Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering) 2022 23(6):458-478
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.A2100325

Towards autonomous and optimal excavation of shield machine: a deep reinforcement learning-based approach

Ya-kun ZHANG[1], Guo-fang GONG[1]✉, Hua-yong YANG[1], Yu-xi CHEN[1], Geng-lin CHEN[2]

[1]*State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China*
[2]*School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou 221116, China*

Abstract: Autonomous excavation operation is a major trend in the development of a new generation of intelligent tunnel boring machines (TBMs). However, existing technologies are limited to supervised machine learning and static optimization, which cannot outperform human operation and deal with ever changing geological conditions and the long-term performance measure. The aim of this study is to resolve the problem of dynamic optimization of the shield excavation performance, as well as to achieve autonomous optimal excavation. In this study, a novel autonomous optimal excavation approach that integrates deep reinforcement learning and optimal control is proposed for shield machines. Based on a first-principles analysis of the machine-ground interaction dynamics of the excavation process, a deep neural network model is developed using construction field data consisting of 1.1 million samples. The multi-system coupling mechanism is revealed by establishing an overall system model. Based on the overall system analysis, the autonomous optimal excavation problem is decomposed into a multi-objective dynamic optimization problem and an optimal control problem. Subsequently, a dimensionless multi-objective comprehensive excavation performance measure is proposed. A deep reinforcement learning method is used to solve for the optimal action sequence trajectory, and optimal closed-loop feedback controllers are designed to achieve accurate execution. The performance of the proposed approach is compared to that of human operation by using the construction field data. The simulation results show that the proposed approach not only has the potential to replace human operation but also can significantly improve the comprehensive excavation performance.

Key words: Shield machine; Slurry shield; Intelligent tunnel boring machine (TBM); Deep reinforcement learning; Optimal control; Dynamic optimization; Deep learning

1 Introduction

Tunnel boring machine (TBM) technology has greatly improved the mechanization and automation level of the tunnel construction industry. However, the performance of a TBM still depends heavily on the operational skill of its operators. Manual operation relies on experience and different operators have different operational skill levels. This not only creates quality control issues but also causes similar accidents to happen repeatedly. Specifically, when encountering stratum changes or complex geological conditions, it is very difficult for operators to adjust parameters in a timely and effective manner. Thus, the empirical operation of operators has become one of the main limitations for the further improvement of the TBM performance. This situation has been deteriorating due to a decreasing number of skilled TBM operators.

Because of the huge demand for smart construction and of the rapid development of supporting technologies such as artificial intelligence (AI), big data, and control theory, the intelligent operation of TBMs has attracted significant research interest. Researchers have been working to increase the autonomy in the operation of TBMs since the 1980s. Clearly, to achieve partial or full autonomous operation in complex unstructured environments is a major trend in the development of TBM intelligent control technology. The main tasks of the shield machine operators include face support, excavation, and steering (adjustment of tunneling direction) operations. To automate these

✉ Guo-fang GONG, gfgong@zju.edu.cn
Guo-fang GONG, https://orcid.org/0000-0001-9553-8783

Received July 15, 2021; Revision accepted Dec. 7, 2021;
Crosschecked Apr. 20, 2022

three tasks, in-depth studies have been carried out around the world.

Regarding face support operation, Kuwahara and Harada (1988) developed a fuzzy controller for supporting pressure and excavation direction control to mimic the operation of skilled operators. Yeh (1997) applied a neural network to automate the earth pressure balance (EPB) control process. Liu et al. (2011) presented a predictive EPB control scheme based on a least squares support vector machine to replace manual operation. Shao and Lan (2014) developed an optimal control method that accounts for the tunnel face's stability. Zhang P et al. (2019) proposed a random-forest-based method to automatically control the settlement by regulating the operational parameters. Zhou et al. (2013) proposed a predictive control system for air chamber pressure in a slurry shield using Elman neural network. These studies are only a few examples of automated face support control methods. Previous studies have made valuable contributions towards the realization of autonomous face support pressure control.

In terms of steering operation, Ninić and Meschke (2015) developed a simulation-supported steering method for shield machines. Zhou et al. (2019a) developed a hybrid model integrating wavelet transform noise filter, convolutional neural network feature extractor, and long short-term memory predictor to realize the dynamic prediction of attitude and position in shield tunneling. Wang et al. (2018b) developed an automatic control system for shield pose and trajectory tracking. Xie et al. (2012) developed an automatic trajectory-tracking control system combining corrective trajectory planning with thrust cylinder control. These studies have laid a good foundation for the realization of autonomous steering.

Another significant aspect of shield operation is excavation, which is the main concern of this study. Shield excavation operation involves the decision-making and control of the operational parameters, including thrust force, cutterhead torque, advance speed, and rotational speed of the cutterhead. Numerous studies have been conducted to predict the shield operational parameters using supervised machine learning (ML) (Mahdevari et al., 2014; Namli and Bilgin, 2017; Salimi et al., 2018; Zhou et al., 2018, 2019b; Koopialipoor et al., 2019; Zhang et al., 2020a). These studies have made preliminary research for the realization of autonomous excavation. It is very natural to seek to achieve autonomous excavation by simply replacing the operators with an intelligent agent that can mimic the human operation. However, such methods have difficulty in improving excavation performance, and their potential is limited. Conventional knowledge-based intelligent methods, such as expert systems and fuzzy logic systems, represent the human experience as "if-then" rules or fuzzy reasoning rules. Meanwhile, supervised ML algorithms, such as artificial neural networks, support vector machine, and decision trees, learn from the data labeled or preprocessed by human designers. Essentially, these methods aim to act as humanly as possible. However, human behavior is often slow, empirical, and inaccurate, which makes these methods less likely to grasp the nature and objectives of the problem being addressed. It is therefore difficult for these methods to outperform humans. To further improve the excavation performance, it is necessary to achieve some kind of optimal excavation while achieving autonomous excavation, i.e. autonomous optimal excavation (AOE). Over the years, efforts have been made to optimize various aspects of TBM performance. Previous studies (Huo et al., 2010; Sun et al., 2011; Geng et al., 2015) optimized the cutterhead layout parameters, which are valuable in the design phase of TBMs. Wang et al. (2018a) developed a reliability-based multidisciplinary optimization method for determining the major structural and operating parameters of the hard rock TBM. Sun et al. (2018b) optimized the performance of the hard rock TBM using a collaborative optimization architecture. These methods are static optimization approaches that can only optimize the immediate performance measure under some fixed geological conditions. It is important to note that the long-term performance of the excavation process is significant. For example, the total energy consumption in a specified time interval is much more significant than the energy consumption for a given time step k. However, previous studies are limited to static optimization, and they cannot deal with ever-changing geological conditions and the long-term performance measure. The optimization of long-term excavation performance measure is a typical dynamic optimization problem and has not yet been dealt with.

Previous studies have provided various contributions to intelligent operation of TBMs. However, most of the related research efforts have achieved partial

automation for selected tasks and the various technologies have not been integrated to meet the requirements of a new generation of intelligent TBMs. To further advance the development of intelligent operation of TBMs, our research group has started a project to develop an intelligent operation system (IOS) for shield machines. The long-term goal of the IOS is to improve the efficiency, quality, safety, and degree of autonomy of the tunneling process. The development of the IOS requires significant interdisciplinary research effort as it integrates concepts and methods from areas, such as mechatronics, control, AI, ML, communication, and cyber-physical systems.

The IOS (version 1.0) is built upon four core modules, as shown in Fig. 1. The three modules at the bottom are designed to realize autonomous face support, excavation, and steering operations, respectively. The organizer module is in charge of the cooperation of the lower-level modules and the human-machine interaction. Ideally, the operation could be completed autonomously by the IOS, and the operator would only take control when the IOS operation is obviously inappropriate. Note that the steering issue is not the concern of the AOE module but the concern of the autonomous steering module. In addition to the AOE module, the other modules are currently being developed within our research group and are outside the scope of this paper. Our previous work (Zhang YK et al., 2020) has presented the development of the autonomous face support system for a slurry shield machine. It is important to point out that the operation of the IOS involves complex multi-process coupling effects. For example, the excavation process affects the face support process via the advance speed. The design of each IOS module must carefully consider these coupling effects. In the autonomous face support module, the advance speed was considered as a disturbance input variable for the slurry face-support control process. Accordingly, an autonomous controller was carefully designed such that the influence of the excavation process on the face-support process can be actively compensated (Zhang YK et al., 2020). The idea behind is that the supporting pressure control process should adapt to the advance speed, not the other way around. This paper focuses on the AOE module of IOS. It is assumed that ground surface settlement can be properly controlled by the autonomous face support module.

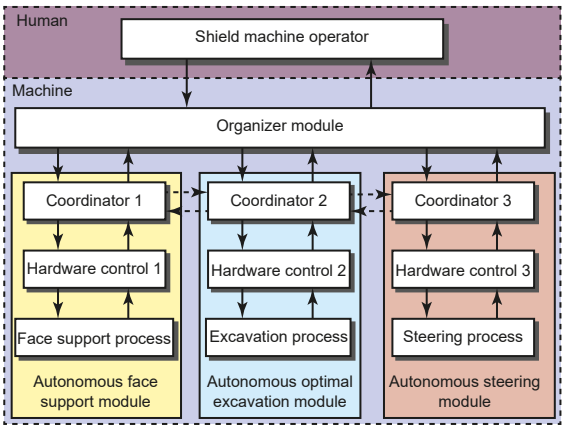

Fig. 1 IOS for shield machines

To resolve the problem of dynamic optimization of the shield excavation performance, as well as to achieve AOE, the following contributions are made in this paper:

(1) A high-accuracy hybrid modeling method that integrates first-principles analysis and a deep neural network is proposed for the machine-ground interaction dynamics of the excavation process.

(2) A dimensionless multi-objective comprehensive excavation performance measure suitable for the IOS is proposed.

(3) A novel AOE approach that integrates deep reinforcement learning (DRL) and the optimal control is proposed for shield machines, and its feasibility and effectiveness are validated.

The numerical results presented in this paper are based on the parameters and construction field data of an actual 6.5 m-diameter slurry shield machine, as shown in Fig. 2. The main parameters of the reference machine are shown in Table 1.

The remainder of the paper is organized as follows. Section 2 conducts a literature review and summarizes the technical difficulties. Section 3 presents

Fig. 2 Slurry shield machine manufactured by the China Railway Engineering Equipment Group Co., Ltd.

Table 1 Main parameters of the reference slurry shield

Parameter	Value
Diameter (m)	6.5
Cutterhead rotational speed (r/min)	0–3
Maximum advance speed (mm/min)	90
Maximum thrust force (kN)	4.26×10^4
Maximum torque (kN·m)	6.30×10^3
Total length (m)	90
Total weight (t)	500
Installed power (kW)	1787
Maximum supporting pressure (MPa)	0.6

the system modeling and analysis. The design of the AOE system is presented in Section 4. The performance evaluation and discussion are presented in Section 5, and finally, the main conclusions are presented in Section 6.

2 Literature review and technical difficulties

Due to the novelty and complexity of the AOE problem, several key issues need to be properly resolved. In this section, a literature review is conducted to justify the related research work, based on which the technical difficulties of achieving AOE are summarized.

Modeling of the machine-ground interactions of the excavation process is a key step in developing an AOE system. However, although operators can adjust the excavation operating parameters based on their intuition and experience, the machine-ground interaction dynamics and the multi-system coupling mechanism for the excavation process have not been well-understood. The existing work related to the modeling of the machine-ground interactions can be divided into two categories. The first category covers the first-principles modeling methods. To accurately predict the dynamic response of the machine-ground interaction, the first-principles modeling methods typically require a knowledge of the external dynamic load. There have been significant research efforts with the aim of establishing the torque and thrust load model for shield machines. The most widely used model in the industry for predicting installed cutterhead torque and thrust force is the empirical model suggested by the Japan Society of Civil Engineers. However, this model provides a rough estimate over a very wide range. To improve on this model, Zhang et al. (2010, 2013, 2014, 2015) developed a static mechanical model for the load acting on the cutterhead. Shi et al. (2011) presented a calculation method for the cutterhead torque based on composition analysis. To further improve accuracy, Wang et al. (2012) introduced penetration per revolution as a dynamic factor in the torque model for EPB shields. Han et al. (2017) developed a 3D finite element model for the dynamic load on the hard rock TBM cutterhead. These methods provide insights into cutterhead load modeling and are beneficial to the design and analysis of TBMs. However, most of these models are static models. Because of its complexity and interdisciplinary nature, it is very difficult to model the ever-changing dynamic load by using analytical approaches. The second category of model covers data-driven modeling methods. In recent years, ML and statistical methods have gradually been used to build cutterhead load models. Ates et al. (2014) developed statistical models to determine the design torque and thrust force for shield machines based on a database including the design parameters of 262 TBMs. Song et al. (2010) proposed a torque model for the EPB shield machine based on nonlinear regression. Sun et al. (2018a) developed a dynamic load prediction model for hard rock TBMs using the random forest algorithm. However, the datasets used to train these models were very small and consisted of hundreds of samples or less, and their generalization abilities are yet to be verified. Nowadays, advances in the field of ML, especially deep learning (DL), have generated new opportunities for the development of accurate dynamic models for machine-ground interaction based on the massive construction data of shields. DL is a class of ML algorithms that gain knowledge for a hierarchy of features (Zhang et al., 2018). The powerful function approximation and representation learning properties have made DL vastly more successful than previous approaches to ML. In contrast to first-principles modeling techniques, DL methods do not rely on the accurate modeling of the load. A generative black-box model for machine-ground interaction dynamics can be directly obtained using DL methods without intermediate calculation of the external load. In conventional ML, how to select the input features (also known as feature engineering) imposes a great challenge whereas DL completely automates the process of feature selection. A typical DL method learns all features in one pass rather than having to engineer them

manually. However, simply passing nearly 1500 variables of a shield machine raw dataset (Qin et al., 2021) as the input features without considering their physical meanings leads to confusion in the causal relationship between input and output variables. If the first-principles analysis could be combined with the powerful DL method, it would greatly improve interpretability while maintaining high accuracy. This type of hybrid modeling method allows the integration of all available knowledge into one approach. However, to the best of our knowledge, few hybrid modeling studies have yet been reported for TBMs. In addition, the excavation process driven by the operators involves multi-system coupling interactions, which is a typical human-cyber-physical system (HCPS) (Zhou J et al., 2019) and also, a hierarchical human-in-the-loop autonomous system (Antsaklis and Rahnama, 2018). To optimize overall performance, it is necessary to obtain insight into the effects of each subsystem and their interactions with each other. However, existing research has focused only on certain aspects of the system and the multi-system coupling mechanism for the excavation process is still unclear.

To achieve optimal excavation, optimization methods typically require an objective function to evaluate the performance of the system. At present, the most commonly used excavation performance specifications include the advance speed, rotational speed of the cutterhead, penetration rate, specific energy, and utilization rate. Each of these reflects an aspect of excavation performance. However, the excavation process typically involves contradictory multiple optimization goals that cannot be described using a single existing performance specification. For example, it is often desirable to have a high advance speed and utilization rate, while minimizing energy consumption. In practice, it does not make sense to optimize a certain specification without considering its constraints and other related specifications. Thus, a multi-objective performance measure that can evaluate the long-term comprehensive excavation performance needs to be defined.

To optimize the long-term excavation performance measure, dynamic optimization techniques such as dynamic programming (DP) and reinforcement learning (RL) are required. The main difference between dynamic and static optimization problems is that in the former, the decision at a current time step affects the possibilities at future time steps, and the optimizing agent needs to consider this effect when making a decision at the current time step (Busoniu et al., 2017). DP and RL have been widely used in the fields of robotics, autonomous driving, and adaptive control, etc. Carreras et al. (2005) presented a hybrid behavior-based scheme using RL for high-level control of autonomous underwater vehicles. Ng et al. (2006) accomplished an autonomous helicopter flight via RL. The recent successes in autonomous driving are fueled by DRL techniques (Shalev-Shwartz et al., 2016; Yu et al., 2016; El Sallab et al., 2017; Pan et al., 2017). However, there are some limitations to applying the existing DP and RL techniques directly to shield machines. Prior to deployment in actual machines, DP and RL typically need to interact with a high-fidelity training environment to learn the optimal policy. This cannot be accomplished without understanding and modeling the machine-ground interaction dynamics. In addition, most existing literature uses the RL agent to directly control the actuators of the system of interest without using feedback control. In this case, the RL agent is responsible for both action planning and actuator control. However, as a complex human-in-the-loop autonomous control system, the excavation process has a natural hierarchical structure. This class of system follows the principle of increasing precision with decreasing intelligence (IPDI) (Saridis, 2001). The principle of IPDI suggests that higher levels are concerned with slower aspects of the system's behavior and with its larger portions, or broader aspects (Antsaklis et al., 1991). There is an increase in the speed of decision-making with the transition from higher to lower levels, i.e., the actuator's control speed is typically faster than the action planning speed. Combining the action planning and actuator control functions into a single agent will greatly increase the complexity and training difficulty of the agent. To address this issue, a new field called hierarchical reinforcement learning (HRL) has attracted increasing attention (Dietterich, 2000). HRL decomposes a complex problem into sub-problems and then uses multiple RL agents to solve them individually. However, for a complex mechatronic system, HRL may not exploit the advantage of conventional feedback control. Because of the barriers between different disciplines, little attention has been focused on the integration of RL and conventional control to effectively exploit both approaches.

Based on the literature review, it could be summarized that there are three technical difficulties that prevent the achievement of AOE:

(1) A high-accuracy modeling method of the machine-ground interaction dynamics has not yet been established.

(2) There is still a lack of a comprehensive excavation performance measure suitable for the IOS.

(3) It is not appropriate to apply the existing dynamic optimization methods directly to shield machines.

This paper provides our solutions to these problems, such that the goal of AOE can be achieved.

3 System modeling and analysis

In this section, the hybrid modeling method for machine-ground interaction dynamics of the excavation process is presented. Then the dynamic models of the electro-hydraulic actuators are developed using first principles. Based on the overall system analysis, the multi-system coupling mechanism and the degrees of freedom (DOFs) for the AOE system design are revealed. The AOE problem is subsequently decomposed into a multi-objective dynamic optimization problem and an optimal control problem.

3.1 Hybrid modeling of the machine-ground interaction dynamics

The equation of rotational motion for the cutterhead can be expressed as

$$T-T_{\mathrm{r}}=\frac{2\pi}{60}J\dot{n}_{\mathrm{c}}, \tag{1}$$

where T is the total driving torque, T_{r} is the total resistant torque, J is the rotational inertia, and n_{c} is the cutterhead rotational speed. The equation of linear motion for the shield machine can be expressed as

$$F-F_{\mathrm{r}}=m\ddot{x}, \tag{2}$$

where F is the total thrust force, F_{r} is the total resistant force, and m and x are the total mass and the displacement of the shield, respectively.

Based on the results of the static analysis in the literature (Zhang et al., 2010, 2013, 2014, 2015; Shi et al., 2011) and given a fixed set of slurry shield machine geometric parameters, T_{r} and F_{r} can be expressed as Eqs. (3) and (4), respectively.

$$T_{\mathrm{r}}=f\left(n_{\mathrm{c}},\dot{x},F,P_{\mathrm{gw}},c,\varphi,t\right), \tag{3}$$

$$F_{\mathrm{r}}=g\left(n_{\mathrm{c}},\dot{x},P_{\mathrm{gw}},c,t\right), \tag{4}$$

where $f(\cdot)$ and $g(\cdot)$ are unknown non-linear time-variant functions, P_{gw} is the total pressure of the ground and water, c is the cohesion of the soil, φ is the internal friction angle of the soil, and t is the time. The specific representations of $f(\cdot)$ and $g(\cdot)$ are very difficult to obtain analytically due to their complexity and interdisciplinary nature. The state vector $\boldsymbol{x}$ is chosen as follows:

$$\boldsymbol{x}=[\,x_1\quad x_2\,]^{\mathrm{T}}=[\,n_{\mathrm{c}}\quad \dot{x}\,]^{\mathrm{T}}. \tag{5}$$

The input vector $\boldsymbol{u}$ and output vector $\boldsymbol{y}$ are defined as Eqs. (6) and (7), respectively. It should be noted that u_1 and u_2 are taken as manipulated inputs $\boldsymbol{u}_{\mathrm{m}}$, while u_3 and u_4 are treated as disturbance inputs $\boldsymbol{u}_{\mathrm{d}}$.

$$\begin{aligned}\boldsymbol{u}&=[\,\boldsymbol{u}_{\mathrm{m}}\quad \boldsymbol{u}_{\mathrm{d}}\,]^{\mathrm{T}}\\&=[\,u_1\quad u_2\quad u_3\quad u_4\quad u_5\,]^{\mathrm{T}}\\&=[\,T\quad F\quad P_{\mathrm{gw}}\quad c\quad \varphi\,]^{\mathrm{T}},\end{aligned} \tag{6}$$

$$\boldsymbol{y}=[\,y_1\quad y_2\,]^{\mathrm{T}}=[\,n_{\mathrm{c}}\quad \dot{x}\,]^{\mathrm{T}}=[\,x_1\quad x_2\,]^{\mathrm{T}}. \tag{7}$$

The non-linear state space representation of the process can then be expressed as

$$\dot{\boldsymbol{x}}=\varphi(\boldsymbol{x},\boldsymbol{u},t), \tag{8}$$

$$\boldsymbol{y}=\begin{bmatrix}1&0\\0&1\end{bmatrix}\boldsymbol{x}, \tag{9}$$

where $\varphi(\cdot)$ is an unknown non-linear time-variant function related to $f(\cdot)$ and $g(\cdot)$.

This first-principles model applies to both EPB and slurry shield machines. Due to the unknown functions, this model cannot be used for quantitative calculations, but only for qualitative analysis. Nevertheless, the input and output variables of the excavation process are fully determined by first-principles analysis. Note that for a specific tunnel constructed by a shield machine, the geometric parameters of the shield are constants. The influence of the geometric parameters is reflected in the term of m in Eq. (2), and thereby m

appears as a constant parameter in Eq. (8). The depth and thickness of the soil stratum are often considered in the literature (Chen et al., 2019; Zhang P et al., 2020b). In this study, its depth and thickness are directly reflected in the term of P_{gw} in Eqs. (3) and (4), and thereby they are implicitly included in Eq. (8).

Using the input and output variables determined by first-principles analysis, a data-driven model can be further developed to calculate the dynamic response, leading to a hybrid model. The power of the proposed hybrid modeling method lies in the improvement of the interpretability of the model, which not only leads to a better understanding of the process, but also helps overcome the curse of dimensionality. To take advantage of the powerful function approximation and representation learning properties of DL, a deep neural network (DNN) with a multi-layer perceptron (MLP) structure was adopted, as indicated in Fig. 3.

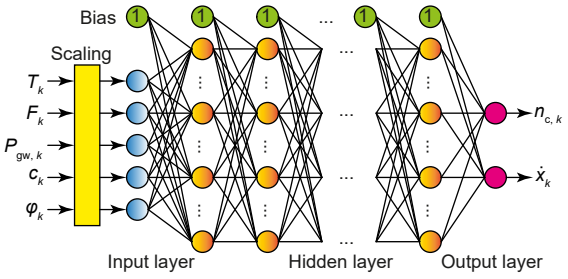

Fig. 3 Structure of the DNN model of the machine-ground interaction dynamics. *k* is the time step

Before being transported to the input layer of the DNN, the input variables are scaled by dividing them using their corresponding maximum values. There are many hyper-parameters for the MLP, including the number of layers, the number of neurons per layer, the type of activation function to use in each layer, and the weight initialization logic. To reduce the number of hyper-parameters that have to be tuned, an equal number of neurons was used in each hidden layer.

The dataset was obtained from the construction field data of a 6.5 m-diameter slurry shield TBM, in which the geological data was extracted from the geological exploration report. The comprehensive classification of the surrounding rock is grade VI, the working face has no self-stability ability, and the surrounding rock is easily collapsed and deformed. According to the geological exploration report, the soil cross-sectional characteristics are relatively uniform, so the influence of soil layering on the cohesion and internal friction angle values is ignored.

The excavation of the slurry shield machine is a periodical process that consists of start-up, stable, and shutdown phases. Given that the excavation performance is unstable during the start-up and shutdown phases, the data collected during these two phases were excluded. On the other hand, the raw construction field data is not only noisy but also contains outliers. To improve the data quality, a Savitzky-Golay filter was used for de-noising, and the isolation forest algorithm was used to remove outliers. For the Savitzky-Golay filter, the filter window length and the polynomial order were set as 1001 and 2, respectively.

After data preprocessing, a total of 1.1 million samples was used to train, validate, and test the DNN. Summary statistics of the dataset are presented in Table 2. The "Mean," "Min," and "Max" rows are self-explanatory, while the "Std" row shows the standard deviation. The "25%," "50%," and "75%" rows show the corresponding percentiles and a percentile indicates the value below which a given percentage of observations in a group of observations falls. The dataset was randomly split into a training set, validation set, and test set in the ratio 70-15-15.

To determine the optimum hyper-parameters, various DNNs were trained and compared. The mean squared error (MSE) was used as the cost function. The "Relu" (rectified linear units) function was used as the activation function because it does not saturate

Table 2 Summary statistics of the dataset

Item	T (kN·m)	F (kN)	n_c (r/min)	$\dot{x}$ (mm/min)	P_{gw} (×10^5 Pa)	c (×10^6 Pa)	φ (°)
Mean	2.03×10^3	2.00×10^4	1.35	13.23	3.49	25.88	12.35
Std	8.65×10^2	5.17×10^3	0.20	17.38	0.56	2.38	2.35
Min	4.17×10^2	4.18×10^3	0.75	1.08	0.69	14.90	7.60
25%	1.22×10^3	1.51×10^4	1.21	3.27	3.05	24.90	9.40
50%	2.20×10^3	2.21×10^4	1.33	4.92	3.28	27.50	14.30
75%	2.73×10^3	2.39×10^4	1.53	12.76	4.00	27.50	14.30
Max	3.94×10^3	2.87×10^4	2.02	90.85	4.63	29.90	16.70

for positive values and also because it is quite fast to compute. In addition, the "He" weight initialization technique (He et al., 2015) was used to prevent the vanishing gradients problem in the DNN, while the adaptive moment estimation (Adam) (Kingma and Ba, 2015) optimizer was used to speed up the training process. The early stopping technique was used to prevent overfitting, i.e., the training process is stopped automatically if the specified number of epochs with no improvement on the validation set MSE is reached.

The performance of the DNNs is illustrated in Table 3 and Fig. 4. It is found that the DNN with four hidden layers and 200 neurons in each layer has the best performance, whereas the DNN with five hidden layers performed poorly. After a comprehensive trade-off between the performance and complexity, the DNN with four hidden layers containing 200 neurons in each layer was finally selected. The performance of the selected DNN for a randomly extracted 400000 construction dataset is shown in Fig. 5 with mean relative error (MRE) and mean absolute error (MAE) values, which demonstrates that the model can describe the machine-ground interaction dynamics and so validates its effectiveness. The definitions of MRE and MAE are given by Eqs. (10) and (11), respectively.

$$\mathrm{MRE}=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\left|y_i-\hat{y}_i\right|}{y_i}\right)\times 100\%, \tag{10}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|y_i-\hat{y}_i\right|, \tag{11}$$

where y_i is the target value, $\hat{y}_i$ is the prediction value, and N is the total number of samples considered.

3.2 Modeling of the actuators

We now consider the processes in which the actuators provide the thrust force and torque.

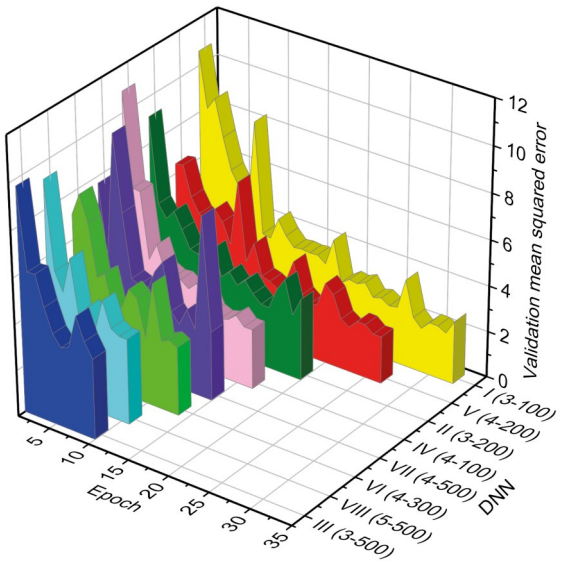

Fig. 4 Learning curves of various DNNs

The thrust force F is provided by 32 hydraulic cylinders connected in parallel. Each hydraulic cylinder is controlled by a three-way proportional pressure reducing valve. Assuming that the load is evenly applied to each hydraulic cylinder and considering a single valve-controlled-cylinder as shown in Fig. 6, the linearized flow equation of the proportional pressure reducing valve is:

$$Q_\mathrm{l}=K_\mathrm{q}x_\mathrm{v}-K_\mathrm{c}P_\mathrm{l}, \tag{12}$$

where Q_l is the load flow, K_q is the flow gain of the valve, x_v is the displacement of the spool, K_c is the flow-pressure coefficient of the valve, and P_l is the load pressure.

The equation of the proportional solenoid coil terminal voltage can be expressed as:

$$u_\mathrm{c}=L\dot{I}+R_\mathrm{c}I+K_\mathrm{e}\dot{x}_\mathrm{v}, \tag{13}$$

Table 3 Performance comparison of various DNNs

DNN	Number of neurons in each layer	Number of layers	Training epochs	MSE		
				Training set	Validation set	Test set
I	100	3	33	2.50	2.74	2.74
II	200	3	20	3.16	3.56	3.43
III	500	3	10	3.71	3.50	3.57
IV	100	4	17	2.94	2.61	2.62
V	200	4	27	2.60	2.06	2.06
VI	300	4	14	3.47	2.98	2.95
VII	500	4	15	3.33	3.01	3.07
VIII	500	5	11	4.09	3.71	3.75

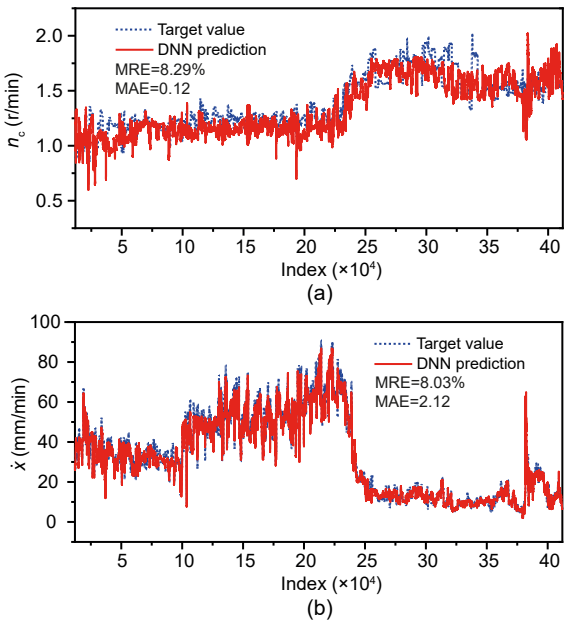

Fig. 5 Comparison of the DNN predictions and the target measurement values: (a) cutterhead rotational speed; (b) advance speed

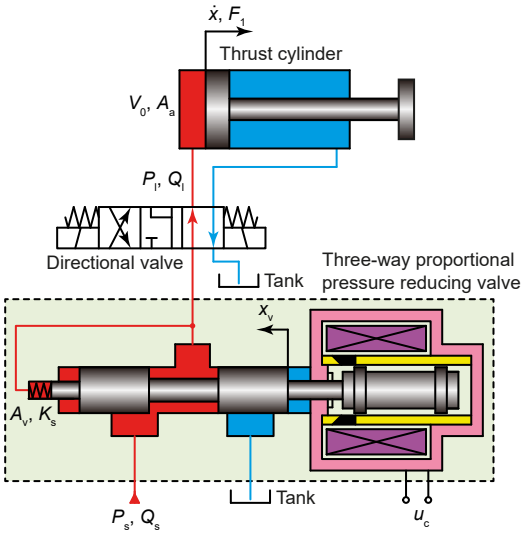

Fig. 6 Schematic diagram of the thrust electro-hydraulic system. P_s and Q_s are the supply pressure and flow of the hydraulic oil, respectively

where u_c is the terminal voltage of the proportional solenoid coil, L is the coil inductance, I is the coil current, R_e is the coil resistance, and K_e is the reverse electromotive force coefficient.

The equation of motion of the armature assembly (including the valve spool) can be expressed as:

$$F_m - K_s x_v - B_v \dot{x}_v - P_l A_v - K_f x_v = m_v \ddot{x}_v, \quad (14)$$

where F_m is the output force of the proportional solenoid, K_s is the spool reset spring stiffness, B_v is the spool viscous damping coefficient, A_v is the effective area, K_f is the steady-state flow force coefficient, and m_v is the total mass of the spool and the armature.

The output force equation of the proportional solenoid is:

$$F_m = K_I I - K_x x_v, \quad (15)$$

where K_I and K_x are respectively the current-force gain and the displacement-force gain of the proportional solenoid.

The continuous equation of the hydraulic cylinder is:

$$Q_l = A_a \dot{x} + \frac{V_0}{\beta_e} \dot{P}_l + C_t P_l, \quad (16)$$

where A_a is the effective area of the piston, V_0 is the volume of the high-pressure chamber of the cylinder and the connecting line, β_e is the effective bulk modulus of the oil, and C_t is the total leakage coefficient.

The thrust force provided by a single hydraulic cylinder is:

$$F_l = P_l A_a = \frac{F}{N_c}, \quad (17)$$

where F_l is the output force for a single thrust cylinder, and N_c is the number of hydraulic cylinders. Taking the Laplace transform of Eqs. (12)–(17), and eliminating the intermediate variables, F_l can be expressed as Eq. (18) using the parameters listed in Table 4, where s is the Laplace variable.

$$F_l(s) = G_1 u_c(s) - G_2 \dot{x}(s), \quad (18)$$

Table 4 Main parameters of the actuators

Parameter	Value	Parameter	Value
m_v (kg)	7.92×10^{-3}	K_c (m^3/(s·Pa))	1.7×10^{-7}
B_v (N·s/m)	7.01	V_0 (m^3)	0.23×10^{-2}
K_s (N/m)	300	β_e (N/m^2)	7×10^{8}
K_f (N/m)	300	C_t (m^3/(s·Pa))	3×10^{-11}
K_x (N/m)	62.25	A_a (m^2)	0.04
K_I (N/A)	7.5	C_{tm} (m^3/(s·Pa))	3×10^{-11}
L	1.2×10^{-3}	D_m (m^3/rad)	7.96×10^{-5}
R_e (Ω)	1.2	k_{mr}	54.44
K_q (m^2/s)	64.7	k_{rc}	7.41
A_v (m^2)	1.77×10^{-6}	η	0.75

where

$$G_1 = 5.91\times10^{17}/(s^4+5.36\times10^4 s^3+9.97\times10^7 s^2+1.16\times10^{11}s+8.73\times10^{12}), \quad (19)$$

$$G_2 = (4.39\times10^8 s^3+8.29\times10^{11}s^2+9.46\times10^{14}s+3.68\times10^{16})/(s^4+5.36\times10^4 s^3+9.97\times10^7 s^2+1.16\times10^{11}s+8.73\times10^{12}). \quad (20)$$

The driving torque T is provided by seven hydraulic motors via reducers and a gear-ring mechanism, and each motor is controlled using a three-way proportional pressure reducing valve. Assuming that the load is evenly applied to the motors and considering a single valve-controlled-motor as shown in Fig. 7, Eqs. (12)–(15) can also be applied to the cutterhead drive system.

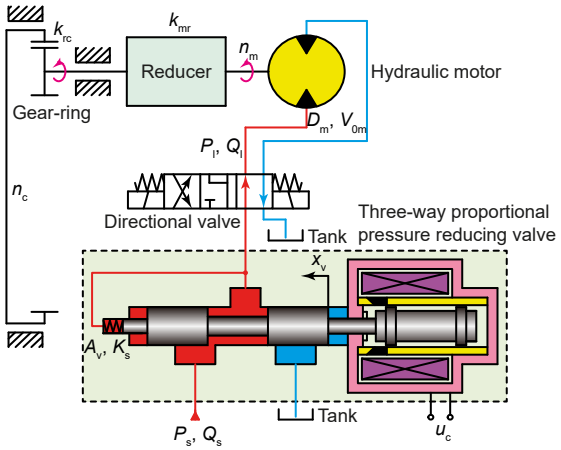

Fig. 7 Schematic diagram of the cutterhead drive electro-hydraulic system

The flow equation of the hydraulic motor is:

$$Q_l = \frac{2\pi}{60}D_m n_m + \frac{V_{0m}}{\beta_e}\dot{P}_l + C_{tm}P_l, \quad (21)$$

where D_m is the motor displacement, n_m is the motor rotational speed, V_{0m} is the volume of the high-pressure chamber of the motor and the connecting line, and C_{tm} is the total leakage coefficient of the motor.

The relationship between the rotational speed of the motor and cutterhead can be expressed as

$$n_m = k_{mr}k_{rc}n_c, \quad (22)$$

where k_{mr} and k_{rc} are the transmission ratios of the reducer and gear-ring, respectively.

The relationship between the output torque of the motor and of the cutterhead can be expressed as

$$T_l = \eta k_{mr}k_{rc}T_m = \frac{T}{N_m} = D_m P_{pl}, \quad (23)$$

where T_l is the output cutterhead torque given by a single motor, η is the transmission efficiency, T_m is the torque of a motor, N_m is the number of motors, and P_{pl} is the inlet port pressure of the motor. Taking the Laplace transform of Eqs. (12)–(15) and (21)–(23) and eliminating the intermediate variables, T_l can be expressed as Eq. (24) using the parameters listed in Table 4.

$$T_l(s) = H_1 u_c(s) - H_2 n_c(s), \quad (24)$$

where

$$H_1 = 1.51\times10^{17}/(s^4+2.21\times10^4 s^3+4.02\times10^7 s^2+4.53\times10^{10}s+3.46\times10^{12}), \quad (25)$$

$$H_2 = (9.61\times10^6 s^3+1.81\times10^{10}s^2+2.07\times10^{13}s+8.04\times10^{14})/(s^4+2.21\times10^4 s^3+4.02\times10^7 s^2+4.53\times10^{10}s+3.46\times10^{12}). \quad (26)$$

In practice, the actuators can be controlled using closed-loop feedback controllers. Even open-loop control could be used when the accuracy requirement is not critical.

3.3 Overall system analysis

Assuming that the actuators are controlled using closed-loop feedback controllers and, based on the models obtained in Sections 3.1 and 3.2, the conventional shield machine excavation process driven by the operator can be illustrated as the block diagram shown in Fig. 8. This is a typical HCPS, in which the operator represents the human part, the actuator feedback controllers represent the cyber part, and the dynamics of the actuators and the machine-ground interaction represent the physical part. This is also a hierarchical human-in-the-loop autonomous control system, in which the operator represents the coordination level, the feedback controllers and the dynamics of the actuators comprise the execution level, and the

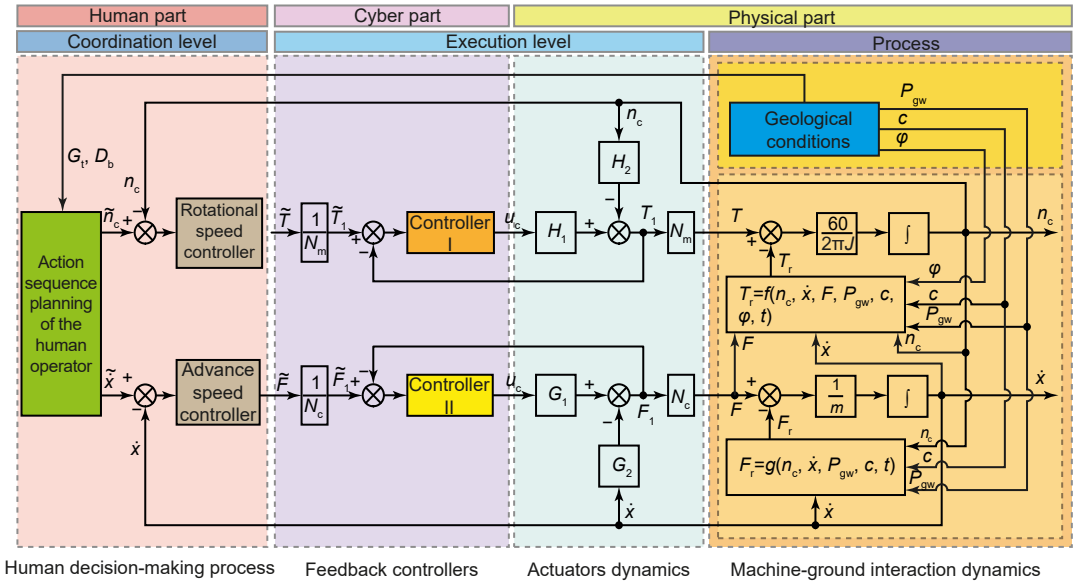

Fig. 8 Block diagram of the conventional excavation process for shield machines

machine-ground interaction is the ultimate process to be controlled.

The coupling relationships between different parts of the system and the effect of each part are clearly illustrated in Fig. 8. The overall effect of the operator as the coordination level is to provide the desired reference inputs of the driving torque $\tilde{T}$ and thrust force $\tilde{F}$ based on the geological conditions and the value of the actual cutterhead rotational speed n_c and advance speed $\dot{x}$. The geological condition information used by the operator typically includes the ground type G_t and the shield burial depth D_b. Specifically, there are three aspects that define the role of the operator. The first is to perform action sequence planning based on the geological conditions to provide the reference inputs for the desired cutterhead rotational speed $\tilde{n}_c$ and advance speed $\tilde{\dot{x}}$. The second is to compare the error between the reference inputs and the actual values of n_c and $\dot{x}$. The third is to determine $\tilde{T}$ and $\tilde{F}$ via mental controllers based on the errors. The task of the execution level is to accurately execute the commands from the coordination level to complete the excavation task. In the case of open-loop control of the actuators, the decision-making requirements of the operator are even more onerous, because the operator must adjust the reference input commands more frequently to deal with execution errors.

Assuming that the hardware and control software of the system are set up and taking everything outside the coordination level as the environment, the conventional excavation process shown in Fig. 4 can be viewed as interactions between the coordination level and its environment. We denote the input variables to the coordination level at time step k as observation $o_{b,k}$, and the output variables from the coordination level at time step k as action a_k. Starting at some initial observation $o_{b,0}$, the operator chooses an action command a_0 for the execution level. As a result, the state of the excavation process transitions to a successor state $o_{b,1}$ according to the dynamics of the execution level and the machine-ground interaction. Then the operator gets to choose another action a_1. As a result of this action, the state transitions again to a state $o_{b,2}$, and so on. The operator makes a sequence of decisions, one for each period. The goal of the AOE is to automatically choose the action values at each step and apply it to its environment, such that some predefined excavation performance objective can be achieved. Supposing that some method has been used to automatically obtain the optimal action sequence to replace human decision-making, the effect of the action execution also depends on the execution accuracy of the execution level. Obviously, the optimal action sequence trajectory obtained from the coordination level must be executed as accurately as possible to guarantee optimum performance.

Based on the abovementioned analysis, there are two equally important DOFs for the AOE system design. The first DOF is to design an intelligent agent for the coordination level to automatically provide the

optimal action (reference input) sequence concerning the task goal. The second DOF is to optimize the design of the feedback controllers for the actuators such that the desired actual actions can be applied to the machine-ground interaction as accurately as possible. Once the hardware and control software of the system are set up, the machine-ground interaction output $\boldsymbol{y}$ is only determined by the reference input commands from the coordination level. As stated in Section 2, the optimization of long-term excavation performance is a typical dynamic optimization problem that involves contradictory multiple optimization goals. Therefore, in this study, the AOE problem is decomposed into a multi-objective dynamic optimization problem and an optimal control problem.

To obtain the optimal action sequence, the excavation process driven by the coordination level can be further modeled as a Markov decision process (MDP). An MDP model is a five-tuple (O_b, A, h, γ, ρ), where O_b is the observation space. For the operator, $O_\mathrm{b}=\{\boldsymbol{x}, G_\mathrm{t}, D_\mathrm{b}\}$. $A=\{\tilde{T}, \tilde{F}\}$ is the action space, $h:O_\mathrm{b}\times A\mapsto O_\mathrm{b}$ is the state transition function of the excavation process. $\rho:O_\mathrm{b}\times A\mapsto\mathbb{R}$ is the reward function that evaluates the immediate excavation performance, and $\gamma\in[0, 1]$ is the discount factor.

The reward evaluates the immediate effect of action a_k, namely the transition from $o_{\mathrm{b},k}$ to $o_{\mathrm{b},k+1}$, but it does not provide any information about its long-term effects. Starting from the initial observation $o_{\mathrm{b},0}$, the long-term performance of the action sequence can be measured by a value function, i.e., the discounted total reward accumulated over the course of interactions, as indicated in Eq. (27).

$$V^h\left(o_{\mathrm{b},0}\right)=\sum_{k=0}^{N}\gamma^k r_{k+1}=\sum_{k=0}^{N}\gamma^k\rho\left(o_{\mathrm{b},k},h\left(o_{\mathrm{b},k}\right)\right). \quad (27)$$

The optimal action sequence trajectory (also called the policy) can be found by solving the constrained discrete dynamic optimization problem defined in Eq. (28).

$$\pi^*\left(o_\mathrm{b}\right)=\arg\max_{a\in A}\left(\sum_{k=0}^{N}\gamma^k r_{k+1}\right). \quad (28)$$

Typically, this class of problems can be solved using DP and RL techniques. It should be noted that the discount factor γ represents the importance of future rewards. If γ<1, the value function will converge to a finite value. If γ=0, the optimization algorithm will have no interest in future rewards but will try to maximize the reward only for the current state. If γ=1, the optimization algorithm will try to increase future rewards even at the expense of the immediate ones.

In this study, a DRL method is used to solve for the optimal action sequence trajectory, and optimal closed-loop feedback controllers are developed to achieve accurate execution.

4 Design of the autonomous optimal excavation system

4.1 Autonomous optimal excavation scheme overview

Based on the two DOFs for the AOE system design revealed in Section 3.3, we propose a hierarchical AOE system scheme that integrates DRL and optimal control to effectively exploit both AI and closed-loop feedback control, as shown in Fig. 9. At the coordination level, the deep deterministic policy gradient (DDPG) algorithm is used as the DRL agent because it can work with continuous action space. The digital optimal feedback controllers are used at the execution level to achieve accurate execution. It should be noted that to simplify the block diagram, Fig. 9 uses a single closed-loop to represent the execution level because the block diagrams of the actuators have the same structure. In particular, N represents N_c and N_m, TF_1 represents G_1 and H_1, and TF_2 represents G_2 and H_2. The working frequencies of the DRL agent and the execution level are set as 1 and 100 Hz, respectively.

The DRL agent learns by interacting with its environment via action, observation, and reward. The observation and reward are the results of the action. Based on the analysis in Section 3.3, the action space is set as $A=\{\tilde{T},\tilde{F}\}$ for the DRL agent. To effectively utilize the geological information and the machine-ground interaction states, the observation space is set as $O_\mathrm{b}=\{\boldsymbol{x}, \boldsymbol{u}_\mathrm{d}\}=\{n_\mathrm{c}, \dot{x}, P_\mathrm{gw}, c, \varphi\}$. At each time step, the DRL agent performs an action on the environment and obtains feedback in terms of observation and reward.

To train the DRL agent, a high-fidelity training environment is required. For the DRL agent, the environment comprises four different components: the execution level, information on the geological condition, the machine-ground interaction dynamics, and

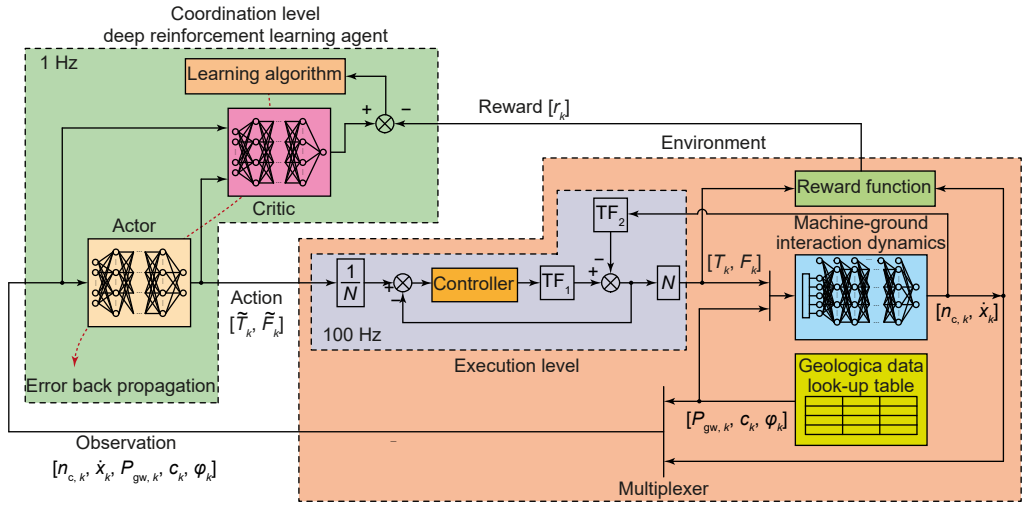

Fig. 9 Autonomous optimal excavation system that integrates deep reinforcement learning and closed-loop feedback optimal control

the reward function. The geological condition information can be implemented as a look-up table using data extracted from the geological exploration report. A high-accuracy DNN model has been developed for the machine-ground interaction dynamics in Section 3.1. The reward function will be defined in Section 4.2. Based on the models of the actuators obtained in Section 3.2, digital optimal controllers can be designed using control theory. The execution level can then be implemented as transfer functions.

After training, the proposed AOE system can be deployed in an actual shield machine, and the machine-ground interaction dynamics block is replaced with the actual machine-ground interaction data obtained from the sensors.

4.2 Definition of the multi-objective comprehensive excavation performance measure

In practice, the advance speed is the most important excavation performance specification because it directly determines the construction time. In addition, a good performance in terms of energy efficiency is also desirable. Excavation energy consumption not only determines the cost of electricity consumption but also is an important indicator of the state of excavation. For example, abnormal conditions such as encountering difficult grounds and a stuck of the cutterhead are often accompanied by abnormally high energy consumption. From the perspective of improving excavation quality, low energy consumption is preferred, however, it should be maintained at a reasonable level in consideration of normal excavation constraints. Obviously, the goals of increasing the advance speed and reducing the energy consumption are contradictory. In this study, we propose a multi-objective excavation performance measure suitable for the IOS as follows:

$$J_k = k_1 \bar{\dot{x}}_k - k_2 \bar{E}_k, \tag{29}$$

where

$$\bar{E}_k = \frac{4}{\pi D^2}\left(\frac{\bar{n}_{c,k}\bar{T}_k}{L\left(\bar{\dot{x}}_k\right)} + \bar{F}_k\right), \tag{30}$$

$$L(x) = \begin{cases} x, & x \geqslant 1\times 10^{-5}, \\ 1\times 10^{-5}, & x < 1\times 10^{-5}, \end{cases} \tag{31}$$

$$\begin{bmatrix} \bar{\dot{x}}_k \\ \bar{n}_{c,k} \\ \bar{T}_k \\ \bar{F}_k \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\dot{x}_{\max}} & 0 & 0 & 0 \\ 0 & \dfrac{1}{n_{c,\max}} & 0 & 0 \\ 0 & 0 & \dfrac{1}{T_{\max}} & 0 \\ 0 & 0 & 0 & \dfrac{1}{F_{\max}} \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ n_{c,k} \\ T_k \\ F_k \end{bmatrix}, \tag{32}$$

where $\bar{\dot{x}}_k$, $\bar{n}_{c,k}$, $\bar{T}_k$, and $\bar{F}_k$ are the dimensionless advance speed, cutterhead rotational speed, total torque, and total thrust force at time step k, respectively. These variables are scaled to the interval (0, 1] by dividing by their corresponding maximum values as shown in

Eq. (32). D is the diameter of the shield. Taking $4/(\pi D^2)$ as a weighting constant, $\bar{E}_k$ is the dimensionless specific energy consumption. The physical meaning of $\bar{E}_k$ is the energy consumption per meter weighted by $4/(\pi D^2)$. To prevent the problem of dividing by zero, a piecewise function $L(x)$ is used to restrict the minimum value of the denominator of Eq. (30). The parameters k_1 and k_2 are the relative importance weights.

The value of J_k is in the open interval of $(-1, 1)$. The closer the J_k value is to -1, the worse the excavation quality. By contrast, the closer the value is to 1, the higher the excavation quality. Such a definition of excavation performance measure is not only simple and easy to understand but also has a clear physical meaning. The trade-off between the advance speed and specific energy consumption can be achieved by changing the relative sizes of k_1 and k_2.

In practice, the actual action values applied to the machine-ground interaction, and the resulting state values obtained, are bounded by their corresponding minimum and maximum values. By adding penalization terms to the objective function defined in Eq. (29), the constrained optimization problem defined in Eq. (28) can be reformulated as an unconstrained optimization problem and subsequently solved. This is because the penalization terms have the effect of discouraging constraint violations. Thus, the reward function is defined as:

$$r_k = J_k + P_{\mathrm{sgn},k} + P_{\min,k} + P_{\max,k}, \tag{33}$$

where $P_{\mathrm{sgn},k}$, $P_{\min,k}$, and $P_{\max,k}$ represent the sign, minimum value, and maximum value penalty terms, respectively, as defined in Eqs. (34)–(38).

$$P_{\mathrm{sgn},k} = N\left(\bar{T}_k\right) + N\left(\bar{F}_k\right) + 10N\left(\bar{n}_{\mathrm{c},k}\right), \tag{34}$$

$$P_{\min,k} = 100L\left(\bar{T}_k, \bar{T}_{k\min}\right) + 10L\left(\bar{F}_k, \bar{F}_{k\min}\right), \tag{35}$$

$$P_{\max,k} = N\left(1 - \left|\bar{T}_k\right|\right) + N\left(1 - \left|\bar{F}_k\right|\right) + N\left(1 - \left|\bar{n}_{\mathrm{c},k}\right|\right) + N\left(1 - \left|\dot{\bar{x}}_k\right|\right), \tag{36}$$

$$N(x) = \begin{cases} x, & x < 0, \\ 0, & x \geqslant 0, \end{cases} \tag{37}$$

$$L\left(x, x_{\min}\right) = \begin{cases} x - x_{\min}, & x < x_{\min}, \\ 0, & x \geqslant x_{\min}. \end{cases} \tag{38}$$

Based on the definition of the reward function, the long-term comprehensive excavation performance can be measured by the discounted cumulative reward defined in Eq. (27).

4.3 Design of the optimal controllers for the actuators

The digital optimal controllers of the actuators were designed and automatically tuned using the direct automatic tuning (DAT) method (Zhang YK et al., 2019). The resulting digital optimal controller of the thrust system and the cutterhead drive system is given in Eqs. (39) and (40), respectively.

$$\begin{aligned} C_{\mathrm{T}}(z) = {} & (6.49\times10^{-4}z^3 + 3.15\times10^{-4}z^2 + \\ & 1.11\times10^{-4}z - 1.63\times10^{-4}) \Big/ \{(z-1)(z^3 + \\ & 1.37\times10^{-5}z^2 + 5.12\times10^{-5}z + 5.33\times10^{-5})\}, \end{aligned} \tag{39}$$

$$\begin{aligned} C_{\mathrm{D}}(z) = {} & (8.36\times10^{-4}z^3 + 4.86\times10^{-4}z^2 + \\ & 9.77\times10^{-5}z - 1.91\times10^{-4}) \Big/ \{(z-1)(z^3 + \\ & 4.38\times10^{-6}z^2 + 2.25\times10^{-6}z + 8.32\times10^{-7})\}. \end{aligned} \tag{40}$$

Fig. 10 shows the closed-loop step responses of the actuators. The DAT method yields robust and fast controllers without steady-state errors. Meanwhile, noting that the response speed of the execution level is much higher than that of the coordination level. Based on these facts, the implementation of the execution level can be further simplified by setting $\tilde{T}_k = T_k$ and $\tilde{F}_k = F_k$.

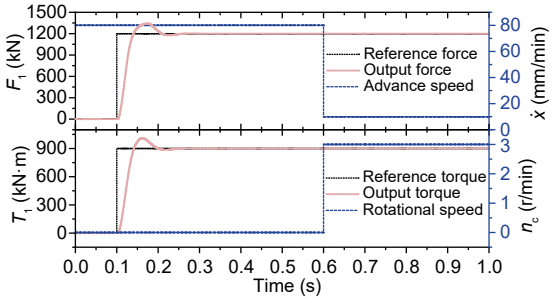

Fig. 10 Closed-loop responses of the actuators

4.4 Implementation of the training environment

The pseudo-code for the implementation of the training environment is given in the electronic supplementary material. During the training of the DRL agent, the flow of execution first goes into the initialization part. The "step" function is then continuously called in a "for" iteration loop until the specified number of training steps is reached. Based on the result

obtained in Section 4.3, the program first assigns the action ($\tilde{T}_k$, $\tilde{F}_k$) given by the DRL agent to the actual action (T_k, F_k) applied to the machine-ground interaction. The program then determines whether the specified number of training steps is reached; if not, the program reads geological data from the look-up table using the current step number. After data scaling, the action and geological data are sent to the DNN model to predict the corresponding rotational speed $n_{c,k}$ and advance speed $\dot{x}_k$. The reward value r_k can then be calculated using Eq. (33). The obtained observation and reward information is then sent to the "Is_done" function to determine if there is a severe constraint violation, and it returns a logical variable: "True" or "False". Finally, the information of the calculated observation, reward value, and "Done" value is returned to the DRL agent. Inside the DRL agent, the "Done" information is a logical flag that indicates that the training episode is complete. When the number of iterations reaches a specified number, the flow of execution exits the iteration loop.

4.5 Design of the deep reinforcement learning agent

The DDPG algorithm is a model-free, online, off-policy, actor-critic RL method proposed by the Deepmind team of Google in 2015 (Lillicrap et al., 2016). A DDPG agent uses an actor to decide which action should be taken and uses a critic to approximate the long-term reward. Typically, the actor and critic are implemented as artificial neural networks. As shown in Fig. 9, the actor takes observation as input and outputs the corresponding action that maximizes the long-term reward. The critic takes observation and action as inputs and outputs the corresponding long-term reward. It is used to assist in training the actor.

A brief description of the DDPG algorithm is as follows. During training, a DDPG agent first utilizes the current observation as an input, selects an action, and perturbs the action using a stochastic noise model at each training step. Then the agent executes the action and observes the reward and the next observation. Subsequently, the agent stores the past experience using a circular experience buffer. Finally, the agent updates the actor and critic using a mini-batch of experiences that are randomly sampled from the buffer. This process is repeated until the specified number of the training steps is achieved.

In this study, the actor and critic were implemented as DNNs using a similar MLP structure as shown in Fig. 3. These DNNs were both designed to have three hidden layers and 200 neurons in each layer.

It is widely appreciated that DRL agents are prone to overfitting (Cobbe et al., 2019). As such, it is a common practice to train and test the DRL agent using the same set of environments. However, the complexity of geological conditions and the high cost of training requires that the DRL agent, as the coordination level, must have a good generalization ability to cope with situations that have not been previously encountered. To overcome this problem, L2 regularization with a factor of 0.01 was used in each layer of the actor and critic DNNs.

5 Performance evaluation and discussion

5.1 Performance evaluation

The performance of the proposed AOE system was compared to that of the human operation by using the construction field data. To facilitate comparisons, the allowed action values for the DRL agent were strictly constrained in the range obtained by the human operators for the same segment of construction field data. In this setting, we trained three DRL agents by using three reward functions with different k_1 and k_2 values to investigate their action characteristics and the resulting excavation performances. Denote these three reward functions as reward I, reward II, and reward III, respectively. The relative importance weights were set as follows: k_1=0.8 and k_2=0.2 for reward I, k_1=0.6 and k_2=0.4 for reward II, and k_1=0.5 and k_2=0.5 for reward III. Denote the resulting AOE systems as AOE system I, AOE system II, and AOE system III, respectively. The same execution level was used for these three AOE systems. The discount factor γ was set as 0.9.

Training a DRL agent is very time-consuming. Thus, the DRL agents were trained on a dataset consisting of a selected representative 150000 samples. The data from the index of 150000 to 300000 shown in Fig. 5 was selected as the training set because it includes both high-speed and low-speed excavation segments. The data from the index of 300000 to 450000 was selected as the test set containing data that had not been previously examined by the DRL agents.

Each of the DRL agents was trained on the training set for four epochs in total.

The reward value comparison results are shown in Fig. 11. For the training set, compared to human operation, the AOE system I results in higher reward values, while the AOE system II results in very similar reward values, and the AOE system III yields a lower reward value. It can be observed that the reward value of the AOE systems on the training set are all within the open interval of (−1, 1). Referring to Eq. (33), penalization terms are added to discourage constraint violations. If the training process of the DRL agents is not convergent, the absolute reward value could be much higher than 1. Thus, from the performance of the AOE system on the training set, it can be inferred that the corresponding agents training processes are convergent. On the test set, the average reward values

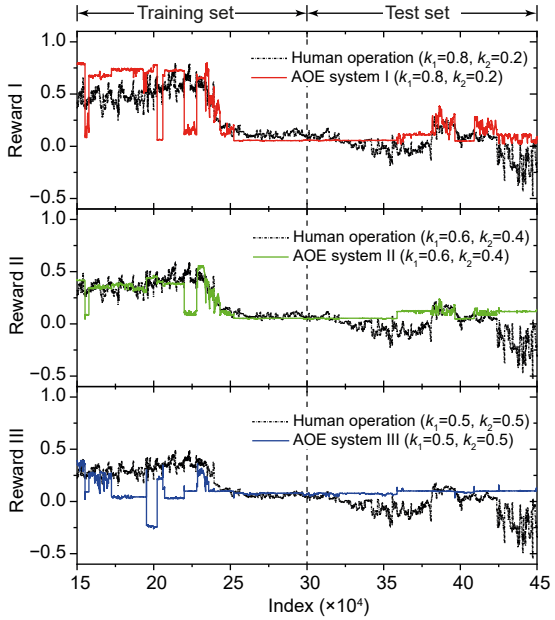

Fig. 11 Reward value comparison of various AOE systems and human operation

of these three AOE systems are all higher than that of the human operation, which means that the DRL agents have good generalization ability. The same phenomenon can be observed for the accumulative reward, as shown in Table 5. It turns out that the trade-off between $\bar{\dot{x}}$ and $\bar{E}$ of the AOE system II is very close to that of human operation, because it yields the closest average reward value to the human operation on the training set. This indicates that the proposed multi-objective comprehensive excavation performance measure is appropriate for shield excavation. On the other hand, it also turns out that when the human operators make decisions on the excavation operating parameters, the relative weight ratio of the excavation speed and the specific energy consumption is close to 6 to 4.

To investigate which AOE system performs better, a performance comparison of the AOE systems should be conducted. However, the reward value cannot be used for this purpose. Because these AOE systems use different reward functions, and different reward functions represent different performance evaluation criteria. It only makes sense to undertake a comparison under the same performance measure. Thus, the performance comparison of the AOE systems was performed from three different perspectives: the dimensionless advance speed $\bar{\dot{x}}$, specific energy consumption $\bar{E}$, and their combination $\bar{\dot{x}}/\bar{E}$, as shown in Fig. 12. It should be noted that $\bar{\dot{x}}/\bar{E}$ is a comprehensive performance measure similar to the reward function, but the effects of k_1 and k_2 are eliminated.

It is evident from Figs. 11 and 12 that both the reward value and advance speed sharply reduced after the index of 250000, which reflects the deterioration of geological conditions. Thus, in Fig. 12, the transverse axis indices between 150000 and 250000 are classified as normal grounds, whereas the indices after

Table 5 Comparison of performance for different datasets

Specification		Training set		Test set	
		Average reward	Accumulative reward	Average reward	Accumulative reward
k_1=0.8, k_2=0.2	Human operation	0.360	54001.950	0.022	3285.268
	AOE system I	0.397	59498.580	0.097	14597.046
k_1=0.6, k_2=0.4	Human operation	0.265	39766.781	−0.016	−2451.870
	AOE system II	0.236	35384.536	0.087	13046.381
k_1=0.5, k_2=0.5	Human operation	0.218	32649.199	−0.035	−5320.527
	AOE system III	0.094	14086.584	0.088	13228.602

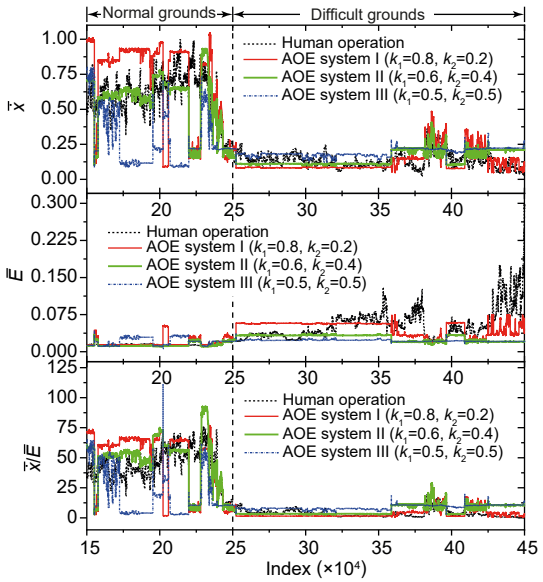

Fig. 12 Performance comparison of various AOE systems and human operation

250000 are classified as difficult grounds. According to the geological exploration report, the normal ground conditions are mainly silty clay and fine sand, with low bearing capacity, high compressibility, low water permeability, and normal engineering performance. In contrast, the difficult ground conditions are mainly silt and fine sand, with low compressibility, high water permeability, and poor engineering performance. Under difficult ground conditions, quicksand damage occurs easily, and the surrounding rock can collapse and deform. Fig. 12 shows that in normal grounds, the specific energy consumption levels of the AOE systems are similar, whereas the AOE system with a higher value of k_1 achieves a higher advance speed. In contrast, in difficult grounds, the advance speed levels of the three AOE systems are similar, whereas the AOE system with a higher value of k_2 shows lower specific energy consumption. Table 6 summarizes the average values of $\bar{\dot{x}}$, $\bar{E}$, and $\bar{\dot{x}}/\bar{E}$ for the human operation and the AOE systems. It is evident from Table 6 that under normal geological conditions, AOE system I has the highest average $\bar{\dot{x}}/\bar{E}$ value and therefore yields the best comprehensive performance. In comparison, under difficult geological conditions, AOE system III has the highest average $\bar{\dot{x}}/\bar{E}$ value, and yields the best comprehensive performance.

Compared to human operation, in normal grounds, the best performing AOE system increased the average excavation speed by 17.33% and the average $\bar{\dot{x}}/\bar{E}$ value by 9.91%. In difficult grounds, the best performing AOE system increased the average excavation speed by 41.91% and the average $\bar{\dot{x}}/\bar{E}$ value by 129%.

To further investigate the decision-making difference between the AOE systems and human operation, the actual actions of the two were compared, as shown in Fig. 13. The AOE system II is considered because it gives the closest comprehensive performance relative to the human operation on the training set. The magnitude of change in the human operator's actual actions is much larger than that in the AOE system. To better observe these respective trends, the first two figures in Fig. 12 use double y coordinates. It turns out that there is a strong correspondence between the actual actions of the AOE system II and geological parameters, whereas human operation does not exhibit the same relationship. This is an important reason why AOE systems can outperform human operation. As stated in Section 3.1, the excavation loads T_r and F_r are complex functions of the machine-ground interaction states (n_c and $\dot{x}$) and the geological parameters (P_{gw}, c, and φ). These variables are provided to the DRL agents for effective utilization. Thus, the DRL agent can explicitly exploit the geological parameters while considering the long-term performance objective for decision-making. In contrast, as stated in Section 3.3, the actual actions of the human operation are mainly dependent on the current n_c and $\dot{x}$. In practice, the main operation strategy of the human operator is to maintain n_c and $\dot{x}$ within their respective allowable ranges. The geological condition information used by

Table 6 Comparison of performance for different geological conditions

Specification	Normal grounds			Difficult grounds		
	Average $\bar{\dot{x}}$	Average $\bar{E}$	Average $\bar{\dot{x}}/\bar{E}$	Average $\bar{\dot{x}}$	Average $\bar{E}$	Average $\bar{\dot{x}}/\bar{E}$
Human operation	0.606	0.014	44.299	0.136	0.054	3.813
AOE system I (k_1=0.8, k_2=0.2)	0.711	0.018	48.692	0.122	0.048	3.548
AOE system II (k_1=0.6, k_2=0.4)	0.554	0.014	46.763	0.150	0.028	6.244
AOE system III (k_1=0.5, k_2=0.5)	0.273	0.021	18.564	0.193	0.022	8.738

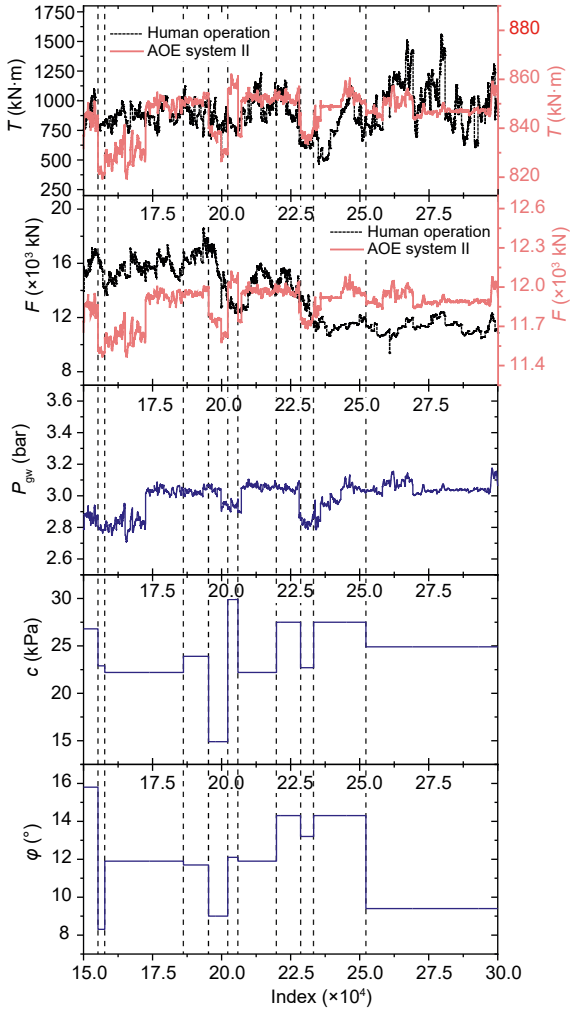

Fig. 13 Correspondence between the actual actions of the AOE system II and the geological parameters (1 bar=1×10^5 Pa)

the human operator (G_t and D_b) is not directly related to the cutterhead load and machine-ground interactions, and thus the correspondence between the actual actions in the human operation and the geological parameters is weak. Moreover, it is very difficult for the human operator to consider the long-term effects of a current action.

5.2 Discussion

Although the AOE system presented in this paper is developed using the construction field data of a selected shield machine, the approach established can be applied to all shield machines. Before its practical application, there are two key issues that need to be further studied.

The first key issue is how to validate the applicability of the AOE approach in real projects. Reliability and safety are the main factors restricting the development of TBM intelligent operation technology, and the AOE approach is no exception. Although the AOE approach has shown excellent performance and generalization ability in the simulated environment, it still has a long way to go before it can be applied to practical engineering. The AOE system needs to complete an excavation test on a real shield machine under the supervision of human operators. Nevertheless, we believe that the AOE approach creates new possibilities for the intelligent operation of shield machines. From the perspective of the development of a new generation of intelligent TBMs, the AOE approach has great potential.

Another key issue is how to generalize the capability of the AOE approach in different tunnel projects. Obviously, one method is to train a DRL agent on the construction field data extracted from different tunnel projects, however, this method is not only very time-consuming, but also has limited potential. A feasible method is to retrain the actor and critic networks of the DRL agent using transfer learning technology. In addition, fusing the similarity theory to improve the generalization ability of the AOE approach is also a method worth considering.

6 Conclusions

In this paper, a novel AOE approach that integrates DRL and optimal control is proposed for shield machines. A hybrid modeling method that integrates the first-principles analysis and DNN is proposed for the machine-ground interaction dynamics of the excavation process, which improves the interpretability of the model and simplifies the feature selection procedure. The mean MRE of the best performing DNN model is less than 8.5%, which validates its effectiveness. By analyzing the overall system, the multi-system coupling mechanism is revealed. The overall system analysis suggests that there are two equally important DOF for the AOE system design, namely the coordination level decision-maker and the execution level closed-loop controller.

The proposed dimensionless multi-objective comprehensive excavation performance measure that

combines the dimensionless advance speed and the specific energy consumption is found to be appropriate for the IOS. By comparison, it is found that the relative weight ratio of the excavation speed and the specific energy consumption is close to 6 to 4 when human operators make decisions on the excavation operating parameters.

The proposed AOE approach not only has the potential to replace human operation, but also can greatly improve the long-term comprehensive excavation performance. In addition, it is found that different decision-making strategies should be used for different geological conditions. To obtain superior comprehensive performance, the AOE system with a higher value of k_1 should be used in normal grounds, whereas the AOE system with a higher value of k_2 should be used in difficult grounds where the reward value and advance speed reduce significantly. Furthermore, there is a strong correspondence between the actual actions of the AOE system and geological parameters, whereas human operation does not exhibit the same relationship. An important reason why AOE systems can outperform human operation is that the AOE systems can make better use of the machine-ground interaction states and geological information while considering the long-term performance objective.

Although training a DRL agent is very time consuming relative to other ML algorithms, it still has a huge advantage over training a skilled operator. In this study, the DRL agent outperforms humans after conducting a total of 7 d of excavation training in the simulated environment. The training process takes about 5 h for each DRL agent. In contrast, training a skilled operator typically requires 15 to 18 months.

In addition, the DDPG agent can be continuously improved after deployment by using its online learning ability. In this study, the grounds with significantly reduced reward value and advance speed are classified as difficult grounds, and the remaining grounds are classified as normal grounds. This empirical and rough classification may pose some difficulties in further determining the switching conditions of different AOE systems, which is also the main limitation of this study. Future work will be focused on the improvement of the proposed approach and industrial testing.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (Nos. 2020YFF0218004 and 2020YFF0218003) and the National Natural Science Foundation of China (No. 52105074). The authors give special thanks to the China Railway Engineering Equipment Group Co., Ltd. for providing construction field data.

Author contributions

Ya-kun ZHANG and Guo-fang GONG designed the research and wrote the first draft of the manuscript. Yu-xi CHEN conducted the literature review. Geng-lin CHEN helped to organize the manuscript. Hua-yong YANG revised the final version and provided the funding support.

Conflict of interest

Ya-kun ZHANG, Guo-fang GONG, Hua-yong YANG, Yu-xi CHEN, and Geng-lin CHEN declare that they have no conflict of interest.

References

Antsaklis PJ, Rahnama A, 2018. Control and machine intelligence for system autonomy. *Journal of Intelligent & Robotic Systems*, 91(1):23-34. https://doi.org/10.1007/s10846-018-0832-6

Antsaklis PJ, Passino KM, Wang SJ, 1991. An introduction to autonomous control systems. *IEEE Control Systems Magazine*, 11(4):5-13. https://doi.org/10.1109/37.88585

Ates U, Bilgin N, Copur H, 2014. Estimating torque, thrust and other design parameters of different type TBMs with some criticism to TBMs used in Turkish tunneling projects. *Tunnelling and Underground Space Technology*, 40:46-63. https://doi.org/10.1016/j.tust.2013.09.004

Busoniu L, Babuska R, de Schutter B, et al., 2017. Reinforcement Learning and Dynamic Programming Using Function Approximators. CRC Press, Boca Raton, USA, p.1-13. https://doi.org/10.1201/9781439821091

Carreras M, Yuh J, Batlle J, et al., 2005. A behavior-based scheme using reinforcement learning for autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, 30(2):416-427. https://doi.org/10.1109/JOE.2004.835805

Chen RP, Zhang P, Kang X, et al., 2019. Prediction of maximum surface settlement caused by earth pressure balance (EPB) shield tunneling with ANN methods. *Soils and Foundations*, 59(2):284-295. https://doi.org/10.1016/j.sandf.2018.11.005

Cobbe K, Klimov O, Hesse C, et al., 2019. Quantifying generalization in reinforcement learning. Proceedings of the 36th International Conference on Machine Learning, p.1282-1289.

Dietterich TG, 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, 13:227-303. https://doi.org/10.1613/jair.639

El Sallab A, Abdou M, Perot E, et al., 2017. Deep reinforcement learning framework for autonomous driving. *Electronic*

Imaging, 2017(19):70-76. https://doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-023

Geng Q, Wei ZY, He F, et al., 2015. Comparison of the mechanical performance between two-stage and flat-face cutter head for the rock tunnel boring machine (TBM). *Journal of Mechanical Science and Technology*, 29(5): 2047-2058. https://doi.org/10.1007/s12206-015-0425-2

Han MD, Cai ZX, Qu CY, et al., 2017. Dynamic numerical simulation of cutterhead loads in TBM tunnelling. *Tunnelling and Underground Space Technology*, 70:286-298. https://doi.org/10.1016/j.tust.2017.08.028

He KM, Zhang XY, Ren SQ, et al., 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. 2015 IEEE International Conference on Computer Vision (ICCV), p.1026-1034. https://doi.org/10.1109/ICCV.2015.123

Huo JZ, Sun W, Chen J, et al., 2010. Optimal disc cutters plane layout design of the full-face rock tunnel boring machine (TBM) based on a multi-objective genetic algorithm. *Journal of Mechanical Science and Technology*, 24(2):521-528. https://doi.org/10.1007/s12206-009-1220-8

Kingma DP, Ba J, 2015. Adam: a method for stochastic optimization. The 3rd International Conference on Learning Representations.

Koopialipoor M, Nikouei SS, Marto A, et al., 2019. Predicting tunnel boring machine performance through a new model based on the group method of data handling. *Bulletin of Engineering Geology and the Environment*, 78(5):3799-3813. https://doi.org/10.1007/s10064-018-1349-8

Kuwahara H, Harada M, 1988. Application of fuzzy reasoning to the control of shield tunnelling. *Journal of the Society of Instrument and Control Engineers*, 27(11):1030-1037. https://doi.org/10.11499/sicejl1962.27.1030

Lillicrap TP, Hunt JJ, Pritzel A, et al., 2016. Continuous control with deep reinforcement learning. The 4th International Conference on Learning Representations.

Liu XY, Shao C, Ma HF, et al., 2011. Optimal earth pressure balance control for shield tunneling based on LS-SVM and PSO. *Automation in Construction*, 20(4):321-327. https://doi.org/10.1016/j.autcon.2010.11.002

Mahdevari S, Shahriar K, Yagiz S, et al., 2014. A support vector regression model for predicting tunnel boring machine penetration rates. *International Journal of Rock Mechanics and Mining Sciences*, 72:214-229. https://doi.org/10.1016/j.ijrmms.2014.09.012

Namli M, Bilgin N, 2017. A model to predict daily advance rates of EPB-TBMs in a complex geology in Istanbul. *Tunnelling and Underground Space Technology*, 62:43-52. https://doi.org/10.1016/j.tust.2016.11.008

Ng AY, Coates A, Diel M, et al., 2006. Autonomous inverted helicopter flight via reinforcement learning. *In*: Ang MH, Khatib O (Eds.), Experimental Robotics IX. Springer, Berlin, Heidelberg, Germany, p.363-372. https://doi.org/10.1007/11552246_35

Ninić J, Meschke G, 2015. Model update and real-time steering of tunnel boring machines using simulation-based meta models. *Tunnelling and Underground Space Technology*, 45:138-152. https://doi.org/10.1016/j.tust.2014.09.013

Pan XL, You YR, Wang ZY, et al., 2017. Virtual to real reinforcement learning for autonomous driving. British Machine Vision Conference.

Qin CJ, Shi G, Tao JF, et al., 2021. Precise cutterhead torque prediction for shield tunneling machines using a novel hybrid deep neural network. *Mechanical Systems and Signal Processing*, 151:107386. https://doi.org/10.1016/j.ymssp.2020.107386

Salimi A, Faradonbeh RS, Monjezi M, et al., 2018. TBM performance estimation using a classification and regression tree (CART) technique. *Bulletin of Engineering Geology and the Environment*, 77(1):429-440. https://doi.org/10.1007/s10064-016-0969-0

Saridis GN, 2001. Hierarchically Intelligent Machines. World Scientific, Hong Kong, China, p.25-32. https://doi.org/10.1142/4846

Shalev-Shwartz S, Shammah S, Shashua A, 2016. Safe, multi-agent, reinforcement learning for autonomous driving. https://arxiv.org/abs/1610.03295v1

Shao C, Lan DS, 2014. Optimal control of an earth pressure balance shield with tunnel face stability. *Automation in Construction*, 46:22-29. https://doi.org/10.1016/j.autcon.2014.07.005

Shi H, Yang HY, Gong GF, et al., 2011. Determination of the cutterhead torque for EPB shield tunneling machine. *Automation in Construction*, 20(8):1087-1095. https://doi.org/10.1016/j.autcon.2011.04.010

Song X, Liu JQ, Guo W, 2010. A cutter head torque forecast model based on multivariate nonlinear regression for EPB shield tunneling. International Conference on Artificial Intelligence and Computational Intelligence, p.104-108. https://doi.org/10.1109/AICI.2010.261

Sun W, Huo JZ, Chen J, et al., 2011. Disc cutters' layout design of the full-face rock tunnel boring machine (TBM) using a cooperative coevolutionary algorithm. *Journal of Mechanical Science and Technology*, 25(2):415. https://doi.org/10.1007/s12206-010-1225-3

Sun W, Shi ML, Zhang C, et al., 2018a. Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data. *Automation in Construction*, 92:23-34. https://doi.org/10.1016/j.autcon.2018.03.030

Sun W, Wang XB, Shi ML, et al., 2018b. Multidisciplinary design optimization of hard rock tunnel boring machine using collaborative optimization. *Advances in Mechanical Engineering*, 10(1):1-12. https://doi.org/10.1177/1687814018754726

Wang LT, Gong GF, Shi H, et al., 2012. A new calculation model of cutterhead torque and investigation of its influencing factors. *Science China Technological Sciences*, 55(6):1581-1588. https://doi.org/10.1007/s11431-012-4749-1

Wang LT, Sun W, Long YY, et al., 2018a. Reliability-based

performance optimization of tunnel boring machine considering geological uncertainties. *IEEE Access*, 6: 19086-19098. https://doi.org/10.1109/ACCESS.2018.2821190

Wang LT, Yang X, Gong GF, et al., 2018b. Pose and trajectory control of shield tunneling machine in complicated stratum. *Automation in Construction*, 93:192-199. https://doi.org/10.1016/j.autcon.2018.05.020

Xie HB, Duan XM, Yang HY, et al., 2012. Automatic trajectory tracking control of shield tunneling machine under complex stratum working condition. *Tunnelling and Underground Space Technology*, 32:87-97. https://doi.org/10.1016/j.tust.2012.06.002

Yeh IC, 1997. Application of neural networks to automatic soil pressure balance control for shield tunneling. *Automation in Construction*, 5(5):421-426. https://doi.org/10.1016/S0926-5805(96)00165-3

Yu A, Palefsky-Smith R, Bedi R, 2016. Deep Reinforcement Learning for Simulated Autonomous Vehicle Control. Technical Report, Stanford University, California, USA.

Zhang P, Chen RP, Wu HN, 2019. Real-time analysis and regulation of EPB shield steering using Random Forest. *Automation in Construction*, 106:102860. https://doi.org/10.1016/j.autcon.2019.102860

Zhang P, Wu HN, Chen RP, et al., 2020a. A critical evaluation of machine learning and deep learning in shield-ground interaction prediction. *Tunnelling and Underground Space Technology*, 106:103593. https://doi.org/10.1016/j.tust.2020.103593

Zhang P, Li H, Ha QP, et al., 2020b. Reinforcement learning based optimizer for improvement of predicting tunneling-induced ground responses. *Advanced Engineering Informatics*, 45:101097. https://doi.org/10.1016/j.aei.2020.101097

Zhang Q, Kang YL, Qu CY, et al., 2010. Mechanical model for operational loads prediction on shield cutter head during excavation. IEEE/ASME International Conference on Advanced Intelligent Mechatronics, p.1252-1256. https://doi.org/10.1109/AIM.2010.5695778

Zhang Q, Huang T, Huang GY, et al., 2013. Theoretical model for loads prediction on shield tunneling machine with consideration of soil-rock interbedded ground. *Science China Technological Sciences*, 56(9):2259-2267. https://doi.org/10.1007/s11431-013-5302-6

Zhang Q, Qu CY, Cai ZX, et al., 2014. Modeling of the thrust and torque acting on shield machines during tunneling. *Automation in Construction*, 40:60-67. https://doi.org/10.1016/j.autcon.2013.12.008

Zhang Q, Hou ZD, Huang GY, et al., 2015. Mechanical characterization of the load distribution on the cutterhead–ground interface of shield tunneling machines. *Tunnelling and Underground Space Technology*, 47:106-113. https://doi.org/10.1016/j.tust.2014.12.009

Zhang WJ, Yang GS, Lin YZ, et al., 2018. On definition of deep learning. World Automation Congress (WAC), p. 1-5. https://doi.org/10.23919/WAC.2018.8430387

Zhang YK, Gong GF, Yang HY, et al., 2019. Data-driven direct automatic tuning scheme for fixed-structure digital controllers of hybrid systems. *IET Control Theory & Applications*, 13(2):248-257. https://doi.org/10.1049/iet-cta.2018.5165

Zhang YK, Gong GF, Yang HY, et al., 2020. Precision versus intelligence: autonomous supporting pressure balance control for slurry shield tunnel boring machines. *Automation in Construction*, 114:103173. https://doi.org/10.1016/j.autcon.2020.103173

Zhou C, Ding LY, He R, 2013. PSO-based Elman neural network model for predictive control of air chamber pressure in slurry shield tunneling under Yangtze River. *Automation in Construction*, 36:208-217. https://doi.org/10.1016/j.autcon.2013.03.001

Zhou C, Ding LY, Skibniewski MJ, et al., 2018. Data based complex network modeling and analysis of shield tunneling performance in metro construction. *Advanced Engineering Informatics*, 38:168-186. https://doi.org/10.1016/j.aei.2018.06.011

Zhou C, Xu HC, Ding LY, et al., 2019a. Dynamic prediction for attitude and position in shield tunneling: a deep learning method. *Automation in Construction*, 105:102840. https://doi.org/10.1016/j.autcon.2019.102840

Zhou C, Ding LY, Zhou Y, et al., 2019b. Hybrid support vector machine optimization model for prediction of energy consumption of cutter head drives in shield tunneling. *Journal of Computing in Civil Engineering*, 33(3): 04019019. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000833

Zhou J, Zhou YH, Wang BC, et al., 2019. Human–cyber–physical systems (HCPSs) in the context of new-generation intelligent manufacturing. *Engineering*, 5(4):624-636. https://doi.org/10.1016/j.eng.2019.07.015

Electronic Supplementary Materials

The pseudo-code for the implementation of the training environment.