Mahjong AI Competition: Exploring AI Application in Complex Real-World Games

Yunlong Lu , **Wenxin Li**

School of Computer Science, Peking University
luyunlong@pku.edu.cn, lwx@pku.edu.cn

Abstract

This paper presents three Mahjong AI competitions we held at IJCAI. We briefly introduce the rule of Mahjong and its challenges to AI algorithms. By showing the results and the application of various algorithms in the competitions, we claim that existing algorithms show promising results in Mahjong, while open problems remain and more efforts are needed towards solving this complex game.

1 Introduction

Games have long been benchmarks and testbeds for AI research. In recent years, with the development of AI algorithms and boost of computational power, AI systems have achieved superhuman performance in many games like Go [Silver *et al.*, 2017], StarCraft [Vinyals *et al.*, 2019], and Texas Hold'em [Zhao *et al.*, 2022]. These games are all popular with seasonal and annual events held throughout the world. Such popularity motivates the academic community to invest efforts and develop new algorithms to solve them.

Mahjong is popular all over the world, especially in China with many regional variants. It poses challenges to AI algorithms due to its imperfect-information and multi-objective nature, but has been overlooked by the community of AI research. To promote AI research and explore AI application in Mahjong, we held three Mahjong AI competitions at IJCAI. Dozens of teams from academia and industry participated and applied a variety of algorithms to build their agents. We organized symposiums every year and invited top teams to give oral presentations to share their methods.

The competition results and their presentations show that modern AI algorithms based on deep learning have huge potential on this game and outperform heuristic methods. However, there are still some open questions that need to be solved in order to further enhance the performance of AI agents. We hope our experiences of the competitions can promote further AI research in complex real-world games like Mahjong.

2 Mahjong

Mahjong is a four-player tabletop game almost as popular in China as Texas Hold'em in America. There are many regional variants of Mahjong rules, but the basic rules are the same.

The game is played with a set of 144 tiles with symbols and characters. Each player begins with 13 tiles which are not shown to other players. They take turns to draw and discard one tile until one completes a winning hand with a 14th tile. Players can also take other's discarded tile to form a meld or declare a winning hand. Winning hands must match some specific patterns, such as four melds and a pair, or seven pairs, which are determined by the rules of each variant. In general, players have to adjust their tiles towards some winning patterns, which requires both skills and luck to win.

Among various regional variants, the rule adopted in our competitions is called Mahjong Competition Rules (MCR), which is recognized and standardized by Mahjong International League (MIL). MCR contains 81 different scoring patterns and each winning hand must match patterns with no less than 8 points. The complete rules of MCR can be found in [Lu *et al.*, 2023]. Its rich winning patterns and complex scoring rules emphasizes strategic thinking and reduces the factor of luck, making it more competitive than other variants and is widely used in Mahjong competitions.

Mahjong presents several challenges for AI algorithms. First, players have to deduce other's private tiles based on their discarded tiles, indicating rich hidden information. As a game with imperfect information, the average size of information sets in Mahjong is around 10^{48}, much larger than games like Poker and Bridge, making it hard to apply CFR-based algorithms which solved Texas Hold'em. Second, the randomness in dealing initial hands and drawing each tile introduce high variance and instability in convergence when applying learning-based algorithms. Third, unlike poker games where each card has its rank, Mahjong tiles are equal in their role in forming patterns, so it is hard to obtain a clear value judgement of each hand. Finally, the strategy of Mahjong involves dynamic choices between various target winning patterns, which can constantly change upon drawing each tile, while other games usually have a single target like making a bigger hand or emptying the hand faster. These unique factors make it difficult to acquire an optimal policy of Mahjong.

3 Competition

To promote MCR as a playground for AI research and explore the potential of modern game AI algorithms in it, we held three Mahjong AI competitions in 2020, 2022, and 2023 at IJCAI. This section gives an introduction of them.

Platform   Our competitions are all held on Botzone [Zhou *et al.*, 2018], an online multi-agent game AI platform built by our lab. This platform provides a universal interface to evaluate AI agents in dozens of games regardless of programming languages. Players can submit their programs and play against others' as long as they follow the input/output protocol specified by the platform. Aside from creating isolated games between AI agents or human players, Botzone also supports to create groups to hold individual tournaments where multiple rounds of games can be played between a fixed set of agents under various formats.

Competition Format   As a four-player game with rich hidden information and a high degree of randomness in dealing tiles, the results of Mahjong games are heavily influenced by luck. To reduce the variance of individual games and get a more accurate ranking of AI agents, we applied a novel competition format, combining Swiss rounds and duplicate format. The idea of duplicate format is to play 24 games as one "duplicate game", with four players exchanging their seats and each seat always being dealt the same tiles. In this way, the advantage from "lucky hands" and the order of play can both be eliminated. We use Swiss round as the high-level format, with each round pitting the closest-ranked players against each other. While typical Swiss-round tournaments have a relatively small number of rounds, hundreds of rounds are needed in our competition to get a stable ranking because of the randomness and imperfect information in Mahjong.

Schedule   As shown in Table 1, the schedules of three competitions are similar and are all divided into three stages: a qualification round to determine the top 16, an elimination round to determine the top 4, and a final round to determine the final winner. We increase the number of Swiss rounds every year as the level of programs is improving and getting closer. To encourage the use of modern AI algorithms like deep learning, we provide the participants with datasets, consisting of either human matches from online game platforms or high-quality self-play matches from top AI agents in previous competitions. Since MCR has quite complex scoring rules and patterns, we also provide an open source library [Lu, 2020] to calculate the scores of winning hands with two versions in Python and C++. With these conveniences, the participants can focus on designing various AI algorithms and many teams used the provided dataset for algorithms based on machine learning, creating powerful agents to play Mahjong.

Results   In the third Mahjong AI Competition, four teams from three companies and twelve teams from five universities advanced to the elimination round, and the team from Tencent Holdings Ltd won the first place eventually. The full results of all three competitions can be found on their homepages[1][2][3]. We organized online symposiums every year where most of the top 16 teams made oral presentations to discuss their algorithms and share their experiences in building their agents. While no offline events were held in the first two competitions because of the COVID-19 pandemic, the awarding

---

[1]https://botzone.org.cn/static/gamecontest2020a.html

[2]https://botzone.org.cn/static/gamecontest2022a.html

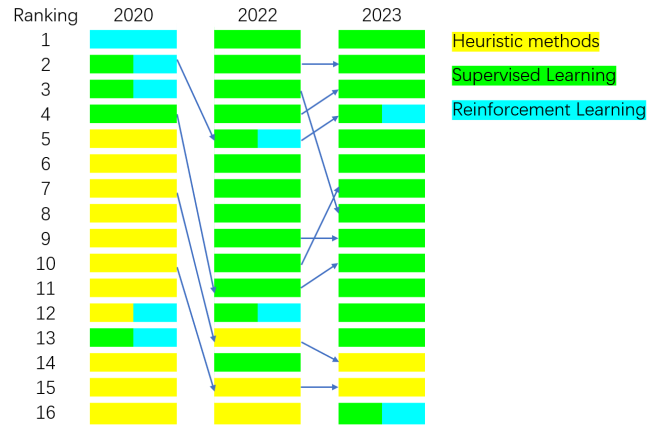[3]https://botzone.org.cn/static/gamecontest2023a.html

Figure 1: The category of algorithms used in the top 16 agents in three Mahjong AI competitions. The arrows indicate that the agents are uploaded by the same team who participated more than once.

ceremony and an offline symposium of the third competition were organized on site in Macao during IJCAI-23, allowing participants to have in-depth exchanges of their methods.

4   Algorithms

In this section, we summarize the algorithms to build Mahjong AI agents in the three competitions based on the presentations given by the participants. Though the details of the algorithms vary, they can be divided into three categories: heuristic methods, supervised learning, and reinforcement learning. An overview of the algorithms of the top 16 agents in three competitions is illustrated in Figure 1.

4.1   Heuristic Methods

Most teams in the first competition use heuristic methods to build their agents, which rely hevily on human knowledge. Human players usually measure their distance from winning based on "shanten value", defined to be the minimal number of tiles to replace from the current hand to a winning pattern. Many teams calculate this value by expanding a search tree from the current hand and trying all possible tiles to replace until it forms a winning pattern, and make actions based on it. When the shanten value is high, predefined rules are defined to choose target patterns based on the experience of human players, like making "Seven Pairs" when there is five pairs in the hand. By considering dozens of scoring patterns as targets and choosing one based on hard-coded rules, a complex behavior tree can be constructed by applying different strategies for each pattern. Some teams further calculate the win-rate by estimating the probability of occurrence of each tile based on visible information, and choose action to both reduce shanten values and maximize win-rates. In all, the performance of these agents is mainly restricted by the strength of heuristics, and the behavior trees designed by better human players tend to be more complex and perform better.

4.2   Supervised Learning

Supervised learning (SL) is also widely used, especially after the first competition where agents trained by SL outper-

Year	Qualification Round			Elimination Round			Final Round			Dataset
	Time	Teams	Rounds	Time	Teams	Rounds	Time	Teams	Rounds	
2020-2021	Nov. 31	37	18	Jan. 1	16	96	Jan. 6	4	128	Human data
2022	May 22	25	128	Jul. 3	16	128	Jul. 4	4	512	AI data
2023	May 21	39	128	Jul. 3	16	256	Jul. 4	4	512	AI data

Table 1: Summary of the schedules of three Mahjong AI competitions we held in IJCAI.

formed those with heuristic methods, indicating its huge potential in this game. More teams turn to SL in the next two competitions, making up most of the top 16 agents. Its basic idea is to train a policy model to predict the actions under each game state by cloning the behavior from the match dataset of human players or AI agents, so the performance of the model depends on the quality of the dataset.

Though many teams use SL, their implementations vary in three components: the design of features, the structure of network model, and the data preprocessing scheme. Many teams encode tile features as images to capture the spatial relationship between suited tiles of the same number, and use convolutional neural networks (CNN) to extract high-level features from the input. Some teams try variants of CNNs such as Resnet [He *et al.*, 2016] and ResNeXt [Xie *et al.*, 2017], which adopt deeper layers by adding shortcut connections or use kernels with more complex structures. Their experiments show the design of image-like features and deeper CNN models can indeed increase the accuracy of the trained model. Some teams use data augmentation by exchanging tiles whose roles are symmetric in forming scoring patterns, which significantly increase the amount of data and improve the generalization ability of the network model. The team ranked first in the third competition adds high-level features such as shanten value and shows it can boost both the accuracy and performance of trained models.

4.3 Reinforcement Learning

The top three teams in the first competition all use a similar paradigm of reinforcement learning (RL) to train their agents. RL learns to choose actions to maximize cumulative rewards by interacting with other players, usually in a self-play manner in multi-agent settings. Their methods combine the algorithm of Proximal Policy Optimization (PPO), a distributed training framework IMPALA, and the self-play of the latest model to collect training data. PPO [Schulman *et al.*, 2017] is a policy-based RL algorithm widely used in game AI because of its stability in training. IMPALA [Espeholt *et al.*, 2018] is an actor-critic training framework consisting of a learner and multiple distributed actors, which can scale to large amount of computational resources. In fact, these teams were all from companies and used hundreds of CPU cores in their training.

Some of them also use SL to train an initial model from the given dataset and run RL algorithms from it to speed up the training in the early phase. Besides, the team ranked first in the first competition adopts reward shaping by clipping game scores to a smaller range to migrate the high variance and stabilize the training. The team ranked fifth in the third competition tries Deep Q-Network (DQN) [Mnih *et al.*, 2013], a

value-based RL algorithm, and claims that it performs better than PPO under limited computational resources, but is still weaker than the models trained by SL. In all, the performance of RL methods depends largely on computational resources used, especially in Mahjong where randomness and hidden information can cause variance and instability in training.

5 Discussion

Though many teams have adopted modern algorithms like SL or RL, we conclude that there still remain some open problems to be solved in building stronger Mahjong agents, based on the presentations of participants in the latest competition.

One of them is the evaluation of agents' performance during training. In the first two competitions, teams using SL tend to measure the model's performance by its accuracy on validation sets, which cannot reflect its level of play from experiments of participants in the third competition. While our competition format with tens of thousands of games cannot be afforded to evaluate multiple checkpoints during a training session, some teams are proposing new metrics like number of turns or win-rates of self-play matches as indications of performance, but their effectiveness has yet to be proven.

Another problem is the instability of RL training. Experiments show that agents trained by RL from scratch never choose some winning patterns, and even pretrained SL models can lose the ability to choose them during RL training, possibly because some patterns cannot be efficiently explored. Since Mahjong strategies involve dynamic choices between various patterns, we claim that such multi-objective nature of Mahjong poses new challenges for existing algorithms, which is not seen in other imperfect-information games, and new techniques are required to tackle it.

6 Conclusion

To explore the potential of AI algorithms in Mahjong, we organized three competitions at IJCAI by providing participants with the platform, game environment, datasets, and a novel competition format to migrate the factor of luck. Dozens of teams from both academia and industry participated and applied a variety of algorithms on this game. The competition results show that modern game AI algorithms based on deep learning tend to perform better than heuristic methods relying on human expertise, while there still remain open problems to be solved due to the complexity of Mahjong with its high randomness, imperfect information, and dynamic choices between multiple targets. From our experiences of these competitions, we believe this game can be a new playground for game AI research and can promote the development of multi-agent AI algorithms in the setting of imperfect information.

Ethical Statement

There are no ethical issues.

References

[Espeholt *et al.*, 2018] Lasse Espeholt, Hubert Soyer, Remi Munos, Karen Simonyan, Vlad Mnih, Tom Ward, Yotam Doron, Vlad Firoiu, Tim Harley, Iain Dunning, et al. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. In *International Conference on Machine Learning*, pages 1407–1416. PMLR, 2018.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Lu *et al.*, 2023] Yunlong Lu, Wenxin Li, and Wenlong Li. Official international mahjong: A new playground for ai research. *Algorithms*, 16(5), 2023.

[Lu, 2020] Yunlong Lu. GitHub - ailab-pku/PyMahjongGB: Python Fan calculator for Chinese Standard Mahjong — github.com. https://github.com/ailab-pku/PyMahjongGB, 2020. [Online; accessed 19-Feb-2024].

[Mnih *et al.*, 2013] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[Silver *et al.*, 2017] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

[Vinyals *et al.*, 2019] Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg, Wojciech M Czarnecki, Andrew Dudzik, Aja Huang, Petko Georgiev, Richard Powell, et al. Alphastar: Mastering the real-time strategy game starcraft ii. *DeepMind blog*, 2, 2019.

[Xie *et al.*, 2017] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[Zhao *et al.*, 2022] Enmin Zhao, Renye Yan, Jinqiu Li, Kai Li, and Junliang Xing. Alphaholdem: High-performance artificial intelligence for heads-up no-limit texas hold'em from end-to-end reinforcement learning. 2022.

[Zhou *et al.*, 2018] Haoyu Zhou, Haifeng Zhang, Yushan Zhou, Xinchao Wang, and Wenxin Li. Botzone: an online multi-agent competitive platform for ai education. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pages 33–38, 2018.