Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)

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

Ya-kun ZHANG, Guo-fang GONG, Hua-yong YANG, Yu-xi CHEN, and Geng-lin CHEN

Cite this as: Ya-kun ZHANG, Guo-fang GONG, Hua-yong YANG, Yu-xi CHEN, Geng-lin CHEN, 2022. Towards autonomous and optimal excavation of shield machine: a deep reinforcement learning-based approach. *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)*, 23(6):458-478.

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

Introduction

Motivations:

- Intelligent TBM is considered as an inexorable development trend.
- Dynamic optimization of a long-term excavation performance is central to intelligent TBM operation.

Long-term research goal:

- Autonomous and optimal excavation of TBMs

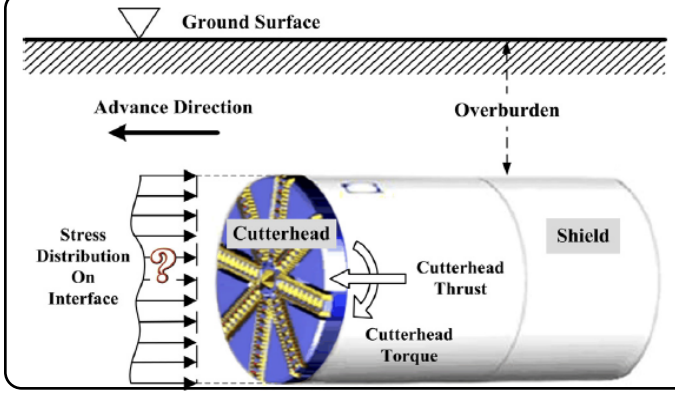
Technical challenges:

- A high-accuracy modeling method of the machine-ground interaction dynamics has not yet been established.
- There is still a lack of a comprehensive excavation performance measure suitable for the IOS.
- It is not appropriate to apply the existing dynamic optimization methods directly to shield machines

System Analysis and Problem Formulation

Modeling of the Machine-ground Interaction Dynamics

- Developed based on the existing static models.
- The input and output variables of the process were fully determined.



$$T - T_r = \frac{2\pi}{60} J\dot{n}_c$$

$$F - F_r = m\ddot{x}$$

$$T_r = f\left(n_c, \dot{x}, F, P_{gw}, c, \varphi, t\right)$$

$$F_r = g\left(n_c, \dot{x}, P_{gw}, c, t\right)$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T = \begin{bmatrix} n_c & \dot{x} \end{bmatrix}^T$$

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{u_m} \mid \boldsymbol{u_d} \end{bmatrix}^T = \begin{bmatrix} u_1 & u_2 \mid u_3 & u_4 & u_5 \end{bmatrix}^T = \begin{bmatrix} T & F \mid P_{gw} & c & \varphi \end{bmatrix}^T$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^T = \begin{bmatrix} n_c & \dot{x} \end{bmatrix}^T = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$$

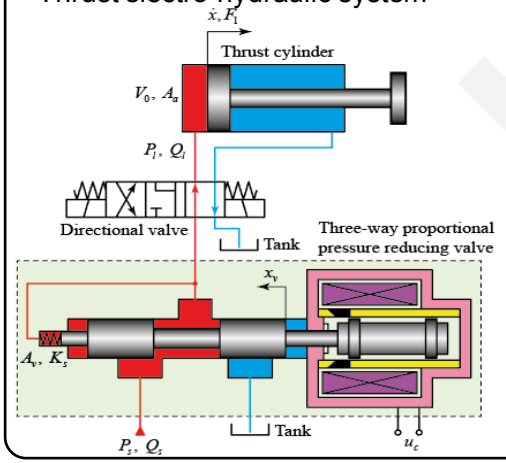
$$\dot{\boldsymbol{x}} = \varphi\left(\boldsymbol{x}, \boldsymbol{u}, t\right)$$

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

Modeling of the actuators

- Developed using first principles.

Thrust electro-hydraulic system



$$Q_l = k_q x_v - k_c P_l$$

$$u_c = L\dot{I} + R_e I + K_e \dot{x}_v$$

$$m_v \ddot{x}_v = F_m - K_s x_v - B_v \dot{x}_v - P_l A_v - K_f x_v$$

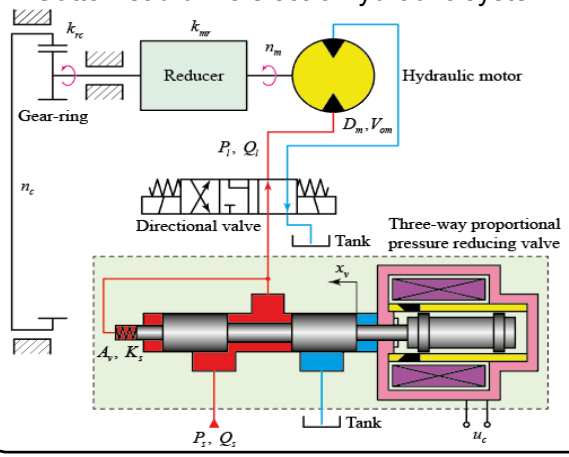
$$F_m = K_I I - K_x x_v$$

$$Q_l = A_a \dot{x} + \frac{V_0}{\beta_e} \dot{P}_l + C_t P_l$$

$$F_l = P_l A_a = \frac{F}{N_c}$$

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

Cutterhead drive electro-hydraulic system



$$Q_l = k_q x_v - k_c P_l$$

$$u_c = L\dot{I} + R_e I + K_e \dot{x}_v$$

$$m_v \ddot{x}_v = F_m - K_s x_v - B_v \dot{x}_v - P_l A_v - K_f x_v$$

$$F_m = K_I I - K_x x_v$$

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

$$n_m = k_{mr} k_{rc} n_c$$

$$T_l = \eta k_{mr} k_{rc} T_m = \frac{T}{N_m} = D_m P_{pl}$$

$$T_l(s) = H_1 u_c(s) - H_2 n_c(s)$$

System Analysis and Problem Formulation

Multi-system coupling mechanism

Markov model for excavation process

$(O_b, A, h, \gamma, \rho)$

$O_b = \{x, G_t, D_b\}$

$A = \{\tilde{T}, \tilde{F}\}$

$h : O_b \times A \mapsto O_b$

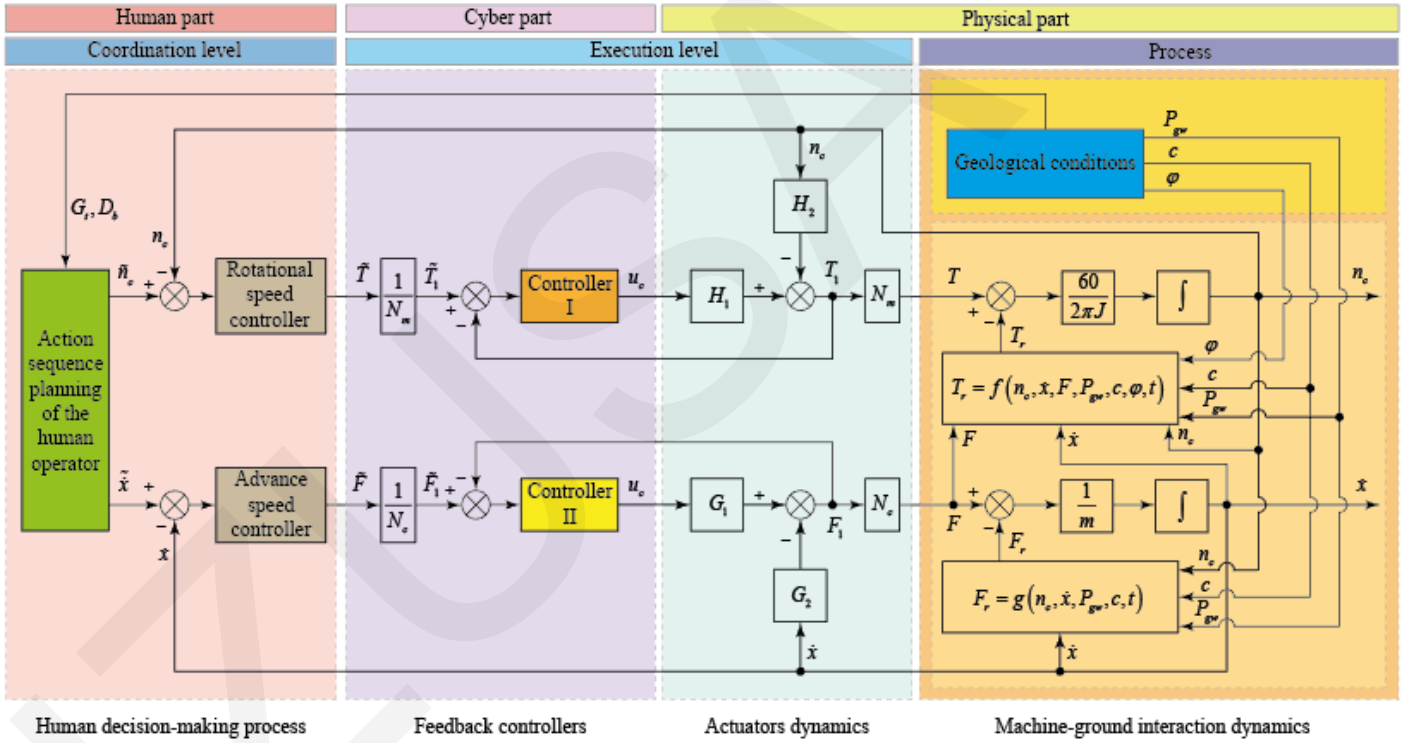
$\gamma \in [0,1]$

$\rho : O_b \times A \mapsto \mathbb{R}$

Long-term performance measure

$$V^h(o_{b,0}) = \sum_{k=0}^{N} \gamma^k r_{k+1}$$

$$= \sum_{k=0}^{N} \gamma^k \rho\left(o_{b,k}, h\left(o_{b,k}\right)\right)$$



The goal of the optimal excavation

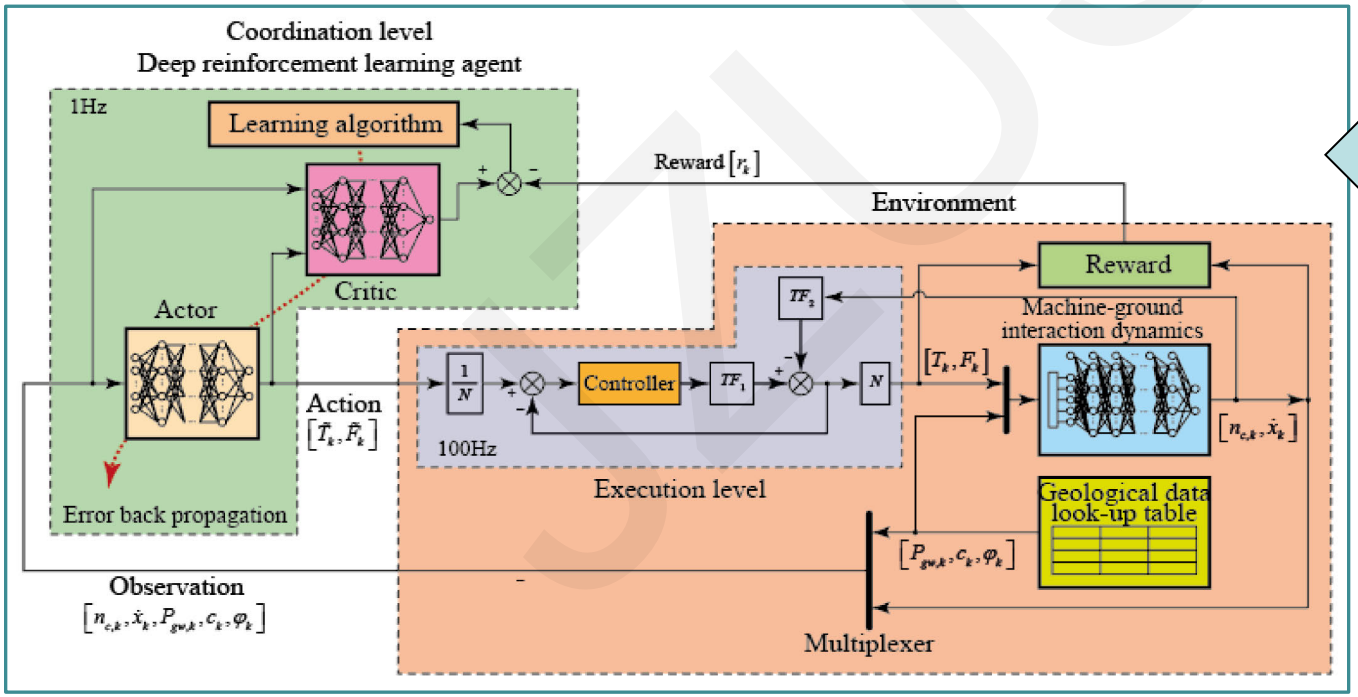
To choose the action values at each step and apply it to its environment, such that the accumulated reward value can be maximized.

$$\pi^*(o_b) = \arg\max_{a \in A}\left(\sum_{k=0}^{N} \gamma^k r_{k+1}\right)$$

The 2 degrees of freedom for the overall system design

Autonomous Optimal Excavation Scheme

- The coordination level was implemented as a deep reinforcement learning agent.
- The execution level digital optimal controllers was designed.
- The machine-ground interaction dynamics was represented as a deep neural network model.



Dimensionless comprehensive excavation performance measure

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

$$\bar{E}_k = \frac{4}{\pi D^2}\left(\frac{\bar{\omega}_k \bar{T}_k}{L(\dot{x}_k)} + \bar{F}_k\right)$$

$$L(x) = \begin{cases} x & \text{if } x \geq 10^{-5} \\ 10^{-5} & \text{if } x < 10^{-5} \end{cases}$$

Definition of the reward function

$$r_k = J_k + P_{\text{sgn},k} + P_{\min,k} + P_{\max,k}$$

$$P_{\text{sgn},k} = N(\bar{T}_k) + N(\bar{F}_k) + 10N(\bar{n}_{c,k})$$

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

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

$$N(x) = \begin{cases} x & \text{if } x < 0 \\ 0 & \text{if } x \geq 0 \end{cases}$$

$$L(x, x_{\min}) = \begin{cases} x - x_{\min} & \text{if } x < x_{\min} \\ 0 & \text{if } x \geq x_{\min} \end{cases}$$

Performance Evaluation and Discussion

Numerical experiment setup

- The allowed action values for the DRL agent were strictly constrained in the range obtained by the human operators for the same segment of field data
- Three DRL agents with different k_1 and k_2 values were trained on a dataset consisting of selected representative 150,000 samples.

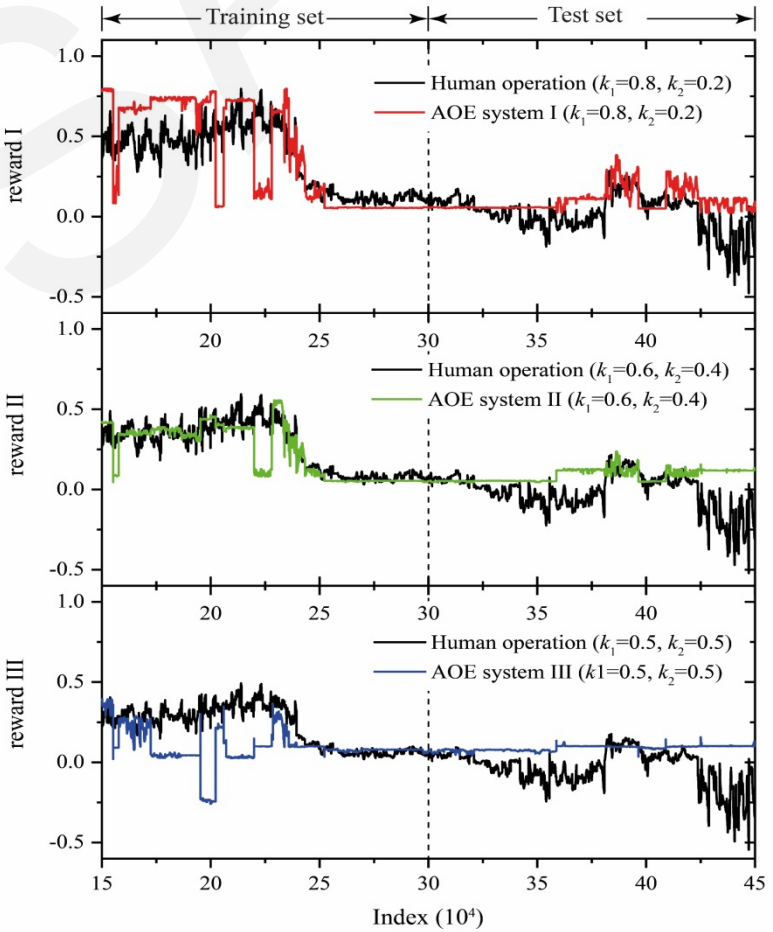
$$r_k = J_k + P_{\mathrm{sgn},k} + P_{\min,k} + P_{\max,k} \qquad J_k = k_1 \bar{\dot{x}}_k - k_2 \bar{E}_k$$

AOE system I: k_1 = 0.8, k_2 = 0.2

AOE system II: k_1 = 0.6, k_2 = 0.4

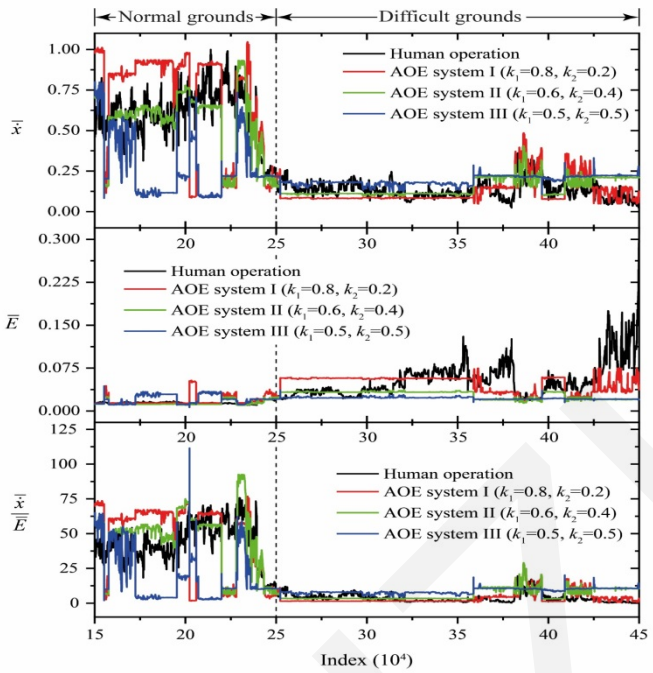
AOE system III: k_1 = 0.5, k_2 = 0.5

AOE system V.S. Human operator



Performance Evaluation and Discussion

Comparison of AOE systems



- In normal grounds, 17.33 % increase in average excavation speed and 9.91% increase in average $\dot{\bar{x}}/\bar{E}$.
- In difficult grounds, 41.91 % increase in average excavation speed and 129 % increase in average $\dot{\bar{x}}/\bar{E}$.

Decision characteristics