Mastering Bidding in Fight the Landlord with Perfect Information Distillation

1st Xiaochuan Zhang
School of Artificial Intelligence
Chongqing University of Technology
Chongqing, China
zxc@cqut.edu.cn

2nd Xin Wang
School of Artificial Intelligence
Chongqing University of Technology
Chongqing, China
2463274417@qq.com

3rd Xiaoman Yang
School of Artificial Intelligence
Chongqing University of Technology
Chongqing, China
1093550752@qq.com

4th Mingzhu Yan
School of Artificial Intelligence
Chongqing University of Technology
Chongqing, China
1724033972@qq.com

5th Piao He
School of Artificial Intelligence
Chongqing University of Technology
Chongqing, China
1910368053@qq.com

2023 35th Chinese Control and Decision Conference (CCDC) | 979-8-3503-3472-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/CCDC58219.2023.10327476

***Abstract*—The popular card game DouDizhu in China has become a research hotspot of computer game because of its unique characteristics. For the bidding stage of DouDizhu game, accurate evaluation of hand cards strength is the key to mastering accurate bidding. In this paper, we propose a method combining convolutional neural network and reinforcement learning, and add the feature of winning distance. Supervised learning is used to train agent, and perfect information distillation technology is used to optimize bidding methods.The experimental results show that the bidding network model proposed in this paper is effective, and by comparing with some open source DouDizhu AI, it proves that the bidding method in this paper is effective.**

***Index Terms*—Bidding in DouDizhu,Convolutional Neural Network,Perfect Information Distillation,Winning Distance**

I. Introduction

In the development of artificial intelligence, computer games serve as an abstraction for many real-world problems. In recent years, the research of perfect information game has made great achievements, such as AlphaGo[1], AlphaGo Zero[2] has become the most advanced AI in Go game. Nowadays, the research on computer games has developed into a more challenging imperfect information game, in which the information observed by each player in the game is different and incomplete. From two-player games, such as two-player limit/non-limit Texas Hold'em[3][4] to multiplayer games, including multiplayer Texas Hold'em[5], StarCraft[6],DOTA[7], and Japanese mahjong Suphx[9], has achieved exciting achievements in various fields.

One of the imperfect information games studied in this paper—Fight the Landlord(a. k. a. DouDizhu),has the characteristics of stages, cooperation and competition, and large action space, which is very similar to the real game scene and brings great challenges to artificial intelligence. In 2019, You et al. [10]compared DQN, A3C and RHCP (recursive disassembly), found that the winning rates of DQN and A3C were only below 20% . In the same year, DeltaDou[11] used an algorithm similar to AlphaZero to combine Fictitious Play MCTS(FPMCTS) with a neural network, and used Bayesian reasoning to make DouDizhu AI reach the level of top human players for the first time. However, DeltaDou's reasoning and calculation are too dependent on heuristic rules, and the calculation is very heavy, and it takes two months to complete the training. In 2021, DouZero[12]has attracted much attention due to its outstanding performance. It uses self-play deep reinforcement learning,and proposes Deep Monte Carlo(DMC) combined with neural networks. Encoding the features of the playing cards perfectly handles the large state space of DouDizhu, and opens a window for such complex and large-scale games. Recently, the emergence of PerfectDou[13] has become the most powerful DouDizhu AI. It proposes the perfect information distillation technology, adopts a perfect training-imperfect execution framework, and achieves the best performance so far.

In the game of DouDizhu, bidding, as the initial and important stage, plays a key role in the final income. An excellent bidding can give the most appropriate bidding score according to one's own hand cards at the beginning, so as to solve the problem of not being able to maximize one's own interests due to bidding. At present, the research on DouDizhu bidding has also achieved certain development, and has been applied on some game platforms[14].

In this paper, aiming at the bidding problem of the agent in DouDizhu,an improved bidding network model based on ResNet is proposed, combined with a method of reinforcement learning. This method integrates the hand cards strength features, sequence features, and specially proposed winning distance features,and uses perfect information distillation to optimize the bidding strategy. Specifically, the agent is allowed to use the global information as a perfect information game to guide the bidding strategy, and use the trained strategy to

This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61702063.

play the imperfect information game in the actual game. The proposed method effectively solves the problem of inaccurate and efficient bidding in DouDizhu.

II. RELATED WORK

A. DouDizhu bidding model

Since the launch of DeltaDou, DouZero, and PerfectDou, Doudizhu AI has reached a level that can defeat top human players. However, these studies are all focused on the playing stage of DouDizhu, and the research on the bidding stage is still relatively scarce. For the study of Doudizhu bidding stage, Li et al.[14]proposed an evaluation model based on classification algorithm, which uses the Adaboost algorithm to train a classifier from DouDizhu's opening hand cards data, and then evaluates its own and opponent's hand cards strength to decide whether to bid. Yuan et al.[15]proposed a bidding recommendation model SeqStgCNN based on convolutional neural network (CNN), which achieved good results in DouDizhu bidding.

B. Reinforcement Learning for Imperfect-Information Games

In recent years, reinforcement learning has been successfully applied to many imperfect information games. For example, the application of reinforcement learning in poker games[16][17],different from the Counterfactual Regret Minimization (CFR) algorithm based on search tree traversal,reinforcement learning is based on sampling and solves the problems of large action space and cooperative games. In this way, OpenAI, DeepMind and Tencent used reinforcement learning to construct their own game AI in DOTA[7],StarCraft[6]and Honor of Kings[8]respectively, and achieved amazing results, proving that reinforcement learning effectiveness in imperfect information game. Recently,some studies have combined reinforcement learning with tree search, and demonstrated the effectiveness of this method in poker games,such as no-limit Texas Hold'em[17][18] and DouDizhu[10][11].

Due to the unique characteristics of the DouDizhu game, traditional reinforcement learning methods such as DQN[19]and A3C[20]do not perform well in this project. Even Combinatorial Q-Networks[9], an improvement on DQN, is not satisfactory. DeltaDou[11]uses Bayesian to reason about hidden information and uses MCTS to combine reinforcement learning with tree search, but its high computational cost limits its practicality and performance. To this end, DouZero[12]used the Deep Monte-Carlo(DMC)to defeat all previous DouDizhu AI. Recently, PerfectDou[13]proposed to use the perfect information distillation to become the most powerful AI in DouDizhu. Compared with other imperfect information games, DouDizhu is more complex, requiring competition and cooperation in a large number of action-state spaces, and the game process is divided into two stages: bidding and playing. The outstanding performance of DouZero and PerfectDou reveals that reinforcement learning can achieve good results in such complex card games, and provides new ideas for future research dealing with complex action-state spaces, imperfect information, and the coexistence of competition and cooperation.

III. METHOD

A. Problem Definition

In a DouDiZhu game, three players take turns drawing from a standard shuffled deck of 54 cards until the remaining three cards are not drawn. At this time, start bidding. According to certain rules, one player will bid first, and then rotate counterclockwise to bid in turn. Each player can only bid once at most, and the score of bidding is ”0” (not call), 1”, ”2”, ”3”. If the score of the player who bids first is lower than 3 points, the players behind can choose higher points to compete for the landlord or not to bid. If a player bids 3 points, the player becomes the landlord, and the three hole cards belong to him. And the remaining two players become peasants to fight against the landlord together.

It can be seen that the advantage of the landlord in the whole game is that he has three extra hole cards, but the hole cards are public information known to all three players, and the disadvantage is that he has no help from his companions. Landlords can get higher rewards after winning the game, and also get greater punishment after losing the game. Therefore, it is very important to evaluate the strength of one's own hand cards before choosing whether to contest the landlord in the bidding stage.

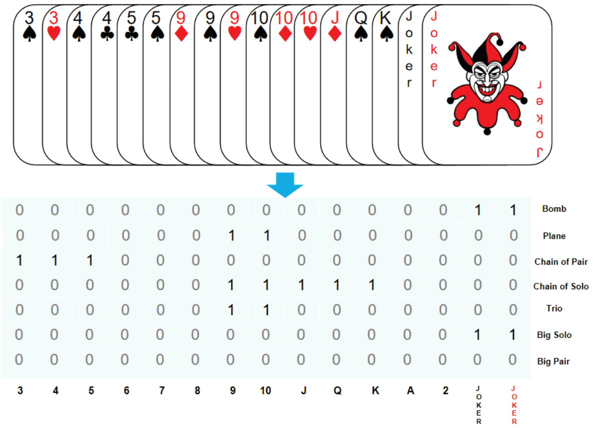
According to the above introduction, the bidding method designed in this paper is to let the agent learn to analyze the current hand cards strength and bidding situation information, and finally make accurate bidding decisions. Specifically, the method of bidding can be abstracted as whether to compete for the role of the landlord, and how many points should be called to maximize the benefits when competing for the role of the landlord. As shown in formula (1):

$$F(X) = \begin{cases} 0, Pass \\ 1, Call \quad 1 \quad contest \quad the \quad landlord \\ 2, Call \quad 2 \quad contest \quad the \quad landlord \\ 3, Call \quad 3 \quad contest \quad the \quad landlord \end{cases} \tag{1}$$

where F(X) is the bidding prediction model, and X is the feature received by the prediction model. When the model output is 0, it means that the agent is not recommended to contest the landlord. When the output is 1, 2, or 3, it means that the model recommends the agent to contest the landlord, but the expectations for the landlord are different. The expected value of 1 is the lowest, and the expected value of 3 is the largest. Where X is defined as the formula (2):

$$X =< X_{sequence,}X_{strength,}X_{winning-distance} > \tag{2}$$

where $X_{sequence}$ is the sequence feature of the hand cards, $X_{strength}$ is the strength feature of the hand cards, and $X_{winning-distance}$ is the winning distance feature of the hand cards.



0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	Bomb
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	Plane
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Chain of Pair
0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	Chain of Solo
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	Trio
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	Big Solo
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Big Pair
3	4	5	6	7	8	9	10	J	Q	K	A	2	JOKER	JOKER	

Fig. 1. Hand strength feature representation matrix. The columns represent 15 different card types, and the rows represent strong legal combinations.

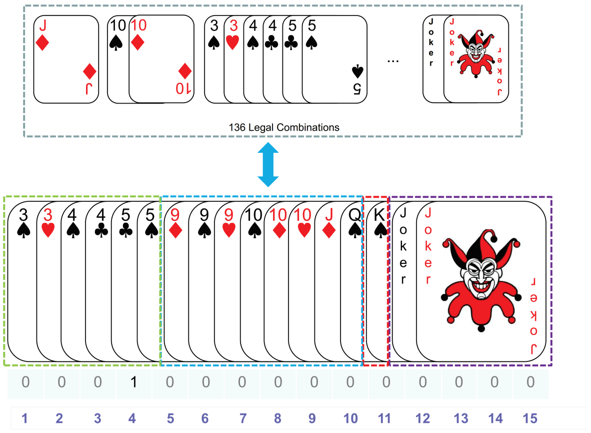
B. Card Representation

In the study of this paper, the One-Hot encoding method is used to encode the strength features, sequence features and winning distance features of the hand cards. All feature designs are shown in Table I. The coding matrix obtained according to the features of hand cards strength is shown in Figure 1, which includes seven combinations of bomb, plane, chain of pair, chain of solo, trio, big pair and big solo. Through 7 combinations of cards of different sizes, it helps the agent to identify the card strength of different combination cards. For the hand cards sequence feature, each hand card corresponds to a 1×15 matrix, for example, ”K” corresponds to the matrix dimension [0,0,0,0,0,0,0,0,0,0,1,0,0,0,0],and finally combine all 17 hand cards into a 17×1×15 coding matrix.

TABLE I
THE FEATURE DESIGN OF THE GAME'S OWN HAND INFORMATION AND PERFECT INFORMATION. PERFECT INFORMATION FEATURES INCLUDE ALL FEATURES.

	Feature	Size
Own Hand Cards Feature	Bomb	1X1X15
	Plane	1X1X15
	Chain of Pair	1X1X15
	Chain of Solo	1X1X15
	Trio	1X1X15
	Big Pair	1X1X15
	Big Solo	1X1X15
	Sequence	17X1X15
	Winning Distance	1X1X15
Perfect Feature	Last Player Hand Cards	25X1X15
	Next Player Hand Cards	25X1X15
	Own Hand Cards	25X1X15
	Hole Card	1X1X15

In particular, this paper proposes the winning distance feature as the input of the network model. The so-called winning distance is the measurement value of the player's 17 hand cards and the state of the end hand cards,that is, how many steps the player needs to use to finish the hand cards. The winning distance proposed in this paper is the number of steps to complete all hand cards considering only your own hand cards. As shown in Figure 2, the hand cards can be divided into 136 different combinations of playing cards, and the number of combinations that can be selected to end the hand the fastest is 4, so 1 is set in 4 positions, and 0 is set in other positions.



0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Fig. 2. Winning distance features matrix.Split the hand cards to get all the different combinations, and get the number of combinations that can finish the hand cards the fastest.

C. Perfect Information Distillation

In the bidding stage of DouDizhu, the imperfect information attribute comes from the 17 hand cards by the other two players and three hole cards. Therefore, the key challenge for the agent is to compare the strength of its own hand cards with that of other players, and at the same time, it is necessary to consider the impact of the hole cards on its own hand after get the landlord.For such a game, we can construct a perfect information game with exactly the same strategy, that is, having the hand cards information of all players and hole cards information at the beginning, there may be more chances to win the game in the end, just like having a cheating software. This allows us to use this perfect information game to train the subject of the imperfect information game, so this paper proposes to use the perfect information distillation technology to optimize the bidding method of DouDizhu.

For perfect information distillation, this paper sets a perfect reward function to improve the learning efficiency of the agent for bidding methods, as shown in formula (3):

$$R_t = AP_{all} - AP_{own} \tag{3}$$

where AP is a reward and punishment value of the relationship between the bidding score, the role and the result, as shown in TableII. Feedback to the corresponding positive and negative feedback to the intelligent agent according to the bidding score of the intelligent agent, the final role and the result. AP_{all}

is the feedback integral under the perfect information game, AP_{own} is the feedback integral under the imperfect information game, calculates the difference R_t between the two, and feeds it back to the agent under the imperfect information game as a reward mechanism, and then continuously optimizes the bidding strategy.

TABLE II
FEEDBACK REWARD VALUE OF DIFFERENT BIDDING SCORES, ROLES AND RESULTS.

Role	Landlord			Peasant		
Bid	1	2	3	0	1	2
Win	+1	+2	+3	+3	+2	0
Lose	-1	-2	-3	0	-1	-2

D. Bidding Network Model

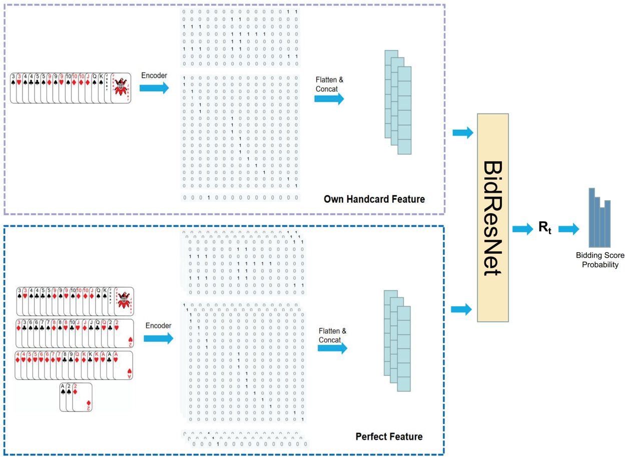


Fig. 3. Bidding strategy overall framework

The overall framework of the bidding strategy in this paper is shown in Figure3. It encodes its own hand cards information and perfect information with strength features, sequence features, and winning distance features, and combines all the features, and then sends them to the network model respectively to obtain prediction of bidding score and AP value under perfect information game and perfect information game. Finally, perfect information distillation is used to optimize the bidding strategy and improve the performance of the bidding method. The proposed convolutional neural network model is inspired by the ResNet18 model and named BidResNet. The model input is the combined feature encoding matrix, which enters a convolution layer with a convolution kernel of 7 and a maximum pooling layer. Then enter 4 modules composed of residual blocks, each of which uses 2 residual blocks with the same output channel, and introduces an additional 1×1 convolutional layer to transform the input into the feature size required in this paper. Do the addition operation. Finally, it enters a layer of average pooling layer and fully connected layer, and obtains the final bidding score prediction after Softmax normalization processing. All activation functions of the network model use ReLU.

This paper uses the Cross Entropy Loss function as the objective function of the neural network gradient descent:

$$\mathrm{L} = -[\mathrm{y}\log\hat{\mathrm{y}} + (1-\mathrm{y})\log(1-\hat{\mathrm{y}})] \tag{4}$$

in order to solve the problem of network overfitting, regularization is introduced. The main reason for overfitting is that the model is complex, there are too many parameters, and there are few training samples. The basic idea of regularization is to introduce penalty terms to reduce the magnitude of parameters. Therefore, the loss function is:

$$\mathrm{L} = -[\mathrm{y}\log\hat{\mathrm{y}} + (1-\mathrm{y})\log(1-\hat{\mathrm{y}})] + \mathrm{c}\,\|\boldsymbol{\theta}\|^2 \tag{5}$$

where $\mathrm{c}\,\|\boldsymbol{\theta}\|^2$ is the regular term of all trainable parameters, and the training effect of the model is verified by ten-fold cross-validation. And the training process is only applied to the training set, and the result of the final trained model on the test set is used as the evaluation index of the model in this paper. During training, the model uses the SGD gradient descent method. The hyperparameters used for training are as follows: learning rate = 2e-6, learning rate decay = 0.001, momentum = 0.9, batch size = 512.

IV. EXPERIMENT

A. Setup

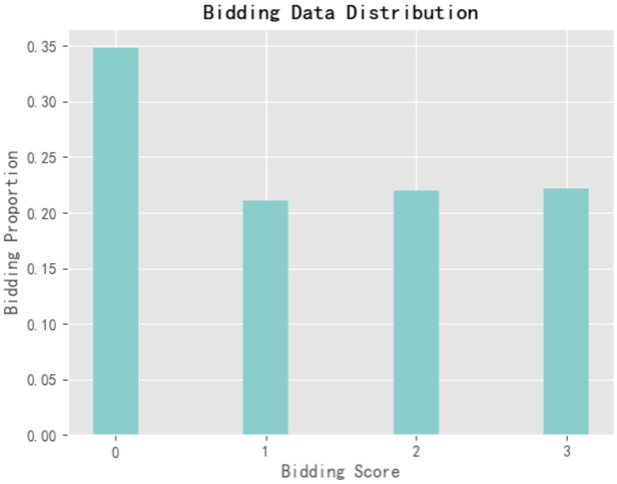
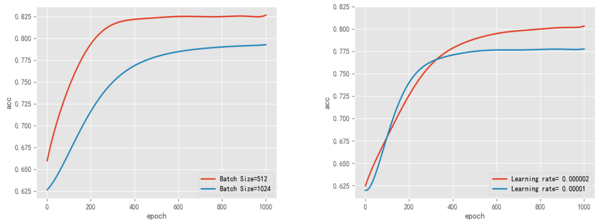


Fig. 4. Distribution of bidding score

The training of the bidding method in this paper includes two stages, namely the supervised learning stage and the reinforcement learning stage using perfect information distillation. The dataset used in the supervised learning phase comes from the training data generated by the agent self-play using heuristics. There are a total of 36,000 bidding points, of which 0 points account for a large proportion, and the remaining bidding points are evenly distributed. Figure 4 shows our game data distribution.

And establish a comparative experiment, compare the model in this paper with the existing DouDizhu bidding model, and compare our bidding model with some open source game AI after adding the DouDizhu agent.

By comparing different learning rates and batch sizes during training, as shown in Figure 5, the learning rate of the final selection model is 2e-6, the batch size is 512, and the total training rounds are 1000 times. The operating environment of all experiments is Intel i5-9500 CPU, 16GB memory, an Nvidia RTX2080 SUPER as GPU, and Ubuntu 20.04.4 LTS as the operating system server. The deep learning framework uses Pytorch (version 1.13.0), and all experiments are implemented in Python (version 3.8).



(a) Batch Size

(b) Learning Rate

Fig. 5. Comparison of training accuracy in different batch sizes and learning rates.

B. Comparison with Existing Bidding Programs

We use the following DouDizhu bidding model as baseline:

Text Convolutional Neural Network (Text CNN): It innovatively applies convolution to the processing of serialized information, realizes convolution operation and feature extraction under different lengths, and has fewer parameters and faster training process than RNN (Recurrent Neural Network).

Long and short term memory network (LSTM): It is a specific structure of recurrent neural network. It has the advantage of being able to transfer sequence data between networks for both long-term and short-term, enabling sequence-based supervised learning tasks.

Support vector machine (SVM): It is a common machine learning model widely used in prediction and classification tasks with few hyperparameters, strong generalization and interpretability.

SeqStgCNN: An improved CNN-based bidding recommendation model for Doudizhu, which is currently the best bidding recommendation model.

The performance comparison between the proposed model and the baseline model is shown in Table III and Figure 6. It can be seen from the results that the model in this paper has the best prediction performance in the three indicators of accuracy, recall and F1. Among them, the poor performance of SVM indicates that traditional machine learning methods may have problems such as overfitting when dealing with massive data such as the game of DouDizhu, resulting in poor results. In the deep learning model, the performance of LSTM and TextCNN is obviously due to the traditional machine learning method. Compared with the ordinary single deep learning model, the improved SeqStgCNN based on CNN also has a good performance improvement, but the performance of these models is weaker than model proposed in this paper. It shows that this model can effectively extract the hand information in DouDizhu bidding, so as to evaluate the strength of the hand, and finally achieve more accurate prediction of the bidding score.

TABLE III
COMPARISON OF PERFORMANCE INDICATORS OF EACH MODEL

Model	Accuracy	Recall	F1
BidResNet(Ours)	**0.8189**	**0.8107**	**0.8148**
SeqStgCNN	0.8005	0.7980	0.8000
TextCNN	0.7824	0.7814	0.7822
LSTM	0.7721	0.7718	0.7720
SVM	0.6049	0.6035	0.6044

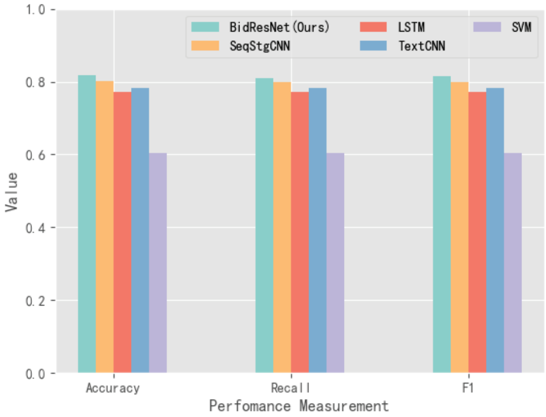


Fig. 6. Three classification indexes of the model in the bidding task of DouDizhu

C. Comparison with Existing AI Programs

We use the following Fight the Landlord AI program as our baseline:

XDou: A powerful AI program. It integrates complex heuristics and PIMC algorithms. We compared it on Botzone.[21]

CQN: Based on two stages reinforcement learning[22]. We used the pre-trained model provided by them for comparison.

The metrics in the experiment are as follows:

- **WP (Winning Percentage):** The number of the games won by A divided by the total number of games.
- **ADP (Average Difference in Points):** The average difference of points scored per game between A and B. Each bomb or special card type will double the score.

We play 100 games with each baseline program. Table IV shows the results of matches with different DouDizhu AI programs with and without BidResNet added.

TABLE IV
COMPARISON RESULT WITH THE BASELINE AI PROGRAM. THE COMPARISON RESULTS OF WITH BIDRESNET AND WITHOUT BIDRESNET.

	WP	ADP
With BidResNet vs XDou	**0.69**	**+0.81**
Without BidResNet vs XDou	0.66	+0.59
With BidResNet vs CQN	**0.80**	**+1.31**
Without BidResNet vs CQN	0.75	+1.04

The results of the game against the public Doudizhu AI show that the performance of our proposed BidResNet is excellent. In 100 games, BidResNet can provide more accurate bidding score recommendations for the agent. In the games against XDou and CQN, after adding BidResNet, WP increased by 3% and 5% respectively, and ADP increased by 22 and 27 in 100 rounds respectively. It is not difficult to see that the improvement of BidResNet on ADP is obvious, which proves its accuracy in bidding strategy of the agent and refinement of bidding score prediction.

V. CONCLUSION AND FUTURE WORK

This paper proposes a new method for training bidding in DouDizhu agent, combine the convolutional neural network with reinforcement learning, optimize the agent's bidding strategy by using perfect information distillation, and increase the winning distance in the hand cards feature. This allows agent to take the lead in the bidding stage.In the experiment, we found that BidResNet will make a high score bid when it has a rocket or the hand cards type is neat, which will lead to losing the game. This may be because our weight on rockets and winning distance is too high in the training process, which leads to agent choosing a high score to compete for the landlord and losing the game. In the future work, we will reduce the weight of rockets and winning distance, and combine card type combination with card strength to extract strength features, so that agent can bid more cautiously.

REFERENCES

[1] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search. " Nature 529. 7587 (2016): 484-489.

[2] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledgeJ]. Nature, 2017, 550(7676):354-359.

[3] Zinkevich M A , Johanson M , Bowling M H , et al. Regret Minimization in Games with Incomplete Information[C]// Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007. Curran Associates Inc. 2007.

[4] Schmid, Martin, Lisy, et al. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker[J]. Science, 2017.

[5] Brown N , Sandholm T . Superhuman AI for heads-up no-limit poker: Libratus beats top professionals[J]. Science, 2017:eaao1733. W János, V István. Solving Renju[J]. Icga Journal, 2001, 24(1):30–34.

[6] Vinyals, Oriol et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature (2019): 1-5.

[7] Berner C , Brockman G , Chan B , et al. Dota 2 with Large Scale Deep Reinforcement Learning:, 10. 48550/arXiv. 1912. 06680[P]. 2019.

[8] Deheng Ye et al. "Mastering Complex Control in MOBA Games with Deep Reinforcement Learning" national conference on artificial intelligence(2019): n. pag

[9] Li J , Koyamada S , Ye Q , et al. Suphx: Mastering Mahjong with Deep Reinforcement Learning[J]. 2020. Mao Limin, Zhu Peiyi, Lu Zhenli, et al. Gobang game algorithm based on LabVIEW [J]. Computer applications, 2016, 36 (6): 1630-1633

[10] You Y , Li L , Guo B , et al. Combinational Q-Learning for Dou Di Zhu[J]. 2019

[11] Jiang Q , Li K , Du B , et al. DeltaDou: Expert-level Doudizhu AI through Self-play[C] Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19. 2019

[12] Zha D , Xie J , Ma W , et al. DouZero: Mastering DouDizhu with Self-Play Deep Reinforcement Learning[J]. 2021.

[13] Guan YangMinghuan Liuet al. PerfectDou:Dominating DouDizhu with Perfect Information Distillation[J]. 2022

[14] Li S , Meng D , Peng L , et al. Design and implementation of an adaboost-based landlord bidding strategy[C]// Control Decision Conference. IEEE, 2017.

[15] Yuan B , Li S . Recommending Bids on Dou-DiZhu Poker Games: A Deep Learning Approach[C]// 2020 Chinese Automation Congress (CAC). 2020.

[16] Heinrich J , Silver D . Deep Reinforcement Learning from Self-Play in Imperfect-Information Games:, 10. 48550/arXiv. 1603. 01121[P]. 2016.

[17] Lanctot M , Zambaldi V , Gruslys A , et al. A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning[C]// 31st Conference on Neural Information Processing Systems (NIPS), 4-9 December 2017, Long Beach, CA, USA. 2017.

[18] Brown N , Bakhtin A , Lerer A , et al. Combining Deep Reinforcement Learning and Search for Imperfect-Information Games:, 10. 48550/arXiv. 2007. 13544[P]. 2020.

[19] Volodymyr, Mnih, Koray,et al. Human-level control through deep reinforcement learning. [J]. Nature, 2015.

[20] Mnih V , Badia A P , Mirza M , et al. Asynchronous Methods for Deep Reinforcement Learning:, 10. 48550/arXiv. 1602. 01783[P]. 2016

[21] "Botzone", https://en. botzone. org. cn/, 2022.

[22] "CQN", https://github. com/qq456cvb/doudizhu-C, 2019.