CS 370 Project Two Design Defense

Deep Q-Learning Implementation for Intelligent Agent Pathfinding

Brett Plemons

Southern New Hampshire University

CS 370: Current/Emerging Trends in Computer Science

6/22/2025

Abstract

This design defense analyzes a deep Q-learning algorithm implementation for solving pathfinding problems in a treasure hunt game. The study compares human cognitive problem-solving approaches with machine learning methodologies, examines exploration versus exploitation balance in pathfinding, and evaluates the neural network implementation. The agent successfully achieves 100%-win rates through reinforcement learning, demonstrating effective navigation in complex environments.

Keywords: deep Q-learning, reinforcement learning, pathfinding, artificial intelligence, neural networks, maze navigation

Introduction

Developing intelligent agents capable of autonomous navigation represents a fundamental challenge in artificial intelligence. This design defense examines the implementation of a deep Q-learning algorithm for solving pathfinding problems in a treasure hunt game environment. The intelligent agent (pirate) must navigate through an 8x8 maze to reach the treasure while avoiding obstacles and maximizing rewards.

Deep Q-learning combines traditional reinforcement learning with neural networks, enabling agents to handle complex state spaces effectively (Mnih et al., 2015). This approach has proven particularly successful in navigation tasks where traditional rule-based systems struggle with dynamic environments and reward optimization.

This analysis addresses three key areas: comparing human cognitive approaches with machine learning methodologies for maze navigation, examining the balance between exploration and exploitation in pathfinding contexts, and evaluating the effectiveness of the neural network implementation for this specific problem domain.

Human vs. Machine Problem-Solving Approaches

Human Approach

Humans solve this maze through intuitive spatial reasoning and strategic planning. They visually scan the entire maze, identify obvious paths and dead ends, and mentally trace potential routes before moving. They apply heuristics like "move toward the target" and remember failed paths to avoid repeating mistakes. This cognitive approach leverages evolved spatial intelligence for immediate understanding and rapid adaptation to new environments.

Machine Approach

The deep Q-learning agent uses systematic trial-and-error learning combined with mathematical optimization. Starting with random behavior, the agent gradually learns through reinforcement which actions yield rewards. The algorithm employs experience replay to store and learn from past episodes, uses epsilon-greedy exploration to balance trying new actions versus using learned knowledge, and utilizes neural networks to recognize patterns in state-action-reward relationships.

Key Differences

The primary differences include processing speed (humans understand immediately but decide slowly, while agents require training but process rapidly), learning methods (cognitive reasoning versus mathematical optimization), and generalization (humans adapt instantly to new mazes, while agents need retraining for different configurations). However, both approaches learn from experience and optimize performance over time.

Intelligent Agent's Purpose in Pathfinding

Exploration vs. Exploitation Paradigm

The fundamental challenge in pathfinding involves balancing exploration of unknown paths with exploitation of learned successful strategies. Exploration allows for discovering alternative routes and gathering environmental information, while exploitation uses known optimal paths to maximize rewards.

For this maze problem, the optimal balance evolves during training. Initially, epsilon = 0.1 provides 10% exploration, ensuring comprehensive path discovery while building on emerging knowledge. When win rates exceed 90%, epsilon reduces to 0.05, allowing policy refinement while maintaining flexibility for edge cases. This adaptive approach ensures thorough state-space coverage early, enabling optimization as competence develops.

Reinforcement Learning’s Role

Reinforcement learning determines optimal paths through several mechanisms. The reward structure (+1.0 for treasure, -0.75 for obstacles) creates clear incentives for goal-directed behavior. Experience replay enables learning from multiple episodes simultaneously, while Q-value learning helps the agent predict expected future rewards for each action. Through iterative training, the agent develops policies that consistently guide navigation toward the treasure while avoiding obstacles.

Algorithm Evaluation

Neural Network Implementation

The implementation uses a three-layer neural network: a 64-neuron input layer accepting the flattened maze state, two hidden layers with 64 neurons each using PReLU activation, and a 4-neuron output layer representing Q-values for each possible action (left, up, right, down). This architecture provides sufficient complexity for maze navigation while maintaining training efficiency.

Deep-Q Learning Features

The algorithm incorporates several key components for effective learning. Experience replay stores episodes in memory and samples them randomly for training, preventing correlation between consecutive updates and improving stability. The epsilon-greedy policy balances exploration and exploitation through adaptive scheduling. Q-values are updated using the Bellman equation: , where γ = 0.95 represents the discount factor. Mean Squared Error loss between predicted and target Q-values provides stable gradient updates.

Performance Results

The implementation successfully achieves 100% win rates within approximately 1000 training epochs, demonstrating effective convergence. The completion check verifies that the agent can navigate to the treasure from any starting position, indicating robust policy learning rather than memorization. Training remains stable throughout the process without catastrophic forgetting or oscillatory behavior.

The algorithm effectively addresses potential local optima through sustained exploration and experience replay, while maintaining computational efficiency with reasonable training times. Current limitations include specificity to the given maze configuration, requiring retraining for different layouts.

Conclusions

This deep Q-learning implementation successfully demonstrates reinforcement learning principles applied to pathfinding problems. The analysis reveals fundamental differences between human spatial reasoning and machine learning optimization, highlighting complementary strengths of each approach. The careful balance of exploration and exploitation, combined with appropriate neural network architecture, results in a practical solution achieving optimal navigation performance.

The project illustrates practical applications of deep reinforcement learning in autonomous navigation, providing insights into intelligent agent development. The systematic approach to learning optimal policies through environmental interaction showcases how AI systems can solve complex problems that traditional rule-based algorithms might struggle with in dynamic environments.

References

Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review*, 25(3), 1212–1219. https://doi.org/10.3758/s13423-017-1347-z

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. https://doi.org/10.1038/nature14236

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 30, No. 1, pp. 2094-2100). https://doi.org/10.1609/aaai.v30i1.10295

Wang, Z., Schaul, T., Hessel, M., Hasselt, H. V., Lanctot, M., & Freitas, N. D. (2016). Dueling network architectures for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning* (pp. 1995-2003).