Blackdice

Introduction

We designed a game named Blackdice which is similar to Blackjack with the following two main differences:

- If a player's hand is less than 12, then player is required to draw another card.
- If a player's hand is greater than or equal to 12, then player can decide whether to roll a die. If the player decides to roll a die, then the value from the die is added to the player's hand. Player can roll a die multiple times until busted.

The rest of the Blackdice is the same as the standard Blackjack including the dealer hits on soft 17.

We came up with 6 candidate policies and developed a reinforcement learning algorithm (MC) to figure out the best policy of playing this game. For simplicity, we also assume there is only one player in each round and the probability of getting any card is completely random, i.e. card is not drawn from a standard deck.

Figure 1: The environment of BlackDice game.

Framework

Based on the rules above, we can frame this game in the following way:

- **Agent:** the player
- **Environment:** a deck of cards and a die
- **State:** the value in player's hand
- **Reward:** 1 for winning, -1 for losing, 0 for tie
- **Action:** stay or roll a die

Method

A model free method with Monte Carlo appears to be a logical solution because the agent would be able to learn from complete episode rather than relying on the prior knowledge.

Policies

We defined a naive kind of policy which is applying one threshold on the current sum value of cards on player's hand. The value of threshold ranges from 16 to 21. For instance, if we set the threshold as 16, it means that the agent will roll the die if the sum value is lower than 16.

Policy evaluation

We first used 3D visualization to qualitatively evaluate the policy. After 500,000 times run, we computed the winning frequency of each policy and linearly mapped the numeric value from [0, 1] to [-1, 1]. The Red and Blue colorbar was then applied which show the winning probability (in Red) and losing probability (in Blue). Visually, a warmer color of 3D surface indicates a better policy. From Fig 2 we can tell that for lower or higher threshold value, the surfaces become colder color, so the optimal should located at the medium.

Besides the qualitative evaluation, we also did a quantitative evaluation by computing the arithmetic average of the winning probabilities as shown in Fig 3. Similarly, the highest value we found was 18 which is located at the medium.

In summary, the best (one of six) policy was threshold 18, which means the agent will keep rolling the die until the value sum is equal or higher than 18. This makes sense because the expectation of rolling a die once is adding 3.5 point, in order to reach 21 as close as possible but no exceeding, the best base point should be 17.5, which is 18 after rounding.

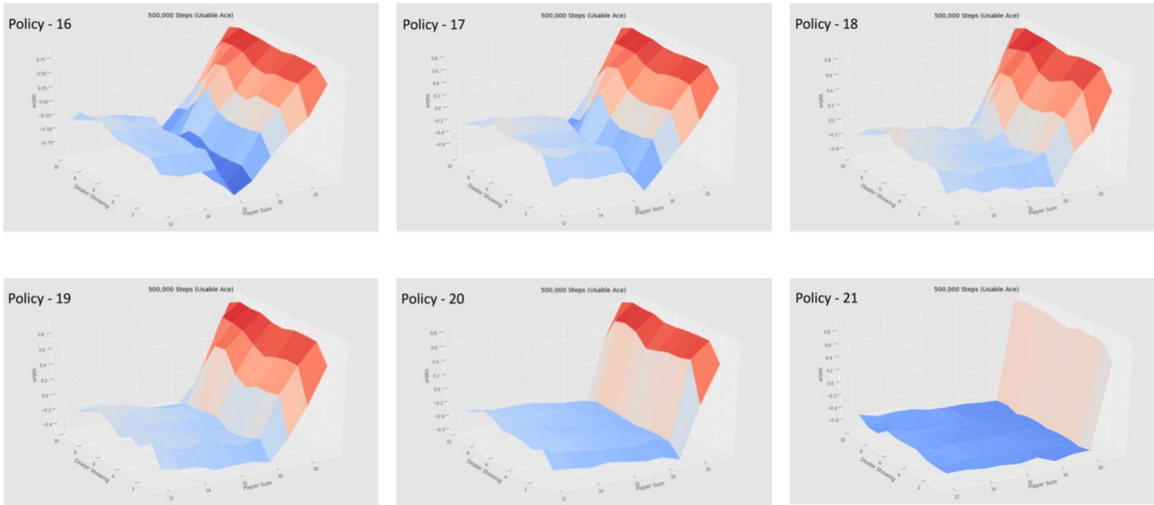

Figure 2: Qualitative evaluation of 6 policies.

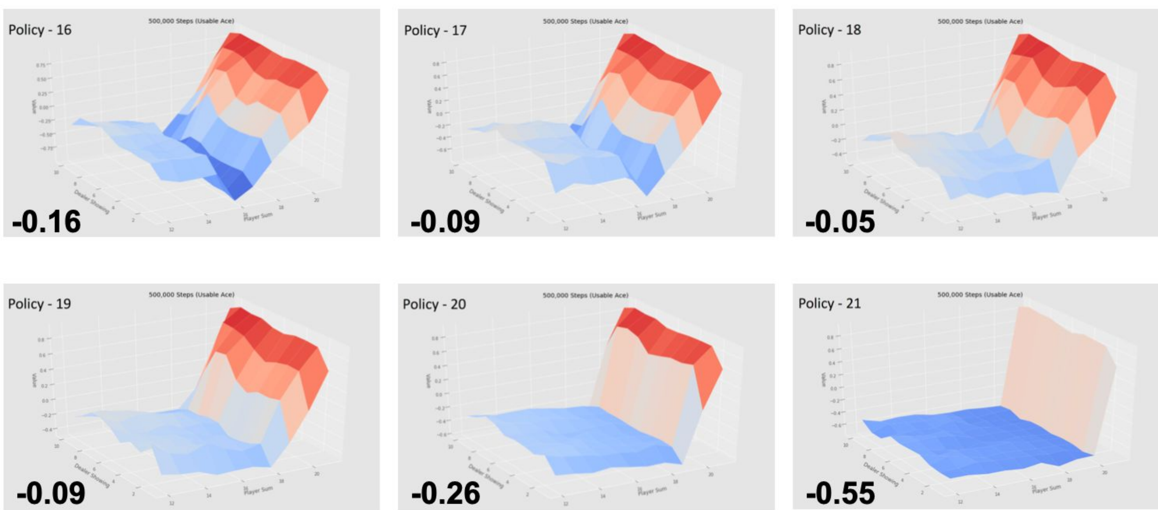

Figure 2: Quantitative evaluation of 6 policies.

Future Work

The policies we experimented are only focusing the numbers in player's hand without considering the number in dealer's hand. Some effective Blackjack strategies do have to consider the dealer's hand. Ideally, we could experiment player policies based on different value in dealer's hand.

We could also try other algorithms, such as epsilon greedy, to see if it can outperform Monte Carlo.