Reinforcement Learning Approaches to Blackjack

Sharik Anwar Zahir Hussain
MAI Faculty of Computer Science
Technical University of Applied Science Wurzburg-Schweinfurt
sharikanwar.zahirhussain@study.thws.de

***Abstract*—This paper investigates the application of combining reinforcement learning with counting systems to optimize blackjack. The principal aim is to utilize reinforcement learning to enhance the basic strategy for blackjack and determine its viability for implementing an effective algorithm. First, we have implemented a Q-learning algorithm to enhance the basic blackjack strategy, which proved to be dramatically improved in performance. Then we have implemented two rule variations derived from [3]. Next, we study the effects of adding rule variations into the basic strategy. Then, the high-low counting system was implemented for the complete point count system. To enhance performance even more, we added the Zen count system. By now, two different rule variations are being acted on and results are examined. Our experiments report the pros and cons of the approaches, offering an understanding of the difference in rule sets. The results indicate that incorporating of reinforcement learning into blackjack can greatly improve the performance of a blackjack player.**

***Keywords*—Strategic decision-making, Q-learning algorithm, Counting systems, Rule variations, Optimization.**

I. Introduction

Blackjack is one of the more popular card games in casinos. Players play against the dealer to get a total as close to 21 as possible. The player tries to stay under 21. This involves a lot of strategic decision-making, dependent on the cards in the player's hand and one shown card on the dealer's side. This makes it quite interesting to use mathematical and computational techniques to maximize play.

In this paper, we explore how reinforcement learning might improve and extend over traditionally used strategies in the blackjack game. Reinforcement learning (RL) is a sub-field of machine learning in which an agent learns to make decisions by observing its environment and maximizing rewards through actions. It enables the development of such algorithms that can adapt and improve over time, as a human does from interaction with the environment. RL has proven to be very successful when applied to complex decision-making problems in various domains, including games, robotics, and autonomous systems.

This has motivated this study where reinforcement learning has been applied to blackjack to give an outline of how it can improve conventional strategies. The dynamic nature of the game is motivating for applying reinforcement learning, where the state variables keep changing. On the other hand, RL will be able to develop adaptive strategies taking into consideration the current state of the game, which is expected to result in higher success rates.

The goal of this paper is to understand the performance of the Q-learning algorithm in blackjack playing strategies compared to traditional counting methods while also examining its effectiveness in various rule variations. Our work here tries to take a step by incorporating counting systems with reinforcement learning (RL), which then helps us understand how well such methods possibly perform across different sets of game rule conditions.

The paper is structured as follows: The Literature Review section summarizes relevant prior research on blackjack strategies, reinforcement learning applications, and counting systems. The Methodology section details the implementation of Q-learning in blackjack, including the theoretical background, the basic point count system, the Zen count system, and the rule variations considered. The Results and Discussion section provides the results, analysis, and comparison of the performance of various strategies. Finally, the Conclusion section summarizes the key findings, discusses their implications, and suggests directions for future research.

II. Literature Review

Edward O. Thorps' "Beat the Dealer" [3] transformed blackjack by introducing the concept of card counting which is based on decision-making statistics. Thus it lays down ground rules for tactics and presents a Hi-Lo count system which depends on fractions of high or low cards.

Despite being a bit dense, Richard S. Sutton and Andrew G. Barto give an extensive view of what is Reinforcement Learning from a principles and algorithmic point of view in [2]. It gives an in-depth view of the theoretical background of Q-learning which is the central algorithm we use in our study.

A framework is provided by "RLCard: A Toolkit for Reinforcement Learning in Card Games" [4] to integrate reinforcement learning into card games, which concentrates much on blackjack. This toolkit makes it easy to create and test out reinforcement learning

strategies. In addition, specific indices are assigned to actions such as 'hit' or 'stand', which make learning and acting possible for the agents.

"Applying Reinforcement Learning to Blackjack using Q-Learning" [1] : The primary objective of the paper is to play blackjack perfectly by deploying a Q-learning algorithm. When set to play against random or those using basic strategy, it was established that an agent that uses Q-learning always continues to perform better for instance in blackjack.

III. Methodology

A. Interactive Blackjack using Basic Strategy

The easy straightforward method of an interactive blackjack game using the basic strategy where the players play against the house by observing the Dealers' card. Two cards are handed out at the beginning of the game. Two cards for the player and two for the dealer. The player uses a basic strategy to decide whether to hit or stand, which is guided by the total value of the card in the player's hand and the dealer's visible card. The player can pick the real-time action in this interactive game. The dealer is only acting according to the typical blackjack rules which means, the player hits until the hand value reaches at least 17. The game can last until the dealer or the player busts out or decides to stand. The outcome is determined by comparing the hand values.

B. Basic Strategy with Q-Learning

Q-learning is applied to the Blackjack game with basic strategy. Q-learning is a model-free reinforcement learning algorithm, which learns the value of action-state pairs by knowledge from feedback. A state space would consist of the player's hand value, the dealer's visible card, and a boolean indicating whether the game is over for simulating the game environment. The actions that are available to the agent: are hit, stand, split pairs, and double down. Q-learning is an iterative update of Q-values through the rewards for taking action. It seeks to maximize the total reward over time. The Q-value is updated by the formula:

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

where α is the learning rate, γ is the discount factor, r is the reward received, s is the current state, a is the action taken, s' is the next state, and a' is the next action. The equation (1) is designed based on the level of the Q- value related to the particular state-action that is to be changed. The development process involves carrying out repeating tasks, checking the system, and taking a different strategy with each step. Based on the results, this approach allows for learning. Gradually, the agent will change itself from an explore to an exploit state as it gains experience of the stacks while trying to find the most efficient strategies.

C. Basic Strategy with Rule Variations

In this implementation, the Q-learning algorithm is applied in a blackjack environment considering two rule variations: "Dealer Hits on Soft 17" and "Dealer Stands on Soft 17." The goal is to understand how these variations affect the agent's learning and strategy. The methodology remains similar to the basic Q-learning approach, for it only changes the game environment to allow for these rule variations. The testing rule does determine the dealer's behavior, and in turn, affects how the agent shall strategize and choose their move. Q-learning algorithm learns the different strategies in the case of every variation of the rule through simulation of many episodes by iteratively updating the Q-values using (1)

D. Complete Point Count System

In this case, a complete point count system has been implemented. The Hi-Lo system has been implemented and keeps track of the running counts of cards, adjusting further strategies in betting and the player's decisions on the remaining count of cards in the deck. State variables for this Q-learning agent include the player's hand value, the dealer's up card, the running point count, and deck penetration. Hitting, standing, doubling down, and splitting are all actions that can be taken.

$$\text{point\_count} += \sum(\text{card\_values}) \tag{2}$$

where card values are adjusted based on a predefined system (e.g., +1 for low cards, -1 for high cards). The count will help the player bet more intelligently and give the ability to optimize the strategy. The equation refines the strategies over several episodes, ensuring that the agent exploits optimally with the information given by the point count system. This will enable the agent to adjust the strategy dynamically due to the current state of the deck.

E. Complete Point Count System with Rule Variations

This implementation extends the previous methodology by incorporating specific rule variations: "Dealer Stands on Soft 17" and "Early Surrender". The complete point count system is combined with a Q-learning agent, which learns to adjust its strategies based on these rule changes. The game environment is modified to reflect the rule variations, impacting the agent's decision-making process. The agent considers the point count, the specific rule in effect, and other game states to decide on actions. The Q-learning algorithm iteratively updates the Q-values, learning the optimal strategies for each rule variation over multiple episodes. The running point count is updated using the equation (2), and the Q-learning equation is applied to learn optimal strategies. The integration of rule variations allows for a more comprehensive evaluation of the agent's adaptability and performance.

F. Enhanced Complete Point Count System with Zen Count

The final methodology consists of utilizing the Zen count system, a specialized card counting technique, in combination with a Q-learning agent and comparing performance on rule variations: "Dealer Stands on Soft 17" and "Early Surrender". The Zen count has a more accurate tracking of the deck composition, so it contains a list of different cards with different points. The player's hand value, the dealer's visible card, the Zen count index, and the deck penetration level are part of the state space for the Q-learning agent.

$$\text{zen\_index} = \frac{\text{point\_count}}{\left(\frac{\text{unseen\_cards}}{52}\right)} \tag{3}$$

The normalized version of the equation (3) is used for the determination of the unseen card numbers. If the agent plays with the Zen count as required by the specific rules then it also bets and plays with the life-stack count. The updated Q-value equation (1) operates to find the optimal strategy through the agent's updates of the Q-values as a response to various episodes. Such a method enables a more insightful reading of the Zen count and eventually allows for the derivation of more suitable strategies corresponding to various moves and changing rules thereby allowing the program to always base its decisions on the most complete information.

IV. Results and Discussion

This section provides the results and comparative analyses of the various strategies implemented as part of this study. To that effect, each of the subsections will correspond to one methodology that is described earlier and compare the same in terms of win percentage results obtained.

A. Interactive Blackjack using Basic Strategy

The baseline for evaluating the efficacy of more sophisticated strategies is provided by the interactive blackjack game that employs basic strategy. The basic strategy decisions are made based on the player's hand and the dealer's visible card, following predefined rules without any adaptation or learning. The results indicate a win percentage of 40%. From the results, we can see that it is less effective than more complex approaches because this strategy is straightforward and simple to use. It does not adapt to changing conditions in the gaming environment.

B. Basic Strategy with Q-Learning

By implementing Q-learning, the performance is much better than the basic strategy. The win percentage increases to 45% with a 5% increase. The Q-learning algorithm learns optimal strategies by iteratively updating Q-values based on the rewards received from different actions. This learning process allows the agent to adapt its strategy based on the outcomes of previous games, leading to better decision-making over time. The comparison with the basic strategy highlights the advantage of using reinforcement learning, which can learn and optimize strategies dynamically.

C. Basic Strategy with Rule Variations

Introducing rule variations, such as "Dealer Hits on Soft 17" and "Dealer Stands on Soft 17," shows how flexible the Q-learning algorithm is. The results show a win percentage of 46% when the dealer hits on soft 17 and 44% when the dealer stands on soft 17. Figure 1 and .Figure 2 shows the winning rate surface for "Dealer Hits on Soft 17" and "Dealer Stands on Soft 17" respectively. When the dealer's up card is low and the player's hand value is near 21, higher winning rates can be seen. It suggests that it is the best time to employ aggressive strategies. When compared to the dealer hitting on soft 17, this scenario shows a typically higher winning rate, with more prominent peaks.

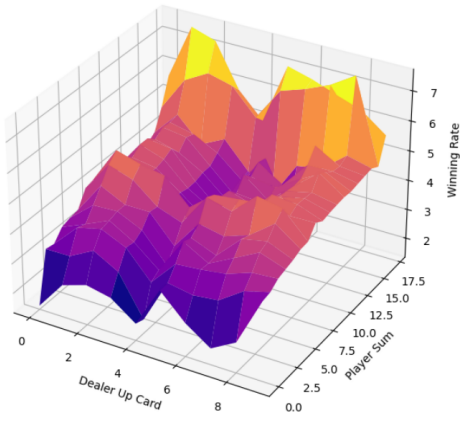

Fig. 1: Winning Rate Surface for Dealer Hits on Soft 17

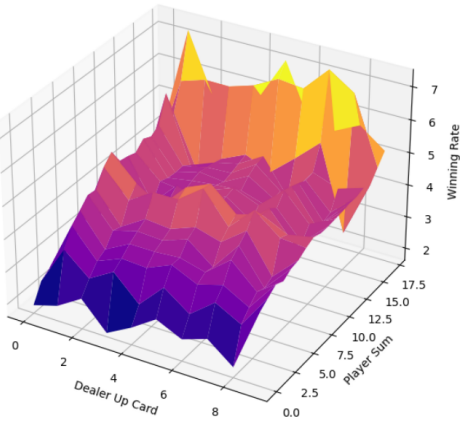

Fig. 2: Winning Rate Surface for Dealer Stands on Soft 17

.Figure 3, compares the win rates by episode block for both rule variations. It shows that over time, the win percentage for "Dealer Hits on Soft 17" consistently outperforms "Dealer Stands on Soft 17".

The probability of the dealer busting while hitting a soft 17 is higher, which explains the variation in win percentages. When the dealer hits soft 17, there is a higher chance that the card value exceeds 21, resulting in a bust. This rule provides a slight advantage to the player, as the increased probability of the dealer busting improves the player's chances of winning. Conversely, when the dealer stands on soft 17, they avoid the risk of busting with an additional card, which reduces the player's advantage. The comparison between the figures demonstrates that the rule where the dealer stands on soft 17 is slightly less favorable for the player.

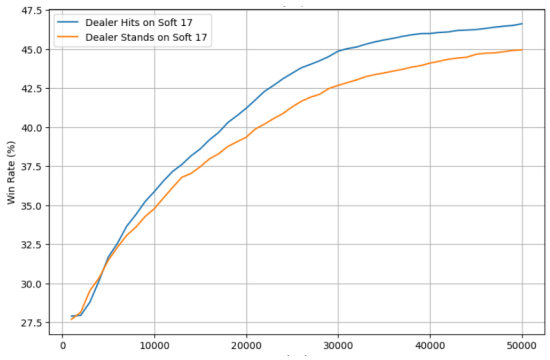

Fig. 3: Win Rate over Episodes

D. Complete Point Count System

Integrating the complete point count system with Q-learning achieves a win percentage of 45%. The point count system keeps track of the running count of cards that are being dealt with, which helps the agent to make better decisions. This additional information allows the agent to adjust its betting and playing strategies based on the remaining cards in the deck. The higher win percentage in results highlights the advantage of combining card counting techniques with reinforcement learning, providing a robust strategy that makes use of comprehensive knowledge about the state of the game.

E. Complete Point Count System with Rule Variations

Incorporating specific rule variations, such as "Dealer Stands on Soft 17" and "Early Surrender," into the complete point count system results in win percentages of 41% and 40%, respectively. These outcomes indicate that the effectiveness of the strategy can be influenced by the specific rules in effect. The Q-learning algorithm is employed consistently to adapt the strategies based on these rule variations. As shown in Figure 4, the "Dealer Stands on Soft 17" rule results in a higher win percentage compared to "Early Surrender". Initially, the win percentages for both variations increase rapidly as the Q-learning algorithm optimizes the strategy. Over time, the win percentages stabilize, reflecting the effectiveness of the learned strategies. The comparison highlights that the "Dealer Stands on Soft 17" and "Early Surrender" rules both have almost similar win percentages.

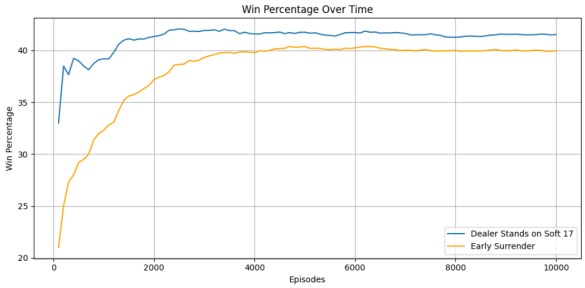

Fig. 4: Win Percent over Time

The almost similar win percentages for the two rule variations can be attributed to the comprehensive nature of the complete point count system, which effectively captures the overall card distribution and aids decision-making. The complete point count system's detailed tracking of cards helps maintain a consistent advantage, regardless of the rule changes, explaining the nearly identical win rates for "Dealer Stands on Soft 17" and "Early Surrender".

F. Enhanced Complete Point Count System with Zen Count

Zen count system is an advanced card counting technique. The implementation of the Zen count system with Q-learning under rule variations achieves win percentages of 41% for "Dealer Stands on Soft 17" and 40% for "Early Surrender". As shown in Figure 5, the win percentage over time for both rule variations demonstrates the performance differences. The graph indicates that the "Dealer Stands on Soft 17" rule consistently results in a higher or similar win percentage compared to "Early Surrender". Initially, the win percentages for both variations increase rapidly as same as the previous counting system. As a result of learned strategies over the period of time, the win percentages eventually stabilize. The comparison highlights that the "Dealer Stands on Soft 17" rule is slightly more favorable for players, emphasizing the value of the Zen count system combined with Q-learning in optimizing strategies under different rule variations.

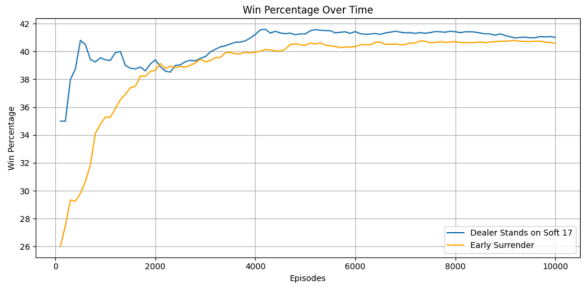

Fig. 5: Win Percent over Time

Despite changing the count system to Zen count, the win percentages remain almost similar to those

achieved with the Hi-Lo count system. This similarity can be explained by the fact that both the Hi-Lo and Zen count systems are balanced card counting techniques designed to provide an advantage by tracking the ratio of high to low cards. The Zen count system is slightly more complex and may offer marginally better accuracy in some scenarios, but the overall impact on win rate is minimal because both systems fundamentally rely on the same principle of tracking card distributions to inform betting and playing decisions. The results indicate that although the Zen count offers finer granularity and may theoretically be superior, in reality, its advantages over the Hi-Lo count are not substantial enough to have an apparent effect on win rates. Both systems improve decision-making and strategy optimization under different blackjack rule variations, but the inherent nature of the game and the effectiveness of Q-learning in leveraging these counts result in comparable win percentages.

Comparing Figures (4) and (5) reveals the detailed performance metrics of the Zen count system over the basic Hi-Lo count system. (Figure4), which represents the win percentages for the Hi-Lo count system with Q-learning under the rule variations "Dealer Stands on Soft 17" and "Early Surrender," shows that the win percentages stabilize around 41% and 40%, respectively. (Figure 5) depicts the win percentages for the Zen count system with the same Q-learning and rule variations, also stabilizes at 41% for "Dealer Stands on Soft 17" and 40% for "Early Surrender". Although the win percentages for both systems are close, the Zen count system offers improved consistency and potentially better decision-making due to its finer granularity in card values. This refined granularity can lead to slight improvements in specific scenarios, making it a theoretically superior choice. However, in practical terms, the marginal improvements are not substantial enough to result in a noticeable difference in win rates compared to the Hi-Lo count system. Therefore, even though the Zen count system might provide better performance in certain contexts, both systems essentially work according to the same ideas and produce results that are comparable when considering every aspect of the game.

V. Conclusion

This research explores various blackjack strategies and evaluates their performance using reinforcement learning and advanced card-counting techniques under different rule variations. The key findings highlight the effectiveness of Q-learning in optimizing blackjack strategies and demonstrate the advantages of incorporating point counting systems like the Hi-Lo and Zen count. A concise summary of different blackjack strategies, their rule variations, corresponding win percentages are presented in Table I.

Code Description	**Rule Variations**	**Win Percentage**
Interactive Blackjack using Basic Strategy	N/A	40%
Blackjack with Q-Learning	N/A	45%
Blackjack with Rule Variations and Q-Learning	Dealer Hits on Soft 17, Dealer Stands on Soft 17	46% (Hits on Soft 17), 44% (Stands on Soft 17)
Blackjack with Complete Point Count System	N/A	45%
Blackjack with Complete Point Count System and Rule Variations	Dealer Stands on Soft 17, Early Surrender	41% (Stands on Soft 17), 40% (Early Surrender)
Enhanced Blackjack with Zen Count and Rule Variations	Dealer Stands on Soft 17, Early Surrender	41% (Stands on Soft 17), 40% (Early Surrender)

TABLE I: **Strategies and Performance Summary**

The implications of these findings suggest that advanced reinforcement learning algorithms and sophisticated card counting systems can significantly improve blackjack strategy outcomes. This research underscores the importance of adapting strategies based on specific game rules while making use of comprehensive information regarding the game state.

Future research could explore the integration of other advanced machine learning techniques and deeper exploration of additional rule variations to further optimize blackjack strategies. Furthermore, real-world testing and validation of these strategies in live casino environments would provide valuable insights into their practical effectiveness and adaptability.

References

[1] Charles De Granville. Applying reinforcement learning to blackjack using q-learning. *University of Oklahoma*, 2005.

[2] Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduction. *Robotica*, 17(2):229–235, 1999.

[3] Edward O Thorp. *Beat the dealer: A winning strategy for the game of twenty-one*. Vintage, 1966.

[4] Daochen Zha, Kwei-Herng Lai, Yuanpu Cao, Songyi Huang, Ruzhe Wei, Junyu Guo, and Xia Hu. Rlcard: A toolkit for reinforcement learning in card games. *arXiv preprint arXiv:1910.04376*, 2019.