



UNIVERSITY OF
CAMBRIDGE

Towards a Competitive 3-Player Mahjong AI using Deep Reinforcement Learning

Xiangyu Zhao, Sean B. Holden
xz398@cantab.ac.uk, sbh11@cl.cam.ac.uk

24 August 2022

Department of Computer Science and Technology

3-player Mahjong (Sanma): a more aggressive game style

	4-Player Mahjong	3-Player Mahjong (Sanma)
Win rate	21.9%	29.7%
Discard-loss rate	11.8%	13.6%
Riichi rate	17.3%	24.9%
Average Han	3.19	4.60

Source: Tenhou.net, Houou table

<http://tenhou.net/ranking.html> <http://tenhou.net/sc/prof.html>



Features

- Features divided into two categories:
 - *Tile features* (sets or sequences of tiles) can be encoded as one-hot vectors
 - *Numerical features* (e.g. player scores) can be binary-encoded into multiple columns
- In addition to the triplet/quad tiles and Riichi status, also include the turn number of the Pon/Kan/Riichi calls
- Leave spaces for 1m–9m tiles and the fourth player, for transferability to 4-player Mahjong
- Use a 34×366 array to represent a state
 - 34 kinds of tiles, 366 columns for 22 features



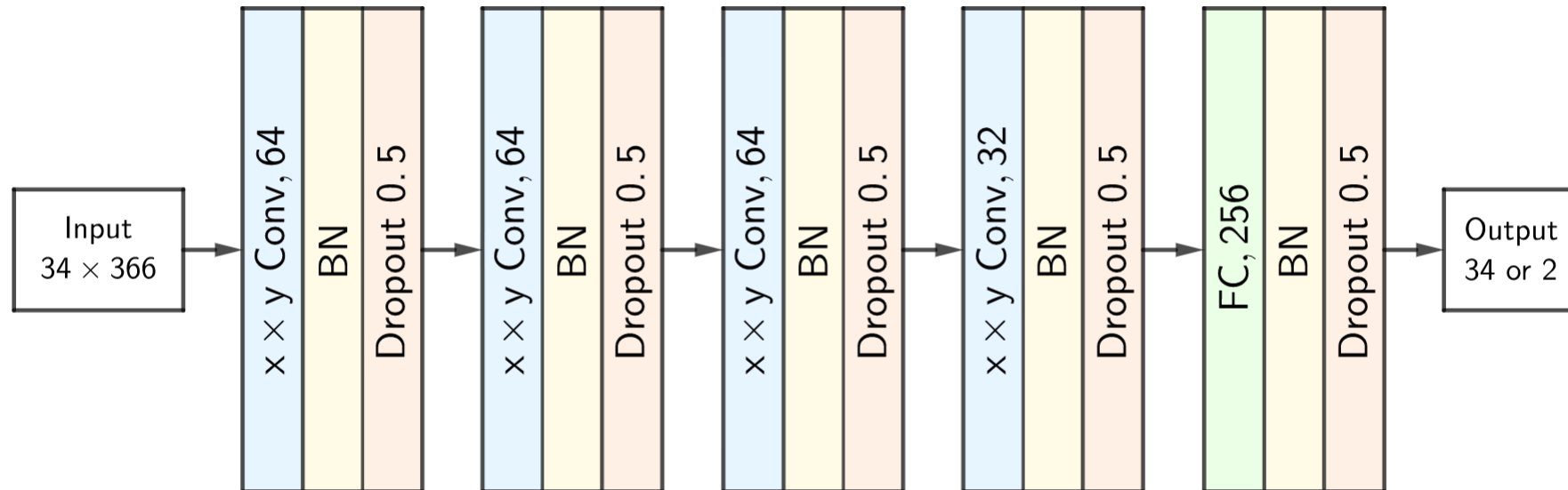
Data

- Use game records from the Houou table (top 0.1% of the ranked players) on Tenhou.net
- Training dataset: 50,000 rounds of Sanma game in 2019
Test dataset: 5,000 rounds of Sanma game in 2020
- 1,033,317 examples for discard
185,052 examples for Pon
41,861 examples for Kan
167,636 examples for Kita
133,949 examples for Riichi



Models

- 4-layer CNN structure
- Convolutional filter sizes tuned for each action
- Enhance major action (discard)'s model through self-play RL, using REINFORCE



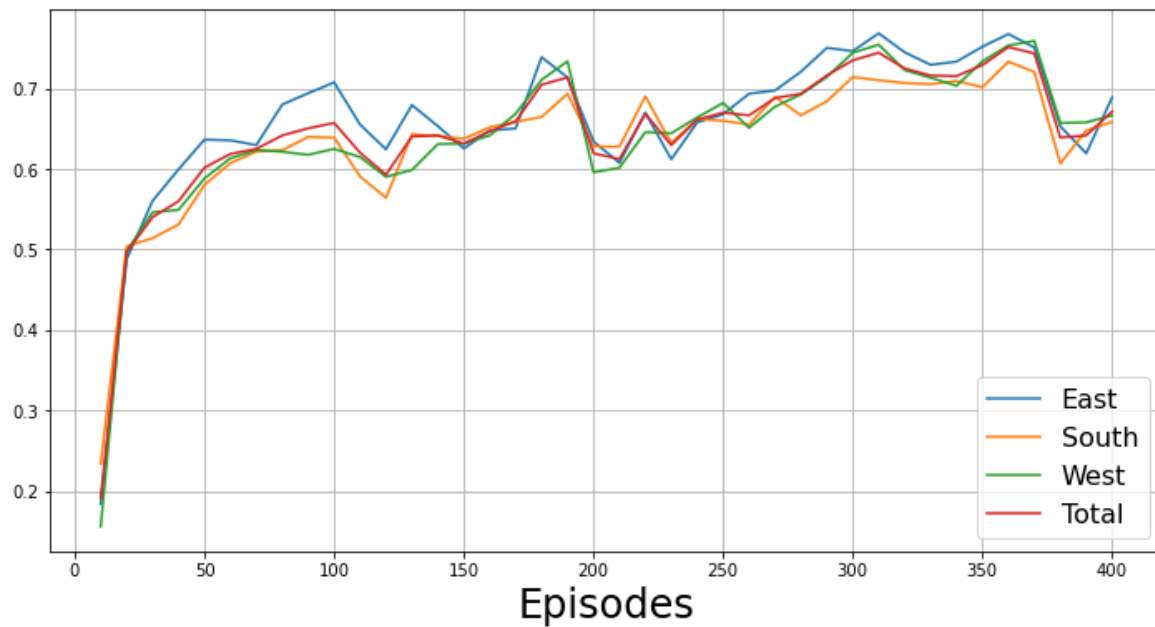
Results: supervised learning

Model	Meowjong	Gao et al. (2018)	Suphx
Discard	65.81%	68.8%	76.7%
Pon	70.95%	88.2%	91.9%
Kan	92.45%	—	94.0%
Kita	94.26%	—	—
Riichi	62.63%	—	85.7%

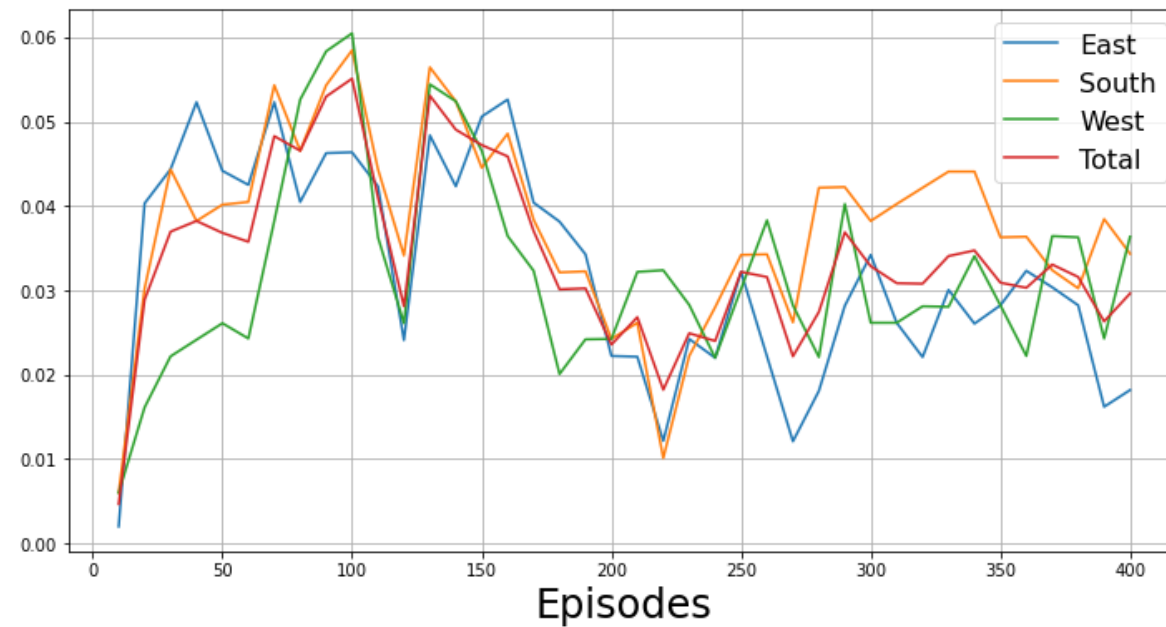


Results: reinforcement learning

Win Rates



Loss Rates



Results: reinforcement learning

Agents (vs. Baseline)	Wind	1 st Place Rate	2 nd Place Rate	3 rd Place Rate	Draw Rate
Baseline	—	0.02%	0.02%	0.02%	99.94%
SL	East	22.00%	0.06%	0.08%	77.86%
	South	22.68%	0.06%	0.02%	77.24%
	West	20.72%	0.16%	0.04%	79.08%
	Total	21.80%	0.09%	0.05%	78.06%
RL	East	73.59%	0.02%	3.27%	23.12%
	South	71.93%	0.08%	3.46%	24.53%
	West	71.61%	0.06%	2.85%	25.48%
	Total	72.38%	0.05%	3.19%	24.38%

Results: reinforcement learning

