MahjongMaster: A General Mahjong AI System

Kuaishou Technology Inc.

Team Introduction

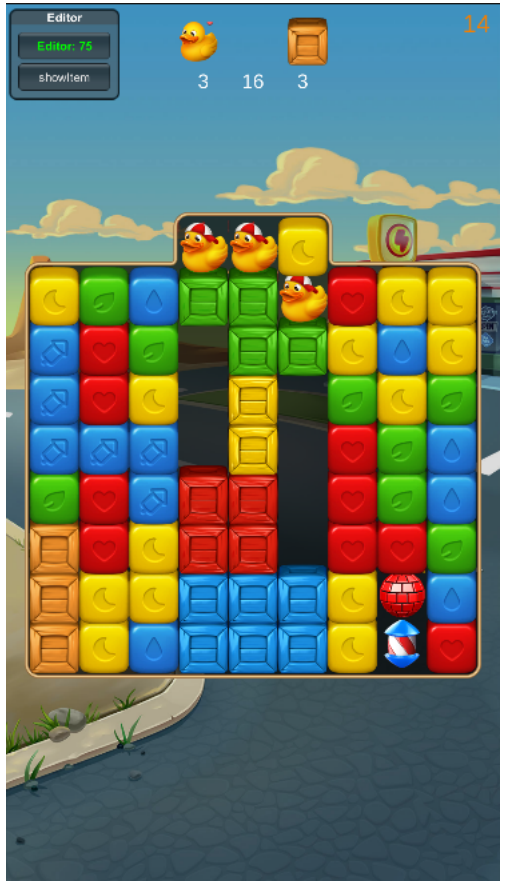
AI Platform & Game Business Department

Empower Games With AI

Throughout all stages

- Game development: automatic test & level generation
- Game promotion: intelligent advertising system
- Game operation strategies: data-driven gaming matching, novice protection, etc.
- Game AI agent: GameAI Platform -> Mahjong, FightTheLandlord, etc.

Team Introduction



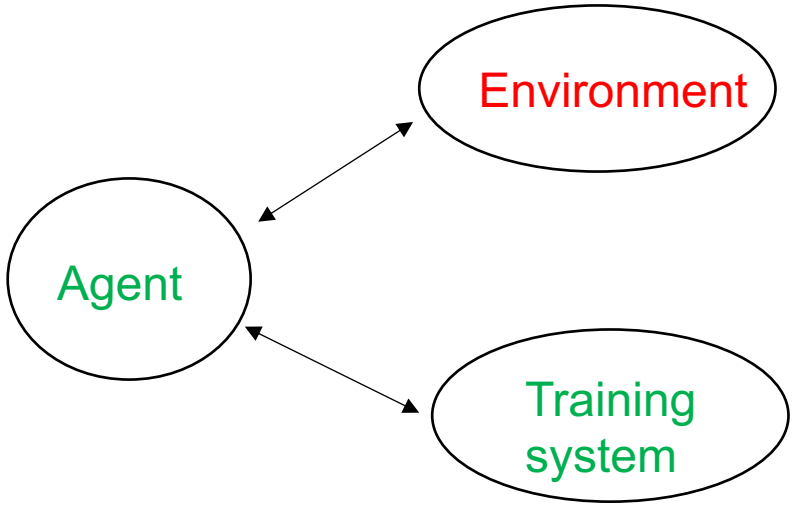
Outline

- Overview of MahjongMaster
  - General Mahjong AI System
  - One-step Decision Making Method
  - Feature & Model Structure
- Training the MahjongMaster
  - Distributed Deep Reinforcement Learning (RL) Framework
  - RL Model Initialization
  - Other Training Techniques

Overview: General Mahjong AI System

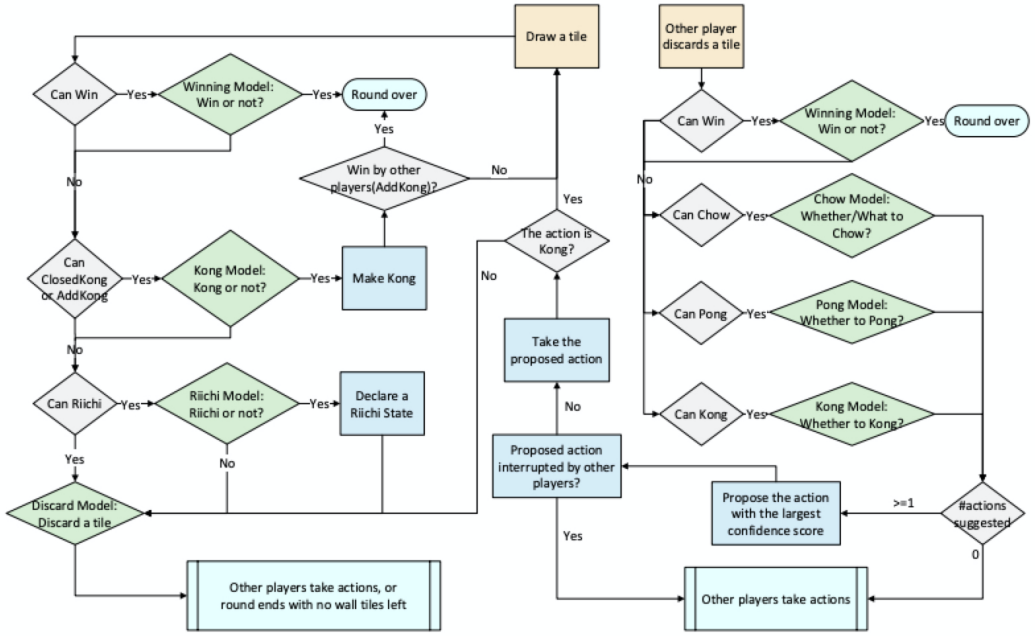
General Mahjong AI System

- Generalized algorithm for different mahjong rules
  - same training and inference framework
  - same agent design
  - only environment (simulator) is rule-specified
- Applied in online Mahjong game
  - serve 6+ mahjong rules with AI rating, surpass top human players
  - rank **1st** in Tournament Round, 3rd in final round, IJCAI2020 Mahjong AI Competition



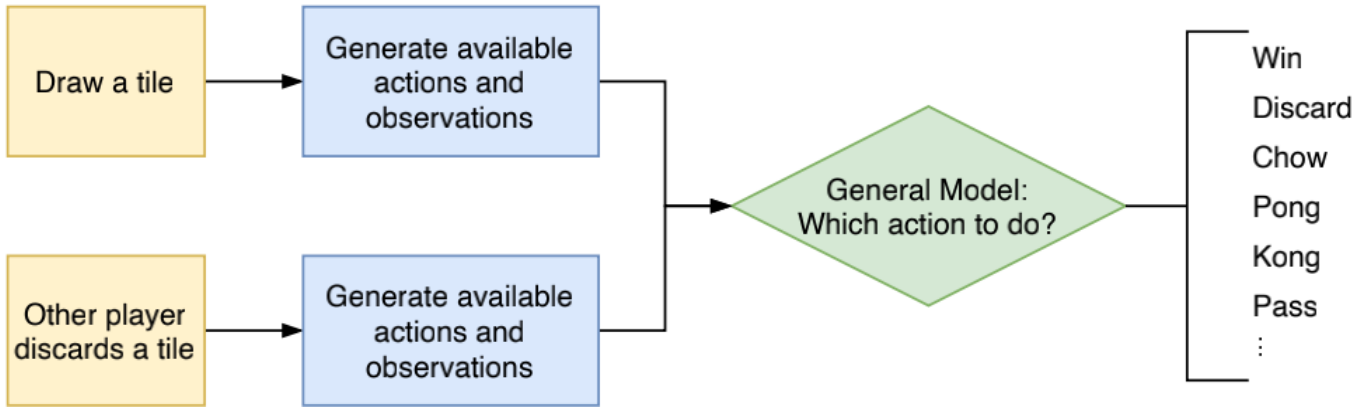
Overview: One-step Decision Making

Decision Flow (Suphx)



- Similar to human player
- Rule-specific / Models are specialized

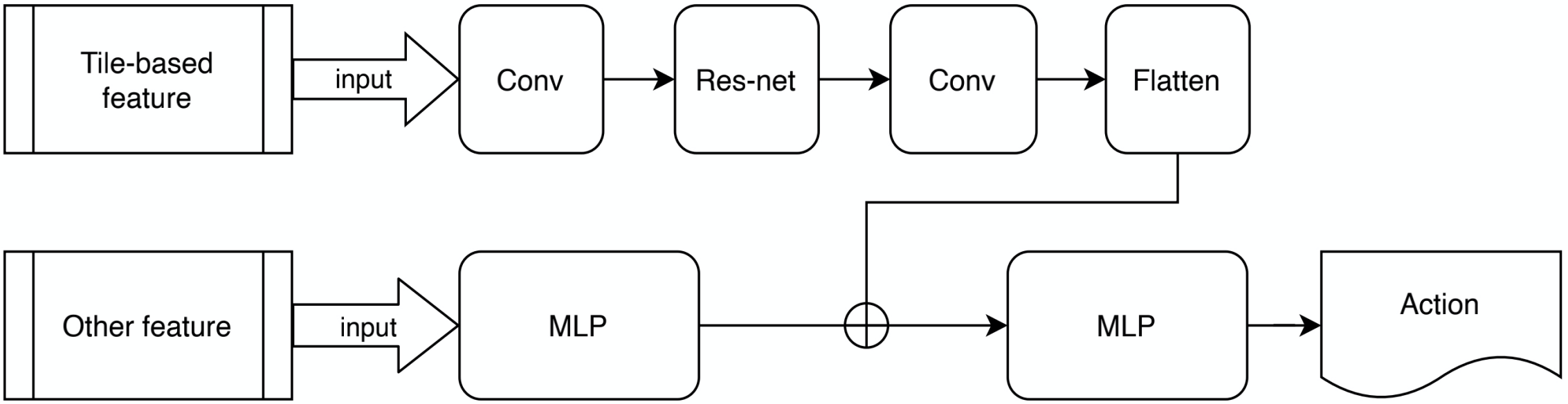
One-step (Ours)



- Suitable for RL training
- Easy to extend
- Low cost for training and inference

Li, Junjie, et al. "Suphx: Mastering Mahjong with Deep Reinforcement Learning." *arXiv preprint arXiv:2003.13590* (2020).

Overview: Feature & Model Structure



- Tile-based feature: handcards, tiles set, etc.
- Other features: dealer position, available actions, etc.

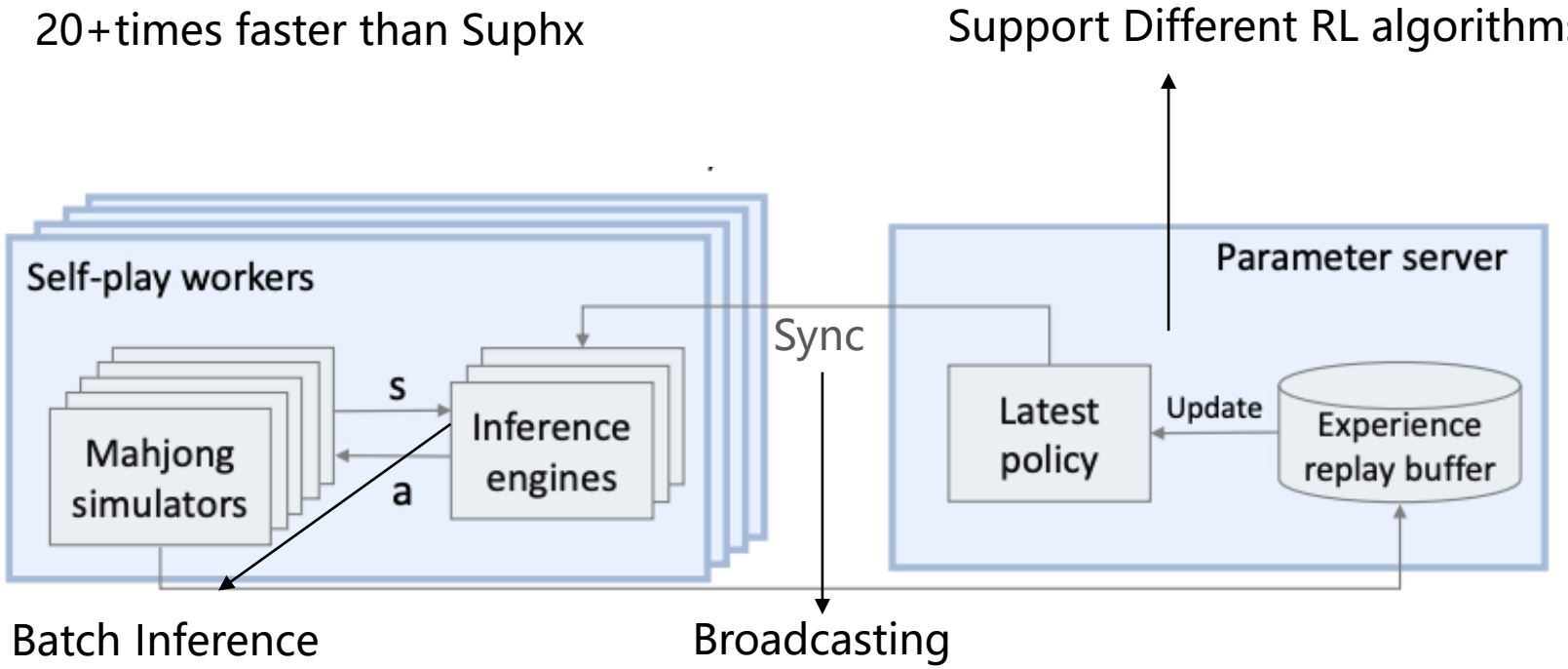
Same model fits well for all rule types, e.g. Chinese-standard, Sichuan, Two-player Mahjong, etc.

Outline

- Overview of MahjongMaster
  - General Mahjong AI System
  - One-step Decision Making Method
  - Feature & Model Structure
- Training the MahjongMaster
  - Distributed Deep Reinforcement Learning (RL) Framework
  - RL Model Initialization
  - Training Techniques

Training: Distributed Deep Reinforcement Learning

- Async training & simulation, deployed on multiple CPU & GPU Machines
- Optimize inference and sync efficiency
- 48 games/second per GPU card (2080ti)
- 20+times faster than Suphx



Li, Junjie, et al. "Suphx: Mastering Mahjong with Deep Reinforcement Learning." *arXiv preprint arXiv:2003.13590* (2020).

Training: RL Model Initialization

RL from random policy is hard

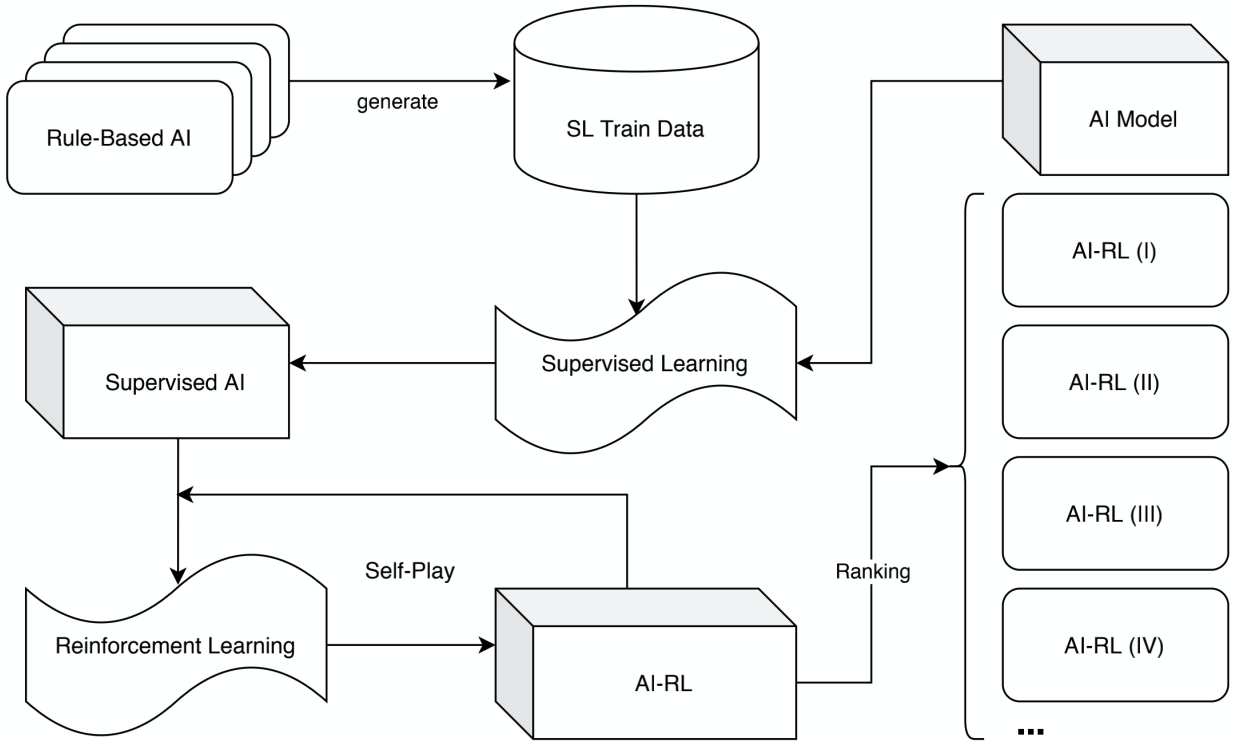
Good initialization is important

Common approach: learn from expert data

Lack of human data: learn from Rule-based AI

Both perform well

Choose suitable approach accordingly



RL based on Rule-base Initialization

Training : Other Techniques

- Reward design
  - Handcrafted reward to avoid sparse reward, useful for early training process
  - adapt the Duplicate Format and use the difference between a player's score as reward, reduce varience
- Entropy control:

$$\nabla_\theta J(\pi_\theta) = \mathop{\mathrm{E}}_{s,a\sim\pi_{\theta'}}\left[\frac{\pi_\theta(s,a)}{\pi_{\theta'}(s,a)}\nabla_\theta \log \pi_\theta(a|s)\, A^{\pi_\theta}(s,a)\right] + \alpha\nabla_\theta H(\pi_\theta)$$

  after certain time, gradually decay alpha to 0
- Oracle guiding: train an oracle model, then decay to normal model by distillation instead of masking

Li, Junjie, et al. "Suphx: Mastering Mahjong with Deep Reinforcement Learning." *arXiv preprint arXiv:2003.13590* (2020).

Thanks!

Contact us: chenzhihan@kuaishou.com