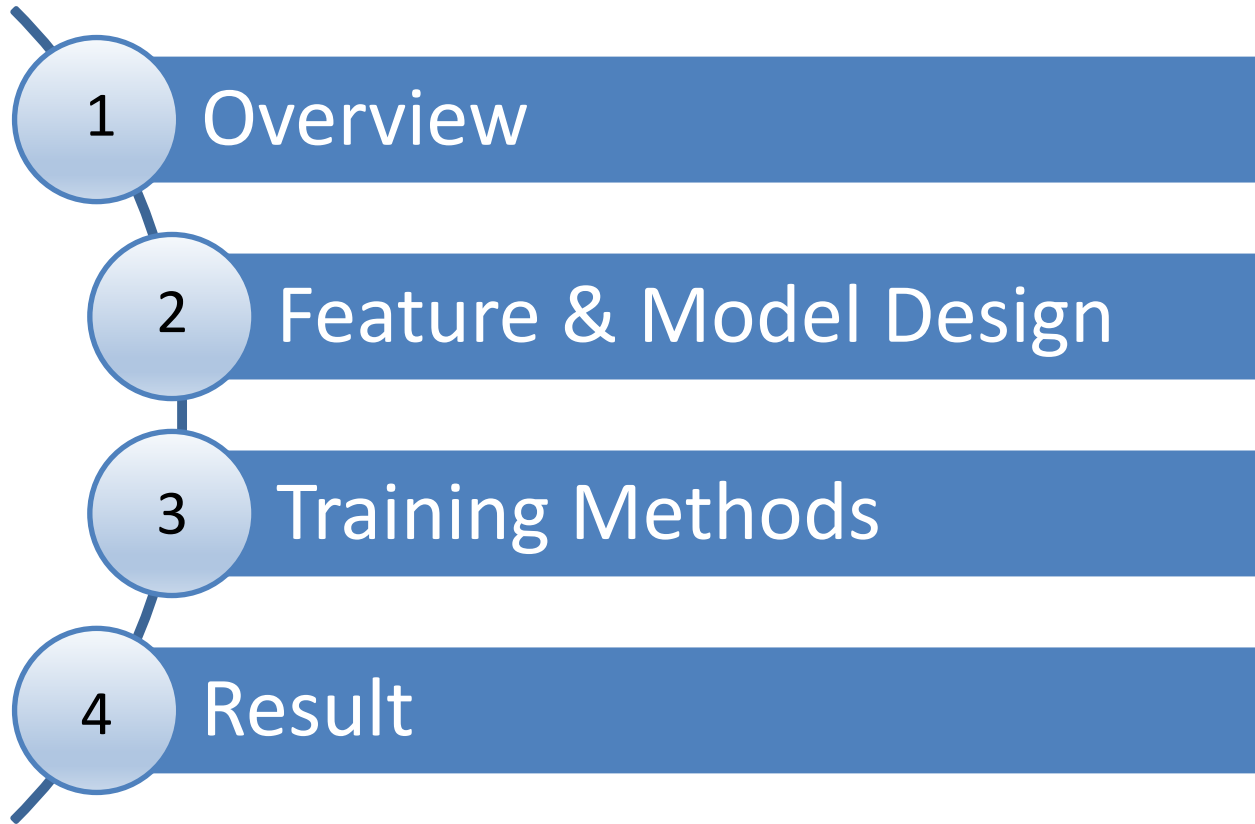


SuperJong

Yata
2021.01

Outline



Overview



Reinforcement
Learning



Deep CNN
Global value



Self-play
Zero start



Asynchronous
training



Feature & Model Design

Image like features

Visible information

Self hands

Self chow

Self pong, kong

Self Drops

Others chow

Others pong, kong

Drops

Invisible information

other hands

Remain deck

Reflect table

	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1	T1	T2	T3	T4	T5	T6	T7	T8	T9
W1	W1	W2	W3	W4	W5	W6	W7	W8	W9
B1	B1	B2	B3	B4	B5	B6	B7	B8	B9
F1	F1	F2	F3	F4	J1	J2	J3	—	—

Hand tiles



Image-like feature

↓

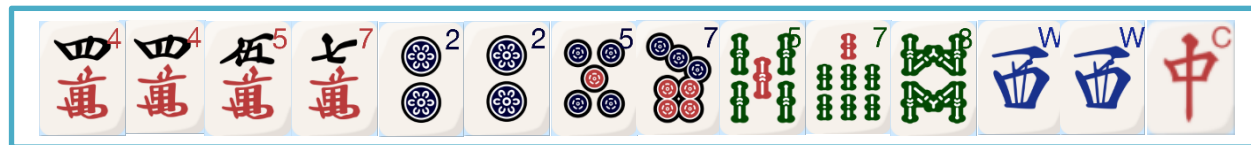
	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	0	0
	0	1	1	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0
1	0	1	0	1	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0

Feature & Model Design

Shanhu features

- Shanhu : the shortest distance from current hand to winning hand
- Need tiles: the needed tiles to change current hand to the winning hand
- No-need tiles: the tiles we don't need in the winning hand

Current hand



Need tiles



No-need tiles



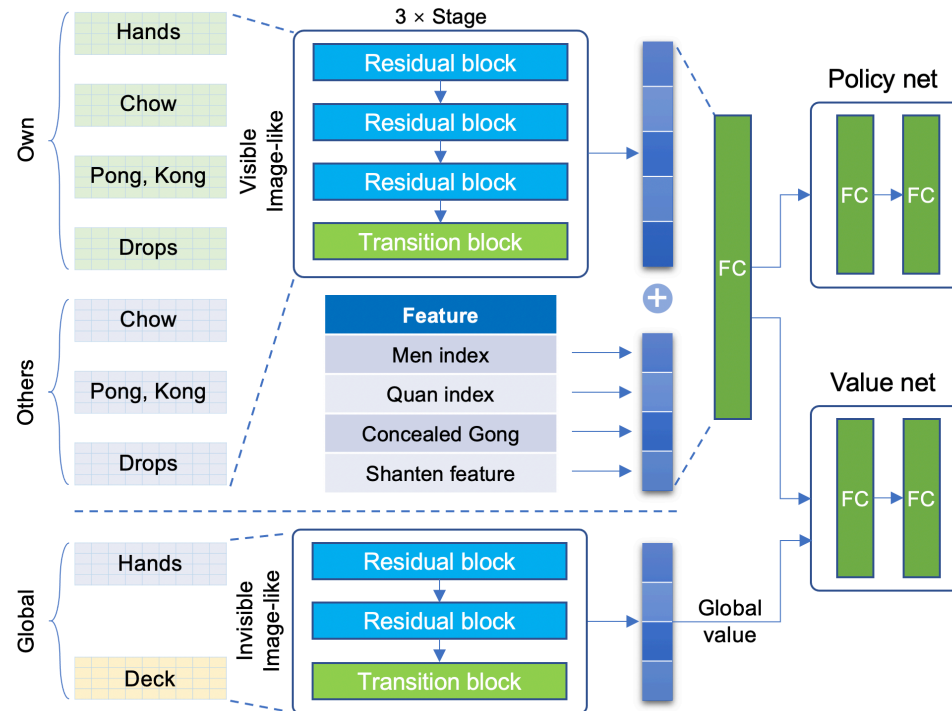
Winning hand



Feature & Model Design

Model Design

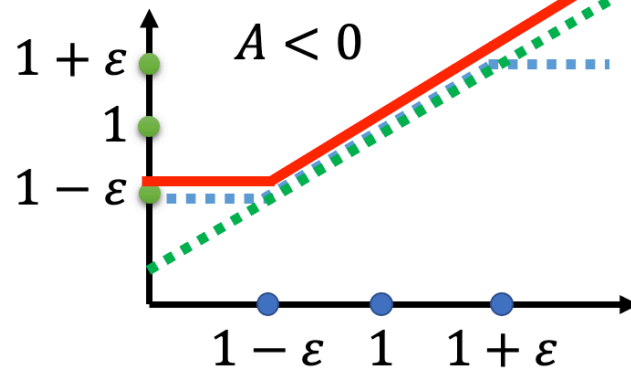
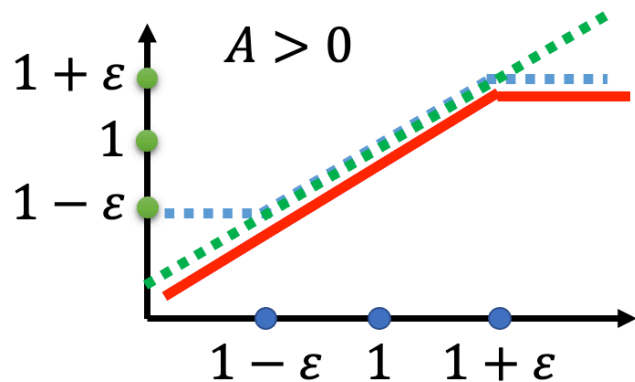
- architecture: a **single** end-to-end neural network
- Input: all relevant information
- Output: probabilities of all actions, the state value



Training Methods

PPO

$$J_{PPO2}^{\theta^k}(\theta) \approx \sum_{(s_t, a_t)} \min \left(\frac{p_{\theta}(a_t|s_t)}{p_{\theta^k}(a_t|s_t)} A^{\theta^k}(s_t, a_t), \text{clip} \left(\frac{p_{\theta}(a_t|s_t)}{p_{\theta^k}(a_t|s_t)}, 1 - \epsilon, 1 + \epsilon \right) A^{\theta^k}(s_t, a_t) \right)$$



Self play from zero

Training Methods

Reward Norm

- make training process stable
- prevent AI too greedy

Input: R

Output: R_norm

function getNormReward

if dianpao:

 R_norm = -2

elif lose:

 R_norm = -1

else:

$R_norm = \sqrt{\frac{R}{2}}$

Result

- rank1 in final round

Results of Final Round - Stage One

Top 16 in Stage One of Final Round are listed here.

Rank	Team	Account	Score
1	SuperJong	yata	1052.01
2	ALONG	yue	991.00
3	清澄高校	gameai_platform	979.00
4	地锅鸡	luyd_cpp	979.00
5	Luma Pools	TheWitness	978.00
6	国土无双	kcznol	973.00
7	雀圣	metaphysics	963.99
8	你能卡掉我当场把屏幕吃掉	infinityedge	962.99
9	重在参与	工行卡十六号嘴	961.00
10	点个大的	humanfy	950.00
11	King of gambler	komqaq	937.99
12	DX	zasfs	936.99
13	Test	huluBrother	935.99
14	自所小分队吧	啊咪咪小熊	928.99
15	just for fun	emmmm	924.99
16	我好菜啊啊	woaixuexi	904.99

Results of Final Round - Stage Two

Top 4 in Stage Two of Final Round are listed here.

Rank	Team	Account	Score
1	SuperJong	yata	1338.00
2	ALONG	yue	1314.00
3	清澄高校	gameai_platform	1281.00
4	地锅鸡	luyd_cpp	1186.98

Thanks

Yata
2021.01