Honor of Kings Arena: an Environment for Generalization in Competitive Reinforcement Learning

Hua Wei*, Jingxiao Chen*, Xiayi Ji†, Hongyang Qin, Minwen Deng, Siqin Li, Liang Wang, Weinian Zheng, Yong Yu, Lin Liu, Lanxiao Huang, Deheng Ye, Qiang Fu, Wei Yang

huanwei@tencent.com, licomm@tencent.com, yugu@tencent.com, yugu@apex.sjtu.edu.cn,
jingxiao.chen@tencent.com, xiyij@tencent.com, yangmin6150@tencent.com

Hua Wei, Jingxiao Chen, Xiayi Ji, Hongyang Qin, Minwen Deng, Siqin Li, Liang Wang, Weinian Zheng, Yong Yu, Lin Liu, Lanxiao Huang, Deheng Ye, Qiang Fu, Wei Yang

huanwei@tencent.com, licomm@tencent.com, yugu@tencent.com, yugu@apex.sjtu.edu.cn, jingxiao.chen@tencent.com, xiyij@tencent.com, yangmin6150@tencent.com

Honor of Kings Arena (HoK): Provided Resources
- HoK: OpenAI Gym-like, authorized game environments
- Baseline models, including behavior-tree (BT) and RL models
- Replay tool

Mechanics from MOBA
- MOBA games have different roles/heroes, and each role has different actions

Challenges for AI
- Multi-agent: Good AI model need to coordinate well between different players
- Generalization: Good AI model needs to perform stably well in controlling the actions of different heroes against different opponent heroes
- Multi-player: MOBA games usually involves in two or more parties, each party consists of one or more players

In the main screen, there are four parts:
- a mini-map (A),
- a dashboard that records the number of KDA(s) (kill/death/assist) (B)
- a movement controller (C.1)
- skill controller buttons (C.2)

Honor of Kings: an Appealing Environment for AI
- Popularity
- Existing research interest

10+ related papers in top AI venues NeurIPS, ICML, AAI, UAI, TNNLS, ...

Generalization across opponents
- Opponent hero: DianChao (B)
- Opponent hero: GanJiangMuye (A)

Generalization across targets
- Target hero: DianChao (B)
- Target hero: GanJiangMuye (A)

Benchmarking Results
- Figure 4: Different evaluation metrics on the Honor of Kings Arena for DQN and PPO w.r.t. the number of training episodes. Error bars represent standard deviation. PPO performs better than DQN.
- Figure 5: Win rate of a well-trained model from task "DianChao (RL) vs. DianChao (BT)" transferred to tasks "DianChao (RL) vs. different opponent heroes (BT)". The agent trained to control DianChao against DianChao controlled by BT, and tested to control DianChao against different heroes controlled by BT. Red: Directly transferring the model to control DianChao and compete with different opponent heroes. Green: Multi-task training on five tasks "DianChao (RL) vs. DianChao[Bartholomew, Anzi,GanJiangMuye,ZhangFei,Bu] (BT)" and testing the model on twenty tasks. Blue: Using the model trained on five tasks "DianChao (RL) vs. DianChao[Bartholomew, Anzi,GanJiangMuye,ZhangFei,Bu] (BT)" and testing the model on twenty tasks. The policy trained on DianChao could not generalize to all tasks with different opponent heroes. Blue rectangles highlights the five tasks used in multi-task and distillation. The error bars indicate the standard deviation under five seeds,