

Team ALONG

Zhongyue Huang, Minggao Wei, Guan Yang
NetEase Games AI Lab

Content

- IJCAI 2020 Mahjong AI Competition
- Features
- Network Structure
- Imitation Learning
- Reinforcement Learning
- Experiments

IJCAI 2020 Mahjong AI Competition

- Chinese Standard Mahjong
- Swiss tournament
- duplicate format
- ranking scores
- 136 tiles, not 8 flowers
- draw tiles from one sub-wall

Features

- base features
 - private hand, concealed Kong, open melds, discarded tiles, and so on
 - unseen tiles and all hand
- look-ahead features
 - distance to each Fan
 - nearest tile pattern for each Fan
- available-action features
 - discard, Chow, Pong, and Kong
 - current tile discarded by others
 - tiles in private hand that are related to Chow

Base Features

The image shows a Mahjong game interface with four players: North, South, East, and West. The dealer is North, and the prevailing wind is East. The game is in progress, with a hand of 15 tiles left for North and 13 tiles left for West. The interface includes annotations for various features:

- dealer's wind**: North
- prevailing wind**: East
- open melds**: Three melds are visible in North's hand: 3-4-5 Ten Thousand, 3-4-5 Thousand, and 3-4-5 Hundred.
- private hand**: The hand of the player whose tiles are being annotated (North).
- discarded tiles**: A central area showing tiles discarded by other players.
- the number of tiles left in each wall**: The number of tiles remaining in each player's wall.

Player	Wind	Tiles Left
North	Dealer	15
South		15
East	Prevailing	12
West		13

Look-ahead Features



- SanSeSanBuGao
- distance 1
- nearest tile pattern



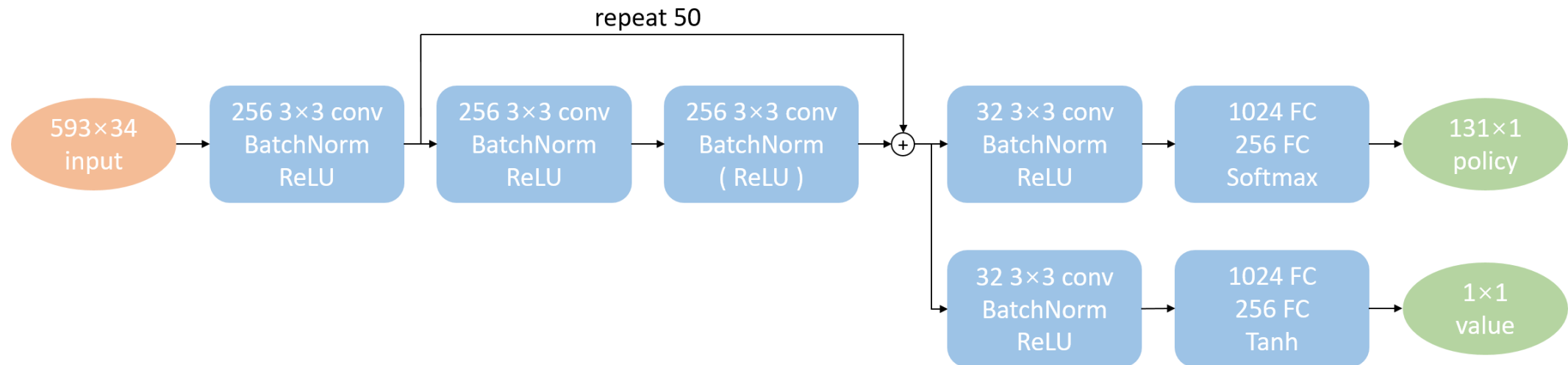
- ZuHeLong
- distance 3
- nearest tile pattern



Network Structure

- one model for all actions
- 50 ResNet blocks
- output policy and value simultaneously

Action	Discard	Chow	Pong	Kong	Pass
Dim	34	45	1	34	1



Imitation Learning

- behaviour clone
- dataset
 - human data provided by competition organizer
 - about half a million games, 32 million steps
 - filter low-quality data
- training
 - learning rate 0.001, ReduceLROnPlateau
 - batch size 2048

Reinforcement Learning

- PPO+GAE
- reward
 - score
 - Ting
 - Huang
- training
 - start with 100 decks
 - double when the winning rate exceeds 80%

Experiments

- all dataset: filter low-quality data based on human data.
- winner dataset: extract data of winners based on all dataset.
- sl_all model: train on all dataset.
- sl_winner model: train on winner dataset.
- rl_random model: train on random decks based on sl_all model.
- rl_double model: train on double decks based on rl_random model.

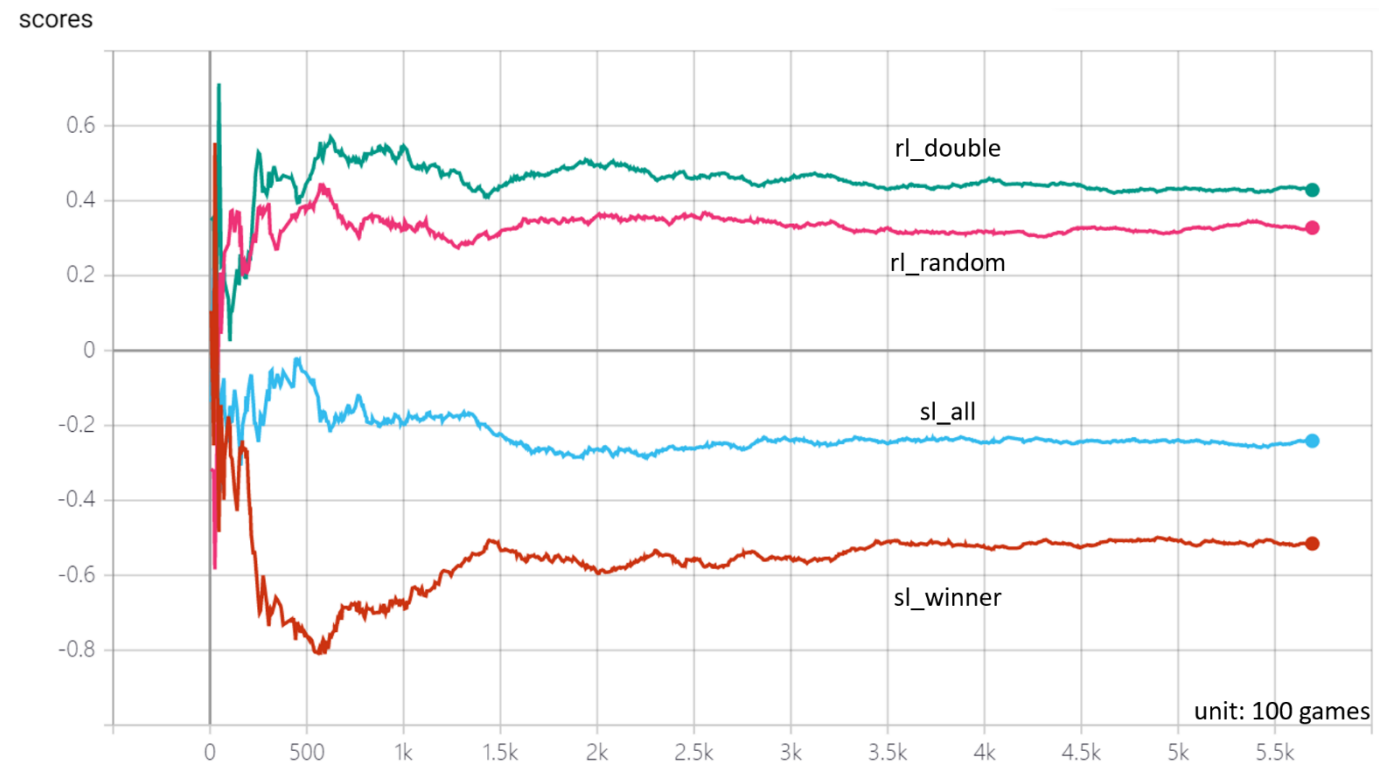
Experiments

- imitation learning

Dataset	Batch Size	Acc	Winner Acc	Score
All	128	71.96	-	-0.31
	512	73.21	-	+0.13
	2048	74.00	-	+0.25
Winner	128	70.90	74.58	-
	512	71.33	76.82	-
	2048	72.75	78.68	-

Experiments

- reinforcement learning



Thanks

Q & A