Training Japanese Mahjong Agent with Two Dimension Feature Representation

Long Honghai[1,a)]   Tomoyuki Kaneko[2,b)]

Abstract: Over recent years, researchers have made many great achievements in the area of Artificial Intelligence. In perfect information games, there are AI agents which have already exceeded the level of top human players. In Japanese Mahjong, an imperfect information game, an agent called Suphx has reached human-level strength. In Suphx, convolutional neural networks have been used and the input of the network is calculated by hand-crafted features part of which are not unveiled yet. The neural networks used in deciding discards and other possible actions such as Chow, Pong and Kong, are large and the time cost in training agent is high. In this paper, we present an alternative set of features that can be obtained by a state observation and is powerful enough to achieve a comparable accuracy through a neural network with a smaller number of layers. We use the data from human records and train the agent by supervised learning. The Riichi network has reached an accuracy of 86.1% and the Kong network has reached an accuracy of 98.1%. The accuracy of discard network has reached 72.3% which is high enough to play effectively in self-play.

Keywords: Supervised Learning, Reinforcement Learning, Imperfect Information Game

1. Introduction

Studies in game AIs are more and more popular in recent years. Strong AIs have been trained in many perfect information games, such as Go [1]. In general, imperfect information games are more complex than perfect information games. Mahjong is an imperfect information game which has its own difficulties and challenges different from those in perfect information games. The huge hidden information and long game lengths make it challenging to train strong AI agents in this game. The score rules and the complicated winning condition also increase the difficulty of the game.

2. Japanese Mahjong

There are usually four players in one game of Japanese Mahjong. The whole game usually includes 4 or 8 rounds. The score of each player is set to 25,000 points at the beginning of the first round. In general, players aim at building a winning hand to get a winning score and increase his or her points. The changes of scores depend on the hand of the winning player of the round. The final goal of the game is to get the highest score among all players at the end of the last round.

2.1 Basic Rules of Japanese Mahjong

There are 136 tiles including 34 different types of tiles and 4 same tiles of each type. At the beginning of each round, all 136 tiles are rearranged into four walls and each player draws 13 tiles from the wall. A round starts from the dealer. A round continues until one or more players declare win, or ran out of walls (resulted in draw). At the end of a round, scores of all players will change according to the result of winning players' hand tiles and how the acquired the last tile in the winning hand. A whole Mahjong game usually includes 4 or 8 rounds but will end earlier if any player's score becomes negative, or continues more if the dealer wins a round. During a round, players take actions one by one in anti-clockwise. However, actions including Chow, Pong and Kong can interrupt and change the order of the players. Additionally, there are some special rules in Japanese Mahjong which are important for us when we design features of observation.

2.1.1 Yaku

Yaku is a condition that determines the value of the player's hand. A winning hand must have at least one yaku. Each yaku has its own *han* value (each han usually doubles the winning points). Details of some yakus will be introduced in 4.2.2 with the design of corresponding features.

2.1.2 Dora, Akadora and Uradora

Dora is a bonus tile that adds a han value to a hand. It may bring big winning points frequently. Players recognize a Dora by dora indicator which is visible to all players through out a round. The number of dora indicators and consequently doras increase when a player declares Kong.

Akadoras are three tiles, which are 5-man, 5-pin and 5-sou. Akadoras are usually colored red and each of them doubles the score as a normal dora does.

There is an uradora indicator under each dora indicator. Uradora has the same function as the normal dora tiles.

[1] Graduate School of Interdisciplinary Information Studies, The University of Tokyo
[2] Interfaculty Initiative in Information Studies, the University of Tokyo
[a)] longhonghai@g.ecc.u-tokyo.ac.jp
[b)] kaneko@graco.c.u-tokyo.ac.jp

Only when a player declared Riichi wins, uradoras become effective by revealing uradora indicators.

2.1.3 Meld

A complete winning hand consists of four melds and a pair. A meld in a player's hand is a group of three or four tiles satisfying one of conditions; a Chow (a sequence of three tiles), a Pong (triplets), or a Kong (quads), in a player's hand.

During the game, a player can form a meld by drawing from the wall or declaring from another player's discard. If a player declares a meld from a discarded tile, he needs to reveal the meld tiles including the tile just discarded on the table and then make his own discard.

A Chow is a set of three tiles of the same suit and consecutive numbers. A player can declare Chow only when the discarded tile is from his left player, who is prior in order.

A Pong is a set of three identical tiles. The tile declared for Pong can be discarded by any other player.

A Kong is a set of four identical tiles. After declaring a Kong, a new dora indicator will be opened and the player who declares Kong will draw one more tile. The discarded tile for declaring Kong can be also from any other player. There are three types of Kong in the game of Mahjong, including closed Kong, exposed Kong, exposed Kong from exposed Pong. Only exposed Kong and exposed Kong from exposed Pong are open melds.

2.1.4 Riichi

A player can declare Riichi if the player needs only one tile to complete a winning hand. The player must not have open melds. Riichi is a kind of yaku and a player can win on an other players' discard or a tile drawing from the wall after declaring Riichi. The player can not change his hand except when forming closed Kong. The Riichi status of players are also important information of the game.

2.1.5 Wind Tiles and Dragon Tiles

There are four types of wind tiles and three types of dragon tiles in each game.

Two type of wind tiles are very important for each player: round wind and player wind. Round wind is set at the beginning of a round and all four players have the same wind in a round. On the other hand, every player has its own player wind in a round. The player who first draw a tile from the wall has East as his player wind. The player wind of the next player is North, and so on. A player can use a triplet or a quad of round wind tiles or player wind tiles as a yaku.

For dragon tiles, a player can use a triplet or a quad of them as a yaku in any round.

2.1.6 Winning

A player can declare win after drawing a tile (the tile is also called 'winning tile') from the wall or an another player's discard. The points of the winning player obtains in this round depends on the winning hand. If a player wins from drawing a tile from the wall, all other players will pay the winner points corresponding to the winning hand, otherwise, if the player wins from another player's discarded tile, the player who discards the winning tile pays all the points.

2.2 Tenhou

Tenhou [2] is a popular online Japanese Majhong site mainly for human players. Game records of top human players are available from Tenhou. It is hard to train an agent of Japanese Mahjong without any imitation of human players because random agents can hardly win a game. In other words, at the beginning of training, supervised learning from human records is important. Agents can also play in Tenhou and the stable rank in Tenhou can be used to evaluate the strength of an agent.

2.3 Challenges of Training AI in Japanese Mahjong

The imperfect information of Japanese Mahjong is from invisibility of other players' private tiles and the tiles in the wall. There are 13 tiles in a player's hand and 70 tiles in the wall at the beginning of each round. When a round starts, a player who is decided by the rule of the game, draws a tile and adds it into his hand tiles. We define this player as the *dealer* of the round and also call this player "oya". Then the dealer has to choose a tile to discard from his hand tiles. The discarded tile is visible for all players. The next player will take the same sequence of actions (draws a tile from the wall and then discards a tile from hand tiles) as the dealer. When a player decides which tile he should discard, apart from his hand tiles, he also has to take the hidden tiles in other players' hand tiles and the wall into consideration. It is because the effectiveness of his tiles in his hands depends on those hidden tiles. Consequently, the agent should consider much more than its hand tiles when discarding a tile.

Furthermore, Japanese Mahjong has its own scoring rules which makes the game more difficult. As mentioned before, 25,000 points initially given to each player will change according to the result of each round. Usually, a player should take a strategy which leads to a win in each round in order to increase his score. However, in some conditions, losing a round can also be considered as a good tactic for the player. For example, the player who has the highest score choose to lose the last round intentionally if he can still keep the highest score among all players after calculating new points. The tactic like this makes learning of Japanese Mahjong more complex because the reward of a round can not be used to train the agent directly. Furthermore, the winning points of a round varies largely depending on the winning hand. The player at the rank 4 position with the lowest points usually aims at large winning points because he needs to get out from the last rank. When a player wants to win more points in a round, he needs a more difficult winning hand.

The actions that player can choose in the game of Mahjong is also complex. As is said above, in general, players draw a tile from the wall and then discard a tile from hand one by one. However, the order of the player may be interrupted by some types of actions including Chow, Pong and Kong.

It is hard to predict these interruptions during the game. Besides, a player can also declare Riichi when he needs only one tile to form a winning hand. A player can not change his hand after declaring Riichi. It means he must have to discard the tile he draws from the wall except when he declares win by the tile. In Mahjong, if a player, say A, forms a winning hands by getting a tile just discarded by another player, say B, then player A can declare win and receives points from player B. Therefore, declaring Riichi will increase the the risk of losing a round because he cannot choose a safer tile to prevent the other players from declaring win.

3. Related Work

In study [3], the proposed agent plays the game using three neural networks; discard, Chow and Pong, each of which is trained by supervised learning. In Suphx [4], five models: discard model, Chow model, Pong model, Kong model and Riichi model, have been trained for playing Mahjong. Suphx has shown stronger performance than most human players in terms of stable rank in the Tenhou platform.

The networks used in Suphx were trained by supervised learning at first. Then reinforcement learning has been added to improve the performance of the discard network. The features used for input of the networks are encoded from the information from the agent's observation. The encoding process is complex. In the feature encoding process, depth first search is used and the details are not clearly described in the paper. The neural networks used in Suphx have more than 30 layers, which makes the efficiency of training low, especially in reinforcement learning.

4. Proposed Method

We train five models which are discard model, Chow model, Pong model, Kong model and Riichi model. We introduce the features of discard model in details first. For features of other models, we used the same feature set with few additional features such as the open meld tiles and the discarded tile of other players.

4.1 Proposed Features

In the game of Japanese Mahjong, the design of the features is very important. The features have to include information from the player's view as much as possible. We present a set of simple but effective features, which are easier for human to understand and reproduce than that in Suphx. There are two types of features, one of private information and one of public information. We use the rules of Japanese Mahjong to extract more information based on our agent's observation. We use convolutional neural networks, so the channel design is very important. We propose to use the channel of 34 by 4, because Japanese Mahjong has 34 different types of tiles and 4 for each type. As a simple example, we encode the hand tiles shown in Fig. 2 into encoding shown in Fig. 1. We use a total of 273 channels to present observation of our agent.

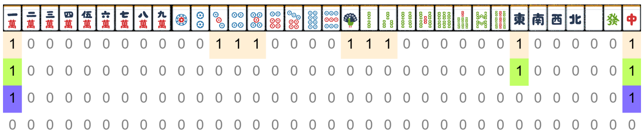

Fig. 1 Representation of the tiles shown in Fig. 2

Fig. 2 An example of current hand tiles

4.2 Features of Private Information

We introduce our features of private information in this subsection. These features of private information are all built from the agent's hand tiles. A player's hand can be divided into closed meld which can be visible only to himself, and open meld which contains all melds that are visible to all players.

4.2.1 Features Represented Hand Tiles

For discard model, we use one channel to encode the information of hand tiles including closed tiles and all meld tiles. In a channel, the horizontal axis means the type of tiles and the vertical axis means the number of the same tiles. We change zero to one from the top to the bottom and the number of ones represent the number of that same tiles in a player's hand. Then we use another channel to save the information of only closed tiles in hand in the same way.

4.2.2 Features Obtained with Respect to Yaku

We use 154 channels obtained from hand tiles according to the melds and yakus in the game of Mahjong.

Sequences, triplets and quads are three important form of tiles. We use 21 channels, 34 channels and 34 channels to represent the information of these three forms of tiles in hand, respectively.

We use channels that save sequences' information as an example. As Fig. 3 shows, we use 21 channels to represent sequences' information because there are totally 21 types of sequences in Japanese Mahjong. We first initialize all channels to zeros. When a sequence appears in the hand, we change the numbers in the corresponding channel from zeros into ones. Suppose that an agent has the same hand tiles in Fig. 2 for example, the 10th of sequences' channels is shown as Fig. 4. Channels of triplets and quads are built in the same way as sequences' channels.

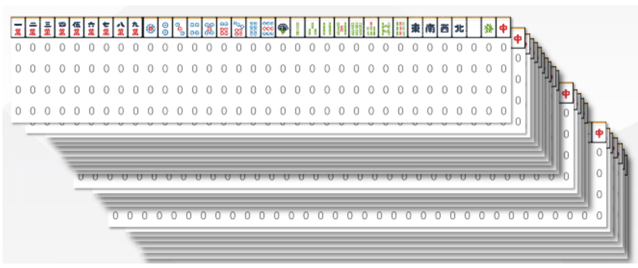

Fig. 3 An example of all sequences saved by 21 different channels

Then we introduce 65 channels to present how the hand

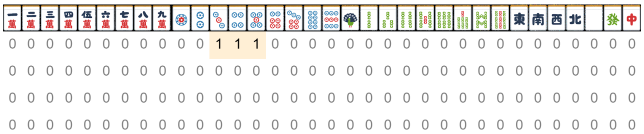

Fig. 4 Channel specialized for 3-4-5 pin sequence

Table 1 Features of Private Information

Feature Name	Number of Channels
Hand tiles	1
Menzen tiles	1
Sequences	21
Triplets	34
Quads	34
Tanyao	2
Seven pairs	2
Toitoi	2
Chin'iisou	6
Hon'iisou	6

tiles are near to each yaku, which are important for a player to win a round.

We designed two categories of channels to represent the distance between a hand and a yaku; (1) tiles in the hand are compatible with the yaku, and (2) tiles not in the hand required for the yaku. For the former category, we consider tanyao, seven pairs, toitoi, chin'iisou and hon'iisou. All these yakus share a common property that if an agent wants to form these yakus in its hand, the agent has to discard all tiles that are not compatible with these yakus. Each channel has an associated yaku where each value represents whether the tile is useful (one) or not (zero). We give an example for seven pairs, a type of hand that the player can declare winning in a round. By this rule, the values in the seven pairs channel for a hand shown in Fig. 2, is shown in Fig. 5. It is obvious that from this channel we can know that there are now 3 pairs in the player's hand tiles. At the same time, we save the remaining tiles in another channel shown by Fig. 6. As we can see in Fig. 6, the channel shows such tiles that appear only once in hand or more than twice in hand. For tanyao, there is a similar pair of channels like that of seven pairs. One of the channels (channel of usable tiles for tanyao) represents number tiles from two to eight in the hand, including single tile not forming a meld or pair. The other one channel (channel of unusable tiles for tanyao) represents other tiles in the hand. We also use 6 channels for chin'iisou and hon'iisou respectively because there are three types of chin'iisou and three types of hon'iisou.

All channels introduced above are shown in **Table 1**.

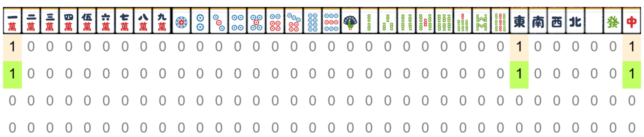

Fig. 5 Seven-pair feature (usable) for the tiles in hand shown in Fig. 2

For the latter category, we consider other yakus that can be calculated in a winning hand with some tiles that are not contained by these yakus. Yakus and the number of channels are shown in **Table 2**. We consider ikkitsuukan only when there are 8 or 9 different tiles of one type in a player's hand tiles. For chanta, we divide tiles in hand tiles into three types: tiles have formed sequences or triplets that can be used in chanta, tiles that can form sequences or triplets in chanta but have not completed now, and tiles that can not be used in chanta. For other yakus in the Table 2, we use intersection of hand tiles to them. For example, there are 21 types of iipeikou and we calculate intersection of each type with the player's hand tiles separately.

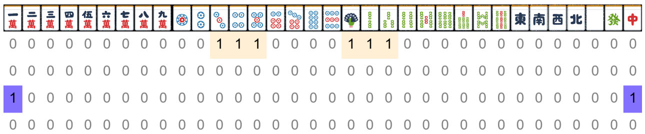

Fig. 6 Seven-pair feature (non-usable) for the tiles in hand shown in Fig. 2

Table 2 Features of Private Information

Feature Name	Number of Channels
Ikkitsuukan	2
Iipeikou	21
Sanshoku doujun	7
Sanshoku doukou	9
Kokushi musou	1
Player wind	1
Round wind	1
Dora	1
Akadora	1
Chanta	3

4.3 Features of Public Information

Details of features of public information are represented in **Table 3**.

4.3.1 Visible Tiles

We use one channel to represent all revealed tiles, including closed melds, discarded tiles, open melds of all players and dora indicators.

4.3.2 Melds

Information of all 4 players' open melds use 16 channels because each player can declare at most 4 melds. The order of channels also indicate the order of declaring melds.

4.3.3 River

River is formed by discarded tiles. Every player has his own river and his discarded tile is added at the end of his river, one by one. We use 20 channels for each player's river and each channel only contains one discarded tile, so that an agent recognizes the order of 20 channels to represent the order of discarded tiles.

4.3.4 Public Information and Agent's Hand

Each of an agent's wind and a round wind is presented in its own channel that indicates the corresponding tiles owned by the player. Dora indicators are also an important information for all players and we also use one channel to save the information. The channel for dora indicators represents all dora indicators at the proper location, i.e., ignores even if there is the same type of tiles in hand or river.

Table 3 Features of Public Information

Feature Name	Number of Channels
Revealed tiles	1
Melds	16
River	80
Player wind	1
Round wind	1
Dora indicator	1

Table 4 Special Channels

Feature Name	Number of Channels
Riichi status	4
Rank	4
Round number	8
Menzen status	1

Table 5 Result

Model	Our Accuracy	Suphx's Accuracy [4]
Discard model	72.3%	76.7%
Riichi model	86.1%	85.7%
Chow model	89.8%	95.0%
Pong model	88.2%	91.9%
Kong model	98.1%	94.0%

Table 6 Details in Training Networks

Model	Training Data Size	Validation Data Size
Discard model	$1.8 \cdot 10^6$	$2.6 \cdot 10^4$
Riichi model	$6.1 \cdot 10^5$	$2.1 \cdot 10^4$
Chow model	$1.4 \cdot 10^6$	$1.7 \cdot 10^4$
Pong model	$2 \cdot 10^6$	$2.4 \cdot 10^4$
Kong model	$1.3 \cdot 10^5$	$1.8 \cdot 10^4$

Table 7 Details in Training Networks

Model	Initial Learning Rate	Decay Rate
Discard model	0.01	0.5
Riichi model	0.001	0.5
Chow model	0.001	0.1
Pong model	0.001	0.1
Kong model	0.001	0.1

4.3.5 Special Channels

We introduce auxiliary status of each player; Riichi, menzen, rank and round number.

The status whether a player declared Riichi or is suitable for menzen tsumohou (a type of yaku) are visible to all players. As shown in **Table 4**, we use four channels of all players' Riichi status and one channel of our agent's menzen tsumohou status. If a player declares Riichi, the channel of his Riichi status will be changed from all zeros to all ones. If our agent does not declare open melds, the channel of our menzen tsumohou status is all ones, otherwise all zeros.

4.4 Additional Channels in Other Networks

The input features of Riichi network and discard network are the same. We added three dense layers before the output in Riich network to transform the output dimension from 34 to 2 like that in Suphx (used two dense layers before the output).

In Chow and Pong networks, there are totally 275 channels. The first 273 channels are the same as that in discard network. Two more channels are used to represent the open Chow (or Pong) meld to be formed and the discarded tile to get by declaration. There are at most three possible types of Chow when a player discards a tile, we need to calculate the output of Chow network three times with changing the channel representing Chow meld. We choose the action with the highest score output by the Chow network.

In Kong network, there are totally 277 channels. The first 273 channels are also the same as that in discard network. Then one channel is used for meld tiles like that in Chow or Pong networks. The last three channels represent three types of Kong. The channel corresponding to current Kong type is set to all ones and the two other channels are set to all zeros.

If there are two or more actions are available at the same time (for example, Chow and Pong are both possible), the agent judges whether to reject these actions separately. When the agent chooses two or more actions, the action with the highest score corresponding to the output of the networks is selected.

4.5 Proposed Network Structures

Suphx used convolutional neural networks with 50 residual blocks, which are a little bit large from our perspective. So we use similar but smaller networks in our training. For computational efficiency, we reduce the number of residual blocks in neural networks from 50 to 30. Weconducted supervised learning to train our models, following the training procedure of Suphx.

5. Result of Supervised Learning

We trained all five networks by supervised learning.

The data are preprocessed from records of human experts download from Tenhou from January 2018 to May 2019. We used cross-entropy for loss function and SGD for optimizer. More details are shown in **Table 6** and **Table 7**.

Our Riichi model has reached an accuracy of 86.1% and our Kong model has reached an accuracy of 98.1%. The training accuracy and validation accuracy of the models are shown in Fig. 7, Fig. 8 , Fig. 9 , Fig. 10 and Fig. 11. Comparison of our result and that in Suphx is shown in **Table 5**. The accuracy of our discard model, Chow model and Pong model has not reached that in Suphx, but the numerical difference are not so large.

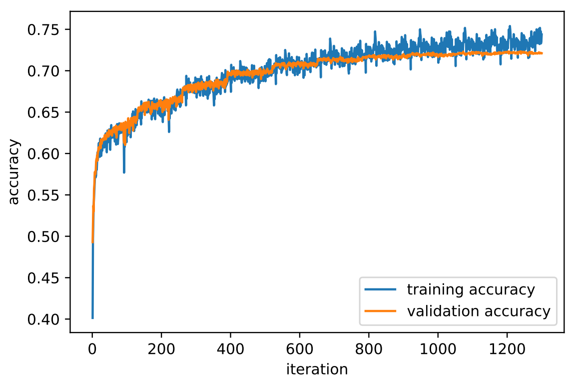

Fig. 7 The accuracy of discard network

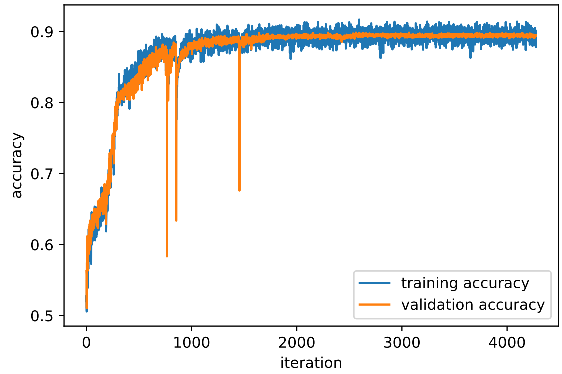

Fig. 8 The accuracy of Chow network

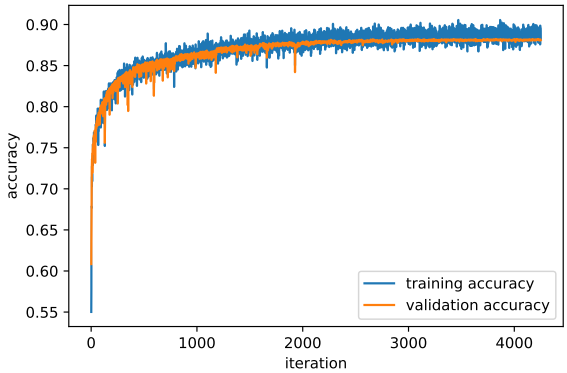

Fig. 9 The accuracy of Pong network

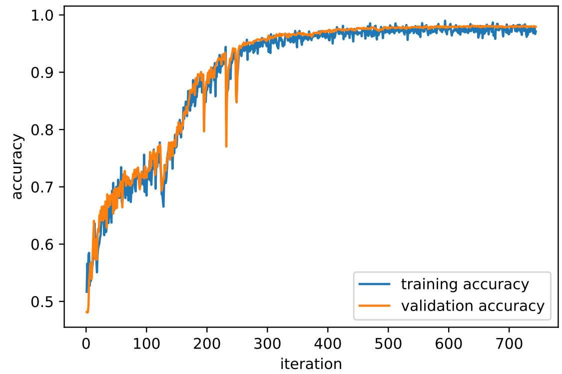

Fig. 10 The accuracy of Kong network

6. Summary and Future Works

6.1 Summary

We proposed a set of features for presenting private and public information of Mahjong. We download human records from Tenhou platform and used them to train neural networks by supervised learning. By supervised learning, the accuracies of our networks have almost reached those of Suphx, even though our networks are smaller. The results indicate the effectiveness of our features.

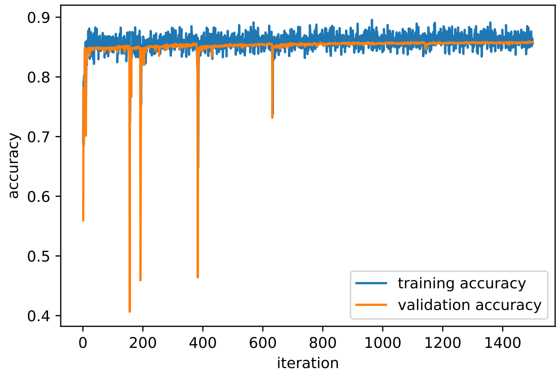

Fig. 11 The accuracy of Riichi network

6.2 Future Works

The accuracy would increase by adding more features such as considering more yakus and current points of all players. Current points are important for a player to decide whether to be offensive or defensive, especially in the last round. Furthermore, reinforcement learning would be beneficial to improve the performance of the networks. We will evaluate the effectiveness of the proposed set of features throughout reinforcement learning.

References

[1] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. et al.: Mastering the game of Go with deep neural networks and tree search, *nature*, Vol. 529, No. 7587, p. 484 (2016).

[2] Tenhou: https://tenhou.net/. (Online; accessed 27-July-2020).

[3] Gao, S., Okuya, F., Kawahara, Y. and Tsuruoka, Y.: Supervised Learning of Imperfect Information Data in the Game of Mahjong via Deep Convolutional Neural Networks, *Information Processing Society of Japan* (2018).

[4] Li, J., Koyamada, S., Ye, Q., Liu, G., Wang, C., Yang, R., Zhao, L., Qin, T., Liu, T.-Y. and Hon, H.-W.: Suphx: Mastering Mahjong with Deep Reinforcement Learning, *arXiv preprint arXiv:2003.13590* (2020).

