Electronic Supplementary Materials

For 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*

✉E-mail: gfgong@zju.edu.cn

The pseudo-code for the implementation of the training environment

Input: Action $\{\tilde{T}_k, \tilde{F}_k\}$

Output: Observation $\{n_{c,k}, \dot{x}_k, P_{gw,k}, c_k, \varphi_k\}$, reward $\{r_k\}$, Done {True or False}

- **Initialization**:
  1. Load the geological data look-up table;
  2. Load the machine-ground interaction DNN model;
  3. Step number $k \leftarrow 0$;
  4. Output the initial observation to the DRL agent;
- **function** step $(\tilde{T}_k, \tilde{F}_k)$:

1. $T_k \leftarrow \tilde{T}_k$, $F_k \leftarrow \tilde{F}_k$;
2. **if** k<the number of training steps:
   step number $k \leftarrow k+1$;
   else:
   $k \leftarrow 0$;
   end if
3. Read geological data $\{P_{gw,k}, c_k, \varphi_k\}$;
4. Scale the actual action $\{T_k, F_k\}$ and geological data $\{P_{gw,k}, c_k, \varphi_k\}$ by dividing by their corresponding maximum values;
5. Calculate $\{n_{c,k}, \dot{x}_k\}$ using the machine-ground interaction DNN model;
6. Scale $\{n_{c,k}, \dot{x}_k\}$ by dividing by their corresponding maximum values;
7. Calculate the reward r_k value using Eq. (31);
8. Done ← Is_done $(T_k, F_k, n_{c,k}, \dot{x}_k)$;

return observation, reward, Done

end function