Blackjack Game using Reinforcement Learning Technique

Krishnapriya Krishnan Santhadevi
MAI, Faculty of Computer Science and Business Information Systems
Technical University of Applied Sciences
Würzburg-Schweinfurt ,Germany
krishnapriya.krishnansanthadevi@study.thws.de

***Abstract*—The objective of this paper explain how different strategies can be utilized to improve the performance of the player playing the Blackjack game. Basic strategy- mimic the dealer is applied to enhance the success of the player. A complete point count system is also used to analyze card values to enhance the accuracy of decision-making. Among the rule variations, the dealer stands on soft 17, and usage of multiple decks is implemented. The results show that every strategy and rule adjustment has a substantial impact on the results and that the entire point count system significantly increases the winning percentage. This paper aims to provide a thorough grasp of how to maximize Blackjack gaming by carefully analyzing various methods. It also illustrates that several strategic variations can optimally improve a player's win rate in the standard version of the game.**

***Keywords*—Blackjack, Q-learning, Card count**

I. Introduction

Blackjack is a popular casino game and it's also known as a 21-card game. The primary objective of a player is to have a hand value of atmost 21 but more than the Dealer's value. The game will have one dealer and a maximum of 7 players. At the beginning of every turn, both the dealer and the player will get two cards where one of the dealer's cards will be faced up and another will be faced down but the player's cards will be open to all [2].

The basic strategies of hitting and standing will guide the players on when to draw additional cards or hold onto their initial cards based on probable outcomes. These are the traditional approaches, but the advancements in machine learning mainly reinforcement learning introduce new strategies for optimizing the game-play. A complete point count system is implemented to enhance the player's ability to predict favorable conditions for drawing cards. By determining how many high-value and low-value (Hi-Lo) cards are still in the deck. This method evaluates various cards and keeps track on them, which enables players to make better decisions.

The paper also analyzes the impact of rule variations on game-play. Specifically, the "dealer stands on soft 17" rule, where the dealer must stand when holding a soft 17 results in benefiting the player. When "Beat the Dealer" book was published by the mathematician Edward O. Thorp, casinos started to play with more than one deck, to minimize the player's winning percentage. The use of multiple decks is a rule practiced in many casinos to maximize the complexity of card counting. The Knock-Out (KO) point count system is a simplified card-counting method used in Blackjack to help players estimate whether the remaining cards in the deck are favorable for the player or the dealer. The paper explains each method has on a player's winning percentage by implementing those strategies and analyzing rule variations.

This paper is structured into various sections, such as the Description of the Blackjack game, Related Works, Methodology and Formulations, Results, and conclusion.

II. Description of the Blackjack game

A. Cases

There are three cases in the Blackjack game, which are win, lose, and push. In the case of winning, the player will get double the bet amount, and in the case of losing, the player will lose the bet amount. The round won't be considered a win or lose if both the player and the dealer have the same point. This is known as a push case. If the player gets the sum of 21 in the initial round then it is natural Blackjack. And if it exceeds 21 then it's Busting.

B. Decisions

The possible decisions are standing, Hitting, Splitting, Double down, Surrendering, and insurance. A player has the option to choose "Stand" when they have a good hand or don't want to bust. A player can "Hit" repeatedly to get a better hand without busting. A player can separate the cards in the initial bet if they get the same face value and it's known as "Splitting". In the original cards, a player can double the original bet before the dealer gives the next card. This is known as Double Down. In surrender, the player has the option to fold a Blackjack hand before picking the

new cards. Insurance is a side bet offered to players when the dealer's up-card is an Ace. It is a strategy used by players to protect against the possibility that the dealer is holding a natural Blackjack. [6].

III. Related Works

1."Beat the Dealer: A Winning Strategy for the Game of Twenty-One" book was written by Edward O. Thorp, and first published in 1962 [5]. It is broadly acknowledged as a fundamental work in the field of Blackjack strategy and has greatly influenced casino gaming rules. Thorp introduced the concept of card counting, a strategy that allows players to keep track of the ratio of high cards to low cards remaining in the deck. Thorp's book made casinos aware that players could gain an advantage by counting cards. Consequently, casinos introduced countermeasures like using multiple decks, shuffling frequently, and barring suspected card counters."Beat the Dealer" has cemented its position as a priceless tool for anyone looking to become a Blackjack expert.

2."Reinforcement Learning: An Introduction" was authored by Richard S. Sutton and Andrew G. Barto [4]. This book provides a comprehensive overview of the ideas, methods, and applications of reinforcement learning (RL). Sutton and Barto examine a wide range of reinforcement learning algorithms, including fundamental approaches like policy gradient methods and Monte Carlo to advanced techniques like Q-learning and Temporal-Difference (TD) learning. The beginner and the experienced reader will find these algorithms understandable because of the book's detailed descriptions.

IV. Methodology and Formulations

A. Q-learning

Q-learning is a model-free reinforcement learning algorithm that focuses on learning the value of an action in a particular state. It is the process by which a player chooses actions that will enable them to maximize the total reward possible in the environment. The simplicity and effectiveness of reinforcement learning tasks are mostly proved by Q-learning.

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \tag{1}$$

In formula (1) [3], $Q(s,a)$ represents the current estimate of the value of taking action a in the state of s. The term r is the immediate reward received after action a in state s, while s' is the resulting state after action. The expression $\max_{a'} Q(s',a')$ represents the maximum estimated Q-value for the next state s' across all possible actions a', signifying the best possible future reward obtained from state s'. The parameters α and γ are the learning rates and discount factor respectively. The learning rate α, which has values between 0 and 1, determines how much new information takes precedence over the old. The significance of future rewards to immediate rewards is determined by the discount factor γ, which has a range of 0 to 1. The term $[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$ is known as temporal difference error, giving a vivid picture of the current estimate and the new estimate. The difference is done with the help of the immediate reward and the reward that is expected in the future. With every update of $Q(s,a)$ using this error, the system firstly updates the Q-values, and then goes through the optimal policy that provides the largest cumulative reward to the player.

B. Complete Point Count System

In Blackjack, the point count system is a strategic tool that keeps track of the proportion of high-value to low-value cards still in the deck. With the help of this approach, players can make informed decisions about playing and betting depending on the changing probabilities as cards are dealt. Card counting strategies can give the player a consistent advantage and significantly decrease the house edge.

The basic concept of card counting is that low cards (2 through 6) benefit the dealer, but some cards- the 10s, face cards, and Aces are advantageous to the player. Players can estimate the ratio of high cards to low cards in the deck by tracking the cards that have been played. This allows them to modify their approach accordingly.

Point count systems in Blackjack such as the Hi-Lo and KO systems give a mathematical way to get an edge over the game. These systems involve conferring point values to cards, maintaining a running count, adjusting betting, and strategy choices based on the count. The KO method streamlines the procedure by doing away with the requirement for real count computations, whereas the Hi-Lo system provides precision through true count conversion. Advanced systems like Hi-Opt, Omega II, and Zen Count offer greater accuracy but require more complexity and skill.

Blackjack can become more of a skill-based game by using these strategies to increase a player's odds of winning and change the game from a matter of chance.

The values for Hi-Lo (High-Low) System Points are explained as from 2 to 6: +1, 7 to 9: 0, and 10, Jack, Queen, King, Ace: -1. This is known as running count. The running count is divided by the expected number of decks remaining to account for the number of cards left. This true count provides a more accurate measure of the deck's favorable condition [1].

The values for the Knock-out point count system are explained as from 2 to 7: +1, 10, Jack, Queen, King, Ace: -1, 8 to 9: 0 respectively. Since the KO system employs an imbalanced count, a full deck's point total will never become zero. This makes the process easier because a true count doesn't need to be converted.

C. Rule Variation

In addition to basic strategies and card counting techniques, Thorp also explains various rule variations that can extensively impact the game's dynamics and the house edge. To maximize the player's advantage and optimize strategy, it is essential to understand these rule variations.

Dealer stands on soft 17 and Multiple Decks are the rule variations explained in this paper. In certain variants of Blackjack, the dealer must stop on a soft 17 which is any combination of cards totaling six, plus an Ace valued as 11. This is advantageous to the player because the dealer has a lower chance of improving their hand by drawing additional cards. It impacts the house edge When the dealer stands on soft 17, the house favor is slightly reduced, resulting in better odds for the player. By this rule, players should adjust their basic strategy to accommodate the dealer standing on soft 17, particularly in circumstances wherein the dealer has a poor upcard.

Initially, Blackjack was played with a single deck, though as strategies such as card counting, from Thorp's book, became more popular, casinos began to increase the number of decks, increasing the challenge for card counters and also increasing the house edge. The number of decks used to play Blackjack can vary; it usually ranges from one to eight.

The single deck provides the lowest house edge because it is simpler for players to manage their cards and make wise choices based on the remaining deck. The double deck is still advantageous for card counters but with a little higher house edge than the single deck. The four to eight decks have a significantly higher house edge because it's harder to keep track of cards, and card-counting strategies don't work either.

In multi-deck games, the effectiveness of card counting decreases because, in a single-deck game, the count's impact is more immediate and accurate, compared to multi-deck games where the dilution effect makes it more difficult to keep a precise count and switch strategies as needed. Players must adjust their play accordingly when handling multiple decks. Significant adjustments to consider include Bet sizing and Playing decisions.

V. Results

This section summarizes the results we obtained by running the players' play through a reinforcement learning algorithm. 1000 to 500000 episodes were given to the player under various techniques.

A. Basic Strategy without Q-learning

This code is very simple and it's a basic implementation of the game. First, the deck of cards is initialized and the values are given to the cards. The game then gets started by shuffling the deck and dealing two cards to the player and two cards to the dealer. The player's hand is shown, but only one of the dealer's cards is shown. The player then gets a turn to draw more cards, stay, double down, or split the hand if the two initial cards are the same. The player chooses which action to take by their input and the functions determine the action. The code then checks to see if the player has

```
Dealer's cards: 4 + hidden
Your cards: ['2', '6']
Hit(H), Stay(S), Double Down(D) or Split(P)? H
Your cards: ['2', '6', 'A']
Hit(H), Stay(S), Double Down(D) or Split(P)? H
Your cards: ['2', '6', 'A', '2']
Black Jack!
Dealer's cards: ['4', '3']
You won!!
Reward: 10
```

Fig. 1. Blackjack Environment Setting

"busted" by seeing if they have exceeded 21 or getting a Blackjack which is exactly 21. The dealer wins if the player busts. If not, it will be the dealer's turn to reveal their hidden card and draw more cards until they have at least 17 in total. The person who wins can be determined by comparing the dealer's total with the player's total. If the player's total exceeds the dealer's without busting, they win. If not, the dealer wins. The outcome of the game (win or loss) affects the reward, which is printed at the end. Figure [1] shows the reward system gives + 10 for a win and -10 for a loss.

B. Basic Strategy with Q-learning

The code implements a Q-learning agent to play Blackjack, using a strategy that mimics the dealer. It begins by defining card values and creating a shuffled deck. The BlackjackEnv class simulates the game by dealing cards, calculating hand totals, and managing player actions, including hitting, standing, and mimicking the dealer's strategy. It increases the chance of defeating the dealer and wins the game. An expectancy of —0.056 exists for the player who follows the dealer's lead by drawing to 16 or less, standing on 17 or more, and never splitting pairs or doubling down. In other words, the dealer has a 5.6% advantage. Using this technique to observe how it impacts the winning percentage. After running the algorithm 500,000 times, we obtained a win percentage of 40, which is not better than our use of the identical method alone without the rule variation. Figure [2] shows the win rate over time plot for this rule modification.

C. Complete point count system

When it comes to the card counting system's strategy, the Q-learning algorithm is employed to maximize rewards. In this strategy, we have implemented one of the most widely used and simplest counting systems in Blackjack. The Hi-Lo card counting method is a useful tool for gamblers trying to outsmart the house. Players can improve their odds of winning by modifying their betting amount and strategies by keeping a running

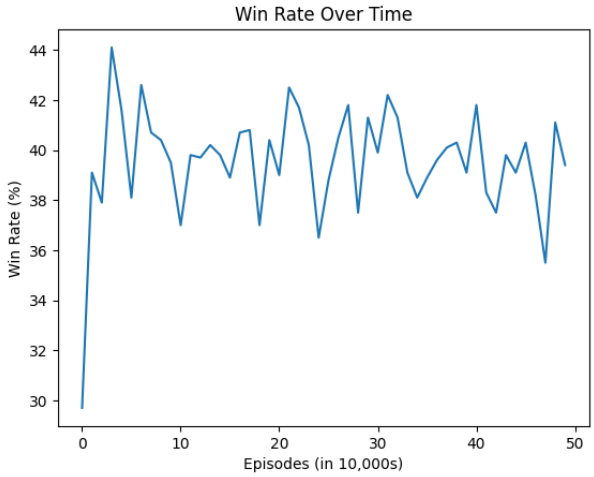

Fig. 2. The Q value optimization when the game is played 500,000 times with Mimicking the Dealer

count of high and low cards and converting this to a true count. We also employed a KO card count system to gain an edge over the house. Players can use KO more easily because it doesn't require a conversion to a true count, unlike more complicated methods. The Q-learning agent was trained with these two rule variations which resulted in the winning percentage. When trained with a normal Complete point count system we are getting 42.70% as the winning percentage. But while implementing the KO card count system the winning percentage is increased by 3% resulting in 45.30%

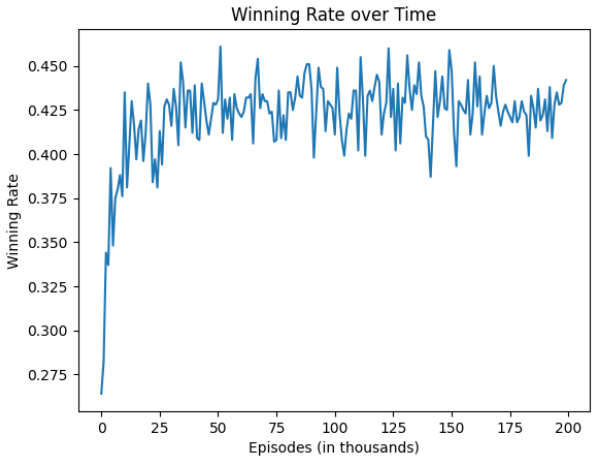

Fig. 3. Winning rate of the basic card counting system over time

Figure[3] shows that the winning rate increases over the no.of episodes, we achieved a significantly higher winning % since the algorithm is attempting to make optimal judgments based on the knowledge gained from the game.

Figure[4] shows that in the first 25,000 episodes, the winning rate rapidly increases from about 27.5% to 42.5%, where it stabilizes and thereafter varies. This

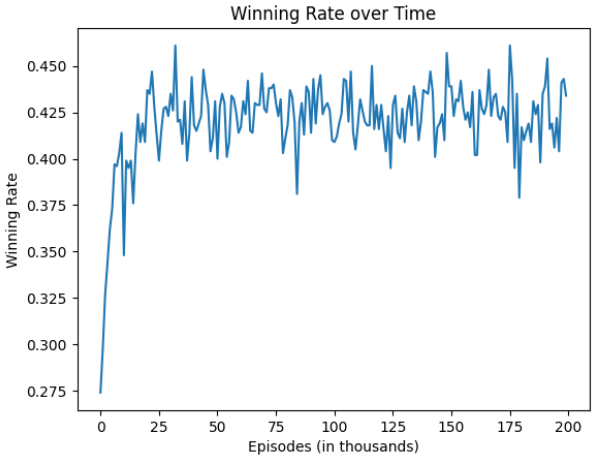

Fig. 4. Winning rate of the KO card counting system over time

demonstrates the strategy's consistency and efficacy in preserving a steady winning percentage. The variations after 75,000 episodes highlight the game's natural variability.

D. Dealer stands on soft 17

In BlackjackEnv, the step function makes sure that the dealer stands on soft 17 by verifying whether the dealer's hand sum is less than 17 or precisely 17 with a usable ace; otherwise, keep drawing. Double Q-learning Agent controls two Q-tables to reduce overestimation bias, selects actions according to an epsilon greedy policy, and updates Q-values given rewards. It runs over many episodes, decaying the exploration rate toward favoring the exploitation of learned strategies. The training function iterates through episodes, updating Q-values based on how the agent interacts with the environment. It then simulates the performance of the trained agent. Finally, it visualizes the Q-value distribution with 3D plots, showing the agent's learned values for the different game states.

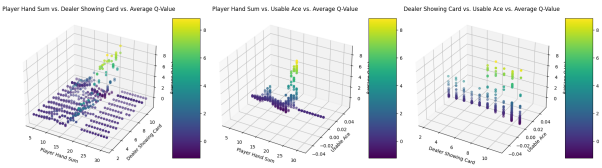

Fig. 5. Q-value optimization for different actions

In Figure[5] Player Hand Sum vs Dealer Showing Card vs Average Q-Value plot visualizes the Greater Q-values, denoted by yellow and green points, suggesting favorable conditions for the player, directing the agent's approach on whether to hit or stand. Player Hand Sum vs Usable Ace vs Average Q-Value shows where Usable aces provide flexibility in hand values, which can significantly impact the player's strategy and are reflected in the higher Q-values. And lastly, Dealer Showing Card vs Usable Ace vs Average Q-Values shows how the dealer's visible card and the

flexibility of a usable ace affect the player's expected rewards, influencing optimal actions.

E. Multiple Decks Rule Variation

The code sets up and trains a Q-learning agent for playing Blackjack using a six-deck shoe. The class BlackjackEnv simulates an environment of the Blackjack according to the rules of hitting and sticking, and the calculation of the value of a hand. The Q-learning agent class manages the Q-learning algorithm by choosing actions and updating Q values according to rewards. The function train agent will run the training loop, while simulated episodes evaluate the performance of a trained agent. Finally, plot q values can be used to visualize the learned Q values.

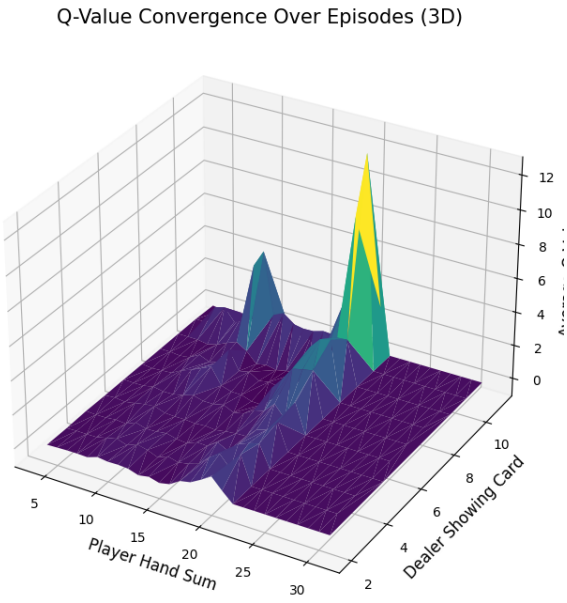

Fig. 6. Q-value convergence over episodes

Figure[5] visualizes the Q-value convergence over episodes in the trained Q-learning agent for Blackjack where the X-axis represents the player's hand sum, the Y-axis represents the dealer's showing card and the Z-axis represents the average Q-values. The peaks in the graph indicate states where the Q-values are higher, indicating that the agent has discovered these to offer greater rewards. These peaks represent the optimal actions that the agent has identified through training for various combinations of the dealer's displaying cards and the player's hand totals. This gives insight into how much the agent values the various states of the game and where it expects to gain the most reward.

F. Conclusion

In this paper, we have tested several strategies and rule changes in Blackjack concerning their effect on winning percentage using reinforcement learning. Our research focused on the basic strategy, Q-learning method, card counting systems, and certain rule changes.

The results show that applying Q-learning and card counting strategies gives a big improvement in the winning chances of the player. Hi-Lo and Knock-Out card counting showed a lot of improvement in the win rates; the latter worked just fine because of its ease and efficiency. We could find that implementation of the Knock-Out card count system in conjunction with Q-learning produced an impressive enhancement of winning percentages, thus showing the importance of more complex counting techniques in perfecting Blackjack performance.

These changes in rules also had huge interactions on the outcomes of gameplay. The "dealer stands on soft 17" rule benefited the player in reducing house edge and thus increased the win rate. On the other hand, multiple decks increased the difficulty in card counting and thus modified the effectiveness of a player's strategy and win rate. Even amidst such challenges in a multi-deck game, players will still benefit from modifying strategy accordingly.

The paper, in general, depicts that though the traditional strategies of Blackjack offer some baseline for playing, advanced techniques of reinforcement learning and card counting works effectively to the strategic imperative of improving the performance of the player. With such high-level approaches, the player can better cope with the intricacies of the game and thus increase the winning potential. Future research will be required to incorporate additional, advanced machine learning algorithms and their real-world applications in playing Blackjack, through which these strategies can be fine-tuned and developed.

In conclusion, our study provides valuable insights into optimizing Blackjack strategies through the use of reinforcement learning and card counting, contributing to a deeper understanding of how these methods can be effectively employed to gain a competitive edge in the game.

REFERENCES

[1] Avish Buramdoyal and Tim Gebbie. Variations on the reinforcement learning performance of blackjack. *arXiv preprint arXiv:2308.07329*, 2023.

[2] David B Fogel. Evolving strategies in blackjack. In *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)*, volume 2, pages 1427–1434. IEEE, 2004.

[3] Raghavendra Srinivasaiah, Vinai George Biju, Santosh Kumar Jankatti, Ravikumar Hodikehosahally Channegowda, and Niranjana Shravanabelagola Jinachandra. Reinforcement learning strategies using monte-carlo to solve the blackjack problem. *International Journal of Electrical and Computer Engineering (IJECE)*, 14(1):904–910, 2024.

[4] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[5] Edward O Thorp. *Beat the dealer: A winning strategy for the game of twenty-one*. Vintage, 2016.

[6] Mózes Vidámi, László Szilágyi, and David Iclanzan. Real valued card counting strategies for the game of blackjack. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23–27, 2020, Proceedings, Part II 27*, pages 63–73. Springer, 2020.