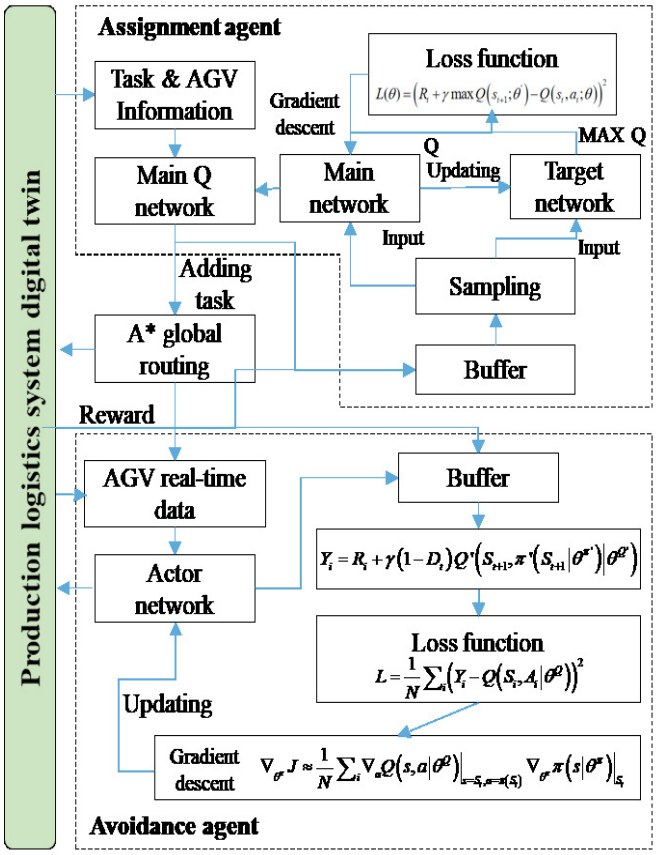

Figure 5. The framework of TL in PLS

to realize real-time avoidance according to the local environmental status. The real-time avoidance problem is formulated as a partially observable MDP, which is defined as (***S, A, R, Ω, O***). ***S*** is the state space of the environment. ***Ω*** is the observation space, which includes the real-time data of the Laser-Lidar, the AGV speed, the global path guidance direction, and the relative position between the AGV position and the local target position. ***O*** is the observation function of the environment, which consists of multiple consecutive observations as input. ***A*** is the action space, including the speed of the AGV on the X-axis and Y-axis respectively. ***R*** is the reward function, which consists of the fixed reward, the tangential running reward, the normal running reward, and the collision penalty.

5.2 TL optimization in PLS

Based on the MDP of the above task assignment problem and AGV avoidance problem, an assignment agent and an avoidance agent are designed to realize the optimization of task assignment and conflict-free routing. Here, a deep Q network algorithm is designed to achieve task assignment, where the main Q network and the target Q network are set with the same networks and parameters. According to the experiments, the neural networks can be designed with two hidden layers and the activation function *ReLU* is used. The output layer uses Sigmoid as the activation function. On the other hand, a deterministic policy gradient algorithm is developed to realize real-time avoidance, including an actor and a critic. As for the actor, the input is the multi-frame sequence observation and the output is the action. For the critic, the input is the observation as well as the action and the output is the estimated value to update the policy. Because there is multi-frame sequence observation data, a recurrent neural network with three hidden layers is designed to connect with the input, where the *Tanh* is set as the activation function. Then, a fully connected neural network with three hidden layers is designed to connect with the output layer, where the *ReLU* is set as the activation function.

Based on the above MDP and designed algorithms, the training processes and the interaction between the agent and the digital space are described as shown in Figure 5, including the production logistics system digital twin, the centralized assignment agent, and the avoidance agent. The digital twin is the real-time virtual mapping of the production logistics system, where the construction processes are shown in Section 3. The assignment agent provides the decision-making according to the real-time information. Then, the A* algorithm is used to plan a global path for the assigned AGV. The avoidance agent provides the real-time order to realize conflict-free routing when the AGV performs the corresponding material handling task. In the meanwhile, the digital twin provides real-time feedback for the agents when the environmental status changes after performing the action, which guides the training processes.

As a result, TL can be applied in a real production logistics system as shown in Figure 6. The main functions consist of task assignment, path planning, avoidance, monitoring, and visualization.

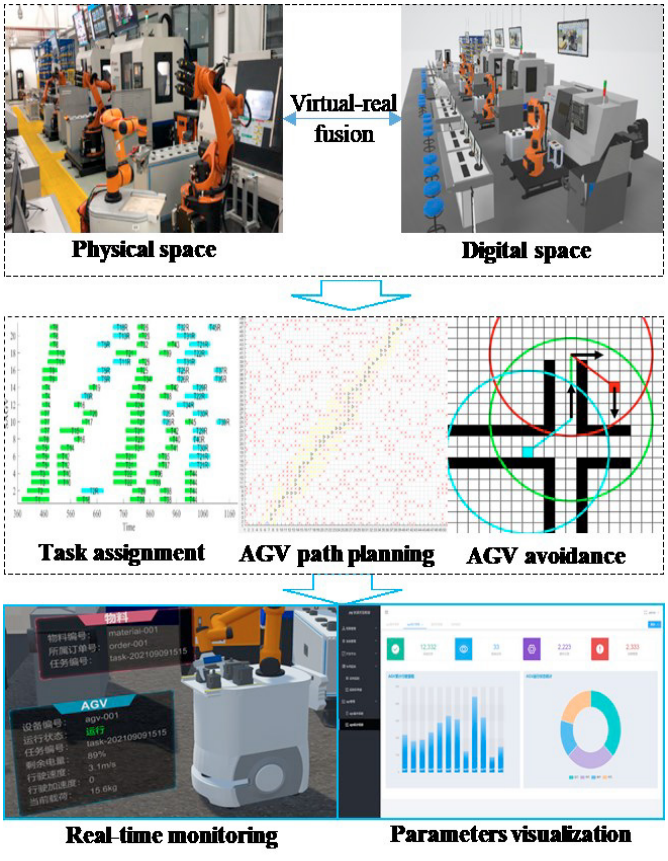

Figure 6. The application of TL in PLS

5.3 Discussion

Based on the above theoretical works and system development, the proposed method has significant management value in practical production as follows.

On the one hand, the state representation of the manufacturing system helps to realize real-time monitoring and also analyze essential features. For example, the utilization of AGVs in PLS can be analyzed to adjust the number of AGVs or the facility layout to improve production performance.

On the other hand, multiple agents can be designed to collaborate with others, which is helpful to eliminate the limitation of the multi-stage method and improve real utilization. For example, the task assignment agent in the case study could adapt to the environment with the avoidance agent, which could obtain better performance in real production environments than without avoidance.

6. CONCLUSION

To task advantage of the digital twin and machine learning to improve production performance, this paper aims at building a bridge between advanced research and industrial application. From the self-learning and self-decision-making perspective, this paper proposes a novel RTS method based on DT and RL, which utilizes the representation capacity of DT and the learning capacity of RL to explore a better policy. Compared with traditional optimization methods, the proposed TL method has better adaptability, because the policy is obtained by interacting with the digital space which is a real-time mapping of the physical space. Besides, the proposed TL method might provide a new perspective for the design of intelligent manufacturing systems. The reason why is the performance of the virtual manufacturing system can be evaluated when the real manufacturing system has not been constructed and also the design can be optimized according to the real-time monitoring of important parameters. In conclusion, the theoretical research and the case study all indicate the proposed TL method has significant potential to be applied in the intelligent manufacturing era.

In the future, based on the TL method, we will explore the real-time scheduling of multiple resources as well as the joint optimization of predictive maintenance and production scheduling.

Acknowledgment

This work was supported by the National Key R&D Program of China (Project No. 2021YFB1715700) and the National Natural Science Foundation of China (Project No. 52175451).

REFERENCE

Fang, Y., Peng, C., Lou, P., Zhou, Z., Hu, J., & Yan, J. (2019). Digital-Twin-Based Job Shop Scheduling Toward Smart Manufacturing. IEEE Transactions on Industrial Informatics, 15(12), 6425–6435.

Gan, Z., & Tao, L. (2013). Control of automated guided vehicles based on multi-attribute dispatching rule. Applied Mechanics and Materials, 278–280, 1432–1435.

Grieves. (2014). Digital twin: manufacturing excellence through virtual factory replication (Issue 12 (761)).

Hu, H., Jia, X., He, Q., Fu, S., & Liu, K. (2020). Deep reinforcement learning based AGVs real-time scheduling with mixed rule for flexible shop floor in industry 4.0. Computers and Industrial Engineering, 149(January), 106749.

Jeong, B. H., & Randhawa, S. U. (2001). A multi-attribute dispatching rule for automated guided vehicle systems. International Journal of Production Research, 39(13), 2817–2832.

Malus, A., Kozjek, D., & Vrabič, R. (2020). Real-time order dispatching for a fleet of autonomous mobile robots using multi-agent reinforcement learning. CIRP Annals, 69(1), 397–400.

Müller-Zhang, Z., Antonino, P. O., & Kuhn, T. (2020). Dynamic Process Planning using Digital Twins and Reinforcement Learning. IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, 2020-Septe(September), 1757–1764.

Negri, E., Ardakani, H. D., Cattaneo, L., Singh, J., MacChi, M., & Lee, J. (2019). A Digital Twin-based scheduling framework including Equipment Health Index and Genetic Algorithms. IFAC-PapersOnLine, 52(10), 43–48.

Reifsnider, K., & Majumdar, P. (2013). Multiphysics stimulated simulation digital twin methods for fleet management. 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 1–11.

Shiue, Y. R., Lee, K. C., & Su, C. T. (2018). Real-time scheduling for a smart factory using a reinforcement learning approach. Computers and Industrial Engineering, 125(101), 604–614.

Tang, H., Wang, A., Xue, F., Yang, J., & Cao, Y. (2021). A Novel Hierarchical Soft Actor-Critic Algorithm for Multi-Logistics Robots Task Allocation. IEEE Access, 9, 42568–42582.

Xue, T., Zeng, P., & Yu, H. (2018). A reinforcement learning method for multi-AGV scheduling in manufacturing. Proceedings of the IEEE International Conference on Industrial Technology, 2018-Febru, 1557–1561.

Zhang, L., Hu, Y., & Guan, Y. (2019). Research on hybrid-load AGV dispatching problem for mixed-model automobile assembly line. Procedia CIRP, 81, 1059–1064.

Zhang, L., Yan, Y., Hu, Y., & Ren, W. (2021). A dynamic scheduling method for self-organized AGVs in production logistics systems. Procedia CIRP, 104, 381–386.