The Prediction of the Best Contract in Non-competitive Bridge Bidding

Ying-Yu Lin[a], Chang-Biau Yang[a] * and Kuo-Si Huang[b]
[a]Department of Computer Science and Engineering
National Sun Yat-sen University, Kaohsiung, Taiwan

[b]Department of Business Computing
National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

Abstract

In a bridge game, bidding is a crucial aspect in determining the outcome, and the choice of the final contract often deeply influences it. Previous studies failed to provide meaningful information because they did not reveal the hidden hand information implied by the convention card. To overcome this issue, we adopt the implications of auctions from the convention card and label the data as features to make decisions. We employ a two-stage model, built by two random forest models, to predict the most suitable contract. The dataset comes from the Bridge Base Online (BBO). For a fixed North-South hand, we shuffle 100 different decks where the East-West side does not bid. Then, we analyze the double-dummy results to determine which achieves the highest score. We define seven indicators to assess the meaning of different contracts in the board. Among them, the same category (SC) provides a more accurate representation of the best contract. Finally, our random forest model achieves an accuracy of 0.773 for predicting the same category contract and wins 0.212 IMPs per board against the BBO system.

Keywords: contract bridge, bidding, convention card, double-dummy solver, random forest

1 Introduction

Game playing is one of the essential fields for artificial intelligence research. These games can be divided into two categories according to the degree of information exposure. A perfect information game means that all players know everyone's game state and its continuation information at any stage of the game. For example, in chess and Go, both players can see all the information in each situation and are therefore called perfect information games. Otherwise, it is an imperfect information game such as bridge. In a game with imperfect information, although each player can see the cards that have been played by the opponent, they cannot fully grasp the remaining cards in the opponent's hand.

Deep Blue, developed by IBM in 1997, became famous when the computer defeated the world chess champion, Garry Kimovich Kasparov. In 2016, AlphaGo, developed by Google DeepMind, defeated the world Go champion, Lee Sedol, by 4:1. In 2017, the new generation AlphaGo Zero beat AlphaGo by 100:0. Despite the high complexity, artificial intelligence has almost overcome the problems of perfect information games.

For the bridge game, computer players can play the double-dummy bridge [6, 8–10], where the four hands of the four players are assumed to be known. However, except the hand of the dummy, the other two hands cannot be seen in the actual bridge game. There remains a need for computer bridge software capable of beating top human bridge players.

From 2006 to 2019, there have been several studies trying to train bidding strategies, and the research on bidding by using machine learning is shown in Table 1. However, all of the previous studies suffer from a fatal flaw: the conclusions of these decision-making methods are not meaningfully informative. That is, the bidding which follows the convention card has not yet revealed the implicit hand information. Failure to accurately

*Corresponding author. E-mail: cbyang@cse.nsysu.edu.tw (Chang-Biau Yang).

Table 1: The bidding models for the bridge game with machine learning.

Year	Authors	Note
2006 [1]	Amit and Markovitch	Use the PIDM to estimate decisions that partners will make
2007 [2]	DeLooze and Downey	Use the SOM to solve the non-competitive bidding sequence
2015 [4]	Ho and Lin	Use the UCB to solve the non-competitive bidding sequence
2018 [12]	Yeh *et al.*	Use the Deep RL to solve the non-competitive bidding sequence
2018 [7]	Legras *et al.*	Use the ILP method to train whether to call an opening bid
2019 [11]	Rong *et al.*	Use the ENN to estimate the cards in the partner's hand and the PNN to take actions in the competitive bidding
2023	This paper	Extract the hidden hand features embedded in the bidding history, and use two-stage random forest models to predict the final contract

estimate hand information may lead to poor final contracts. Therefore, we aim to apply machine learning to train a bidding model for predicting a better contract. This involves referring to the meaning in the convention card, labeling the hand information to aid in deciding the contract, and finally finding the most suitable contract in the bridge game.

In this paper, we extract the experimental dataset from the Bridge Base Online (BBO). For each fixed North-South (NS) hand, we randomly shuffle 100 different East-West (EW) hands, and then analyze the results of the double-dummy solver (DDS) to obtain the most suitable contract. We use two random forest models to predict the suits and game categories, respectively. Our model achieves an accuracy of 0.773 for predicting the same category contract. Our model also outperformed the BBO system by 0.212 IMP per board.

This paper is organized as follows. Section 3 explains the data preprocessing and Section 4 defines the most suitable final contract. Section 5 presents the performance of our proposed model and its predicted results. Finally, Section 6 presents our conclusion.

2 The Main Procedure

The goal of our algorithm is to predict the most suitable contract. Our algorithm consists of the following main stages.

Stage 1: Dataset preparation and preprocessing.

The BBO bidding system [3] has already completed the bidding process in the "Just Declare". The hand information provided includes the hands of the four players, vulnerability, dealer, and bidding history. The bidding history may contain some implied meanings of each bid. We extract approximately 107,000 instances from BBO Solitaire's "Just Declare." By examining the hand information, we delete some hands with unreasonable bidding sequence or interpretation.

Stage 2: Feature extraction of each board

Based on the bidding history, we can extract hand strength, distribution, balanced labels, and stoppers. With our preliminary analysis, the accuracy of HCP and suit length description is above 90%, which is acceptable. But, BBO's descriptions for "balanced" and "stopper" are mostly "unknown," which will be enhanced by our method, explained later. The detailed features include HCP for each suit of the player, the range of total suit lengths for each suit of the player and partner, balanced labels for the player and partner individually, total stopper conditions for each suit of the player and partner, vulnerability for the board, and the bidding history represented in 72-bit one-hot encoding.

Step 3: The definition of the most suitable contract.

For each fixed NS hand, we randomly shuffle 100 different decks for the EW side, where they do not bid. The contract that achieves the highest score most times in the double-dummy results is set as our prediction target. Then, the most suitable contract is further adjusted according to the bidding history. For example, when the opener bids a major suit and the responder bids a Splinter, it clearly tends to play the major suit contract. On the other hand, when the opener bids a no-trump contract and the responder bids a Stayman, the final contract should be a major suit or a no-trump contract. Additionally, when the opener bids a minor suit and the responder supports the minor suit, but the minor suit contract needs to win more tricks, the final contract may be a no-trump contract, instead of a minor suit contract.

Step 4: Two-stage contract prediction with the random forest.

We adopt two random forests independently to predict the most suitable contract: (1) predict the suit (or NT) that the board should choose; (2) predict the game category (partial game, game, slam, or grand slam) with adjusting the suit if it does not comply with the game rules. By combining the suit and the number of winning tricks, we obtain the final predicted contract.

Step 5: Performance calculation with IMP scores.

Based on the prediction target's game category and suit, we define seven indicators, explained later. Finally, we calculate the scores for each set of 100 hands and compare them to BBO, then convert the score difference into IMPs.

3 Enhancement of Feature Extraction

Based on the bidding history, the accuracy of HCP and suit length description is above 90%, which is acceptably high. However, "balanced" and "stopper" are mostly labeled as "unknown," in BBO description, which is unacceptable for prediction. Therefore, with the help of bridge experts, we extract the features of frequently occurring bidding sequences according to the BBO convention card. The meanings of these two labels are given as follows.

- **balanced**. In one hand, the distribution of the four suits as 4333, 4432, 5332, and 5422 is regarded by balanced The remaining distributions are classified as unbalanced. For more precisely, we set the values for balance as follows. 0: balanced; 1: likely balanced; 2: likely unbalanced; 3: unbalanced.

- **stopper**. According to the stopper strength of the suit, it is divided into four grades: 0: no-stopper (such as XXX); 1: unknown; 2: partial-stopper (such as QXX); 3: stopper (such as AJX). A suit with a stopper means that it can prevent at least one attack by the opponents.

W	N	E	S
P	P	P	1N
P	2♢	P	2♡
P	3♣	P	?

Bids	P	1♣	1♢	1♡	1♠	1N	2♣	2♢	2♡	2♠	2N	3♣	3♢	3♡	3♠	3N	...	7N
North	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	...	0
South	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	...	0

Figure 1: An example of 72-bit encoding for the bidding history. Here, the bidding sequence of "North" is `Pass`, 2♢, and 3♣, and 1N and 2♡ for "South".

For expressing the bidding history, since the EW players only choose to pass, there will be no double and redouble situations, so it is only necessary to record the bidding history of the NS players. We can record the bidding sequence with 36 possible contracts, including `P` (`Pass`), 1♣, 1♢, 1♡, 1♠, 1N (no-trump), $\cdots$, and 7N. Figure 1 shows the encoding scheme for a bidding sequence. Table 2 shows all features for representing the hand information of North-South side.

Table 2: The features for the hand information of NS side.

Feature	Description
player_hcp_club player_hcp_diamond player_hcp_heart player_hcp_spade	The player's high card points (HCP) in clubs, diamonds, hearts, spades, range from 0 to 10.
total_hcp_lower total_hcp_upper	The lower/upper limit of total HCP, range from 0 to 40.
total_num_club_lower total_num_club_upper total_num_diamond_lower total_num_diamond_upper total_num_heart_lower total_num_heart_upper total_num_spade_lower total_num_spade_upper	The lower/upper limit of the total number in the four suits, ranges from 0 to 13.
player_balance partner_balance	Balanced label of the player/partner's. 0: balanced; 1: likely balanced; 2: likely unbalanced; 3: unbalanced.
total_stop_club total_stop_diamond total_stop_heart total_stop_spade	The total stopper labels for the four suits. 0: no stopper; 1: unknown; 2: partial-stopper; 3: stopper.
vulnerability	1: Neither side vulnerable; 2: NS vulnerable; 3: EW vulnerable; 4: Both sides vulnerable.
bidding_history	The bidding history sequence, excluding the last bid, encoded by 72 bits.

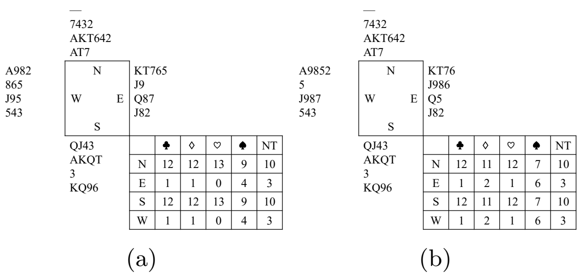

Figure 2: Two examples of different DDS results with the same NS hand, but different EW hand.

4 The Prediction of the Best Contract

Even if the NS hand is the same, the number of winning tricks will be affected by the distribution of the EW hands and the position of the honor cards. Refer to Figure 2 for an example that the same NS hand, paired with different EW hands, yields different trick results. For each fixed NS hand, we shuffle 100 different EW hands to minimize the influence caused by good or bad luck.

We record the contract with the highest score obtained from the DDS. Additionally, contracts that fall within 80 points of the highest score (that is to say, belong to the same type in the four categories of partial game, game, slam, and grand slam) will also be considered as the best contracts for this game. After counting the best contracts of 100 hands, we can obtain the best contracts of this North-South hand according to the following rules.

Rule 1: Select the contract with the largest number of deals.

Rule 2: If there is more than one contract, choose the better one from the total counts of suits.

Rule 3: If the total counts of suits are equal, select a superior game category.

Rule 4: If there are still two or more contracts, first choose by suit order: major suit, no-trump, minor suit; and then by category order: grand slam, slam, game, partial game.

The DDS treats the problem as complete information, but there is hidden information in real bridge games. Consequently, we will adjust the final contract based on the suits determined during the bidding process.

After deciding the most suitable contract, we can classify all situations into seven indicators as follows. We use $\mathbb{U}$ to denote the set of all contracts.

- **Most Suitable** (MS). The highest score most frequently appears in the DDS results, and the most suitable contract is adjusted by the bidding sequence.
- **Same Category** (SC). All contracts that have a similar high score to MS are set in SC.
- **Same Suit** (SS). All contracts that belong to the same suit in MS are set in SS.
- **Same Category and Acceptable Suit Excluding Most Suitable** (SCA). (SCA) consists of the contracts in SC, excluding MS, and the contracts should achieve the highest score at least once. $SCA = SC \setminus (MS \cup SCU)$.
- **Same Category and Unacceptable Suit** (SCU). SCU consists of the contracts in SC, but these contracts have never achieved the highest score. $SCU = SC \setminus (MS \cup SCA)$.
- **Same Suit Excluding Most Suitable** (SSE) SSE consists of the contracts in SC, excluding MS. $SSE = SS \setminus MS$.
- **Others** O. O consists of the contracts that belong to different game categories and suits from MS. $O = \mathbb{U} \setminus (SC \cup SS)$.

To predict the most suitable contract, we first employ one random forest to predict the suit (or NT), and then utilize another random forest to predict the game category. Just as in an actual bridge game, we will communicate first about which suits the partnership's hands are suitable for or whether choosing a no-trump contract is appropriate. Then we consider the game category to be chosen based on the strength of both hands in that suit. The combined prediction result is viewed as our final predicted contract.

5 Experimental Results

Our experimental dataset is taken from the "Just Declare", a solitaire bridge game of *Bridge Base Online* (BBO) [3]. All BBO bidding histories follow the SAYC convention card with slight modification by BBO. To prevent the interference of some noisy data, we first delete some unreasonable instances, resulting in a dataset of 107,660 instances.

Figure 3 illustrates the distribution of the proportions of the seven indicators becoming the highest scoring on a board. It can be observed that the proportion of MS becoming the highest score is not so significant. Considering that there may

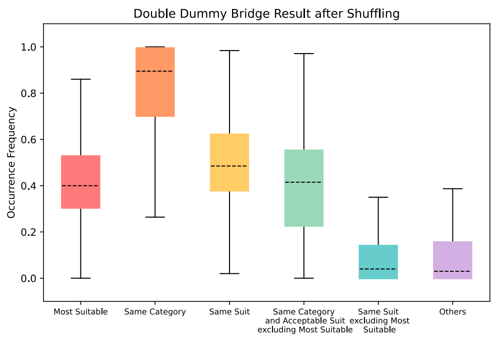

Figure 3: The boxplots of each indicator.

Table 3: The performance comparison of the random forest, the neural network with raw data and the neural network with feature extraction.

indicator	random forest	NN with raw data	NN with feat. ext.
most suitable (MS)	0.583	0.472	0.549
same category and acceptable suit excluding most suitable (SCA)	0.177	0.189	0.175
same category and unacceptable suit (SCU)	0.012	0.028	0.010
same suit excluding most suitable (SSE)	0.140	0.199	0.178
others (O)	0.087	0.113	0.088
total ($\mathbb{U}$)	1.000	1.000	1.000
same suit (SS)	0.723	0.671	0.727
same category (SC)	**0.772**	0.689	0.734

not be a unique optimal contract in bridge and the score difference between different contracts in the same game category is not greater than 80 points, the differences are not large. For example, the score difference between 7C and 7N is 80 points, while 7C and 7N fall within the same category (grand slam). Therefore, we propose the term SC, providing a more accurate representation of the best contract. This representation enhances the accuracy of our approach.

In the preliminary experiment, a neural network model (eight layers of fully connected layers) [11] is used as a baseline to test two kinds of input data separately. The first kind of input data involves encoding the player's cards and the bidding history with one-hot encoding. The second kind involves the extracted features.

Table 3 displays the performance of the random forest, the neural network with raw data, and the neural network with feature extraction. We use the macro-average method to calculate Precision, Recall, and F1. It can be observed that our forest model tends to outperform both neural network models. Thus, we choose the random forest model, rather than the neural network.

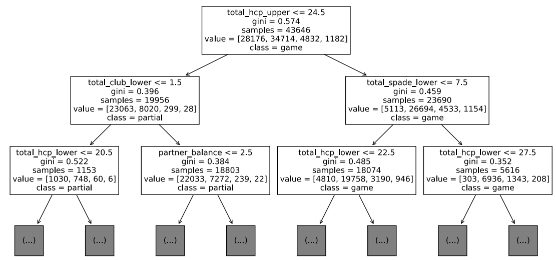

Figure 4: An example of the decision tree generated by the random forest model.

Table 4: Proportion comparison of indicators between BBO and our predicted contract in ten experiments.

indicator	BBO		our model	
	μ	σ	μ	σ
most suitable (MS)	0.568	0.000	0.583	0.001
same category and acceptable suit excluding most suitable (SCA)	0.197	0.000	0.178	0.001
same category and unacceptable suit (SCU)	0.001	0.000	0.012	0.000
same suit excluding most suitable (SSE)	0.150	0.000	0.140	0.001
others (O)	0.084	0.000	0.088	0.000
total ($\mathbb{U}$)	1.000	0.000	1.000	0.000
same suit (SS)	0.718	0.000	0.718	0.001
same category (SC)	0.766	0.000	**0.773**	0.001

The random forest model consists of multiple decision trees, where the nodes are formed by randomly selected features, and the outcome is determined by the majority class output by individual trees. Figure 4 illustrates one example of the decision tree generated during the training process.

In our main experiment, 5-fold cross-validation is performed ten times on the two-stage random forest models. Each performance indicator is calculated by the average of the ten experiments. Table 4 shows the indicator performance of the final prediction results, compared to BBO. Although our model prediction shows little differences from BBO's in the "same category" (SC), it is closer to the "most suitable" (MS) compared to BBO.

To minimize the influence of extreme values on the overall score in real-world bridge games, the point difference of the same board on two different tables is converted into the *International Match Points* (IMP), an integer ranging from 0 to 24. The IMP value is used to measure the performance of two teams. In Table 5, if we consider a fixed NS hand and compare it only once, as in the case of games on the BBO website, the predicted results of our model outperform the final contract of BBO by 0.088 IMP per board. If we fix the NS hand and shuffle 100 EW hands, it improves to 0.212 IMP per board.

Table 5: The IMP gain of our random forest model compared to BBO, averaged by ten runs.

model	μ	σ
playing one board with BBO	0.088 IMP	0.007
playing 100 boards with BBO	0.212 IMP	0.003

6 Conclusion

In today's bridge games, the differences in card-playing skill between experts are not so significant. The key to winning a bridge game is to bid for the best contract based on the combined strength of the two players' hands. In this paper, we define the most suitable contract and utilize a two-stage prediction with random forest models to predict the suit and game category. We incorporate the concept of convention card as a feature in our model, allowing the results to achieve high scores and also consistent with the bidding logic commonly used by bridge players. In other words, our most suitable contract represents a safe contract that can achieve high scores within the capabilities specified by the convention card.

We define seven indicators to assess the meaning of different contracts in the board, and we find that the proportion of the most suitable contract becoming a high-score contract is not significant. Considering that there may not be a unique optimal contract in bridge and the score difference between different contracts in the same game category is not large, we believe that the same category (SC) is a more accurate representation of the best contract. Our model achieves an accuracy of 0.773 in prediction of the same category contract. Moreover, our model achieves an average win of 0.212 IMP per board against the BBO system.

In the future, we may consider employing the hand strength evalution method proposed by Jan *et al.* [5] to enhance the prediction result.

References

[1] A. Amit and S. Markovitch, "Learning to bid in bridge," *Machine Learning*, Vol. 63, No. 3, pp. 287–327, 2006.

[2] L. L. DeLooze and J. Downey, "Bridge bidding with imperfect information," *2007 IEEE Symposium on Computational Intelligence and Games*, pp. 368–373, 2007.

[3] F. Gitelman, "Bridge Base Oline." https://www.bridgebase.com/index.php.

[4] C.-Y. Ho and H.-T. Lin, "Contract bridge bidding by learning," *Proceedings of the 2015 AAAI Workshop*, Hyatt Regency in Austin, Texas, USA, pp. 1–7, 2015.

[5] J.-P. Jan, K.-S. Huang, and C.-B. Yang, "Hand strength evaluation in bridge games with the evolutionary algorithms," *Proceedings of the 40th Workshop on Combinatorial Mathematics and Computation Theory*, Taoyuan, Taiwan, pp. 50–55, 2023.

[6] S. Kowalik and J. Mańdziuk, "Towards human-level performance in solving double dummy bridge problem," *Proceedings of the Neural Information Processing - 28th International Conference*, pp. 15–27, Sanur, Bali, Indonesia, 2021.

[7] S. Legras, C. Rouveirol, and V. Ventos, "The game of bridge: A challenge for ILP," *Proceedings of the International Conference on Inductive Logic Programming*, pp. 72–87, 2018.

[8] J. Mańdziuk and J. Suchan, "Solving the double dummy bridge problem with shallow autoencoders," *Proceedings of the Neural Information Processing - 25th International Conference*, pp. 268–280, Siem Reap, Cambodia, 2018.

[9] J. Mańdziuk and J. Suchan, "Who should bid higher, NS or WE, in a given bridge deal?," *Proceedings of the International Joint Conference on Neural Networks*, pp. 1–8, Budapest, Hungary, 2019.

[10] K. Mossakowski and J. Mańdziuk, "Learning without human expertise: A case study of the double dummy bridge problem," *IEEE Transactions on Neural Networks*, Vol. 20, No. 2, pp. 278–299, 2009.

[11] J. Rong, T. Qin, and B. An, "Competitive bridge bidding with deep neural networks," *Proceedings of the International Conference On Autonomous Agents and Multi-Agent Systems*, pp. 1–9, 2019.

[12] C.-K. Yeh, C.-Y. Hsieh, and H.-T. Lin, "Automatic bridge bidding using deep reinforcement learning," *IEEE Transactions on Games*, Vol. 10, No. 4, pp. 365–377, 2018.