424 Reversi Agent Report

Elya Renom
McGill ID: 261094604

James Kidd
McGill ID: 260276236

Executive Summary

The strongest algorithm we discovered to play Reversi is a combination of **iterative deepening**, **minimax search**, and **alpha-beta pruning**, all designed to work within a strict **2-second time constraint**. This approach provides a balance of **strategic depth**, **computational efficiency**, and **adaptability to different board sizes**, making it effective in competitive play. Our goal was to create an agent capable of not only making **optimal moves in immediate situations** but also **strategically planning ahead**, while maintaining **adaptability for dynamic board states**.

The crux of the agent's strength is the **iterative deepening framework**. This method progressively increases the search depth, ensuring that even if time constraints limit the exploration of the entire search space, the agent will always return the **best move found so far**. This feature was critical in handling the **strict 2-second per move constraint**, allowing our agent to explore **deeper layers of decision-making** when the computational budget allowed. The **minimax algorithm** forms the decision-making backbone, simulating all possible moves for both the player and the opponent to find the **optimal outcome** under the assumption that the opponent plays perfectly. **Alpha-beta pruning** enhances this process by **eliminating branches of the search tree** that cannot influence the final decision, significantly **improving efficiency**.

Together, these techniques allowed our agent to **calculate optimal strategies** while adhering to computational constraints. A key contributor to the agent's success is the **heuristic evaluation function**, which assesses board states at leaf nodes of the search tree. The evaluation function prioritizes critical factors like **coin parity** (difference in the number of discs controlled by the player and opponent), **mobility** (difference in valid moves available to each player), and **corner control** (possession of stable and strategically advantageous corner positions). This heuristic balances **short-term gains with long-term strategic objectives**, ensuring the agent maintains **dominance on the board** while setting up **favorable endgame conditions**.

Detailed Explanation of Agent Design

Core Algorithms

The agent is built on three foundational techniques:

1. **Iterative Deepening Search (IDS):** This technique dynamically adjusts the search depth, progressively increasing it within a **2-second per-move time limit**. IDS ensures that the agent always returns the best move from the deepest completed depth.

2. **Minimax Algorithm:** Minimax simulates all possible moves for both the agent and opponent, assuming **perfect play**. The agent seeks to maximize its score while minimizing the opponent's gains.

3. **Alpha-Beta Pruning:** This optimization reduces the number of evaluated states by **eliminating branches** that cannot affect the decision, enabling deeper searches within the time limit.

Iterative Deepening Search (IDS)

Iterative Deepening Search (IDS): IDS combines the completeness of depth-first search (DFS) with the optimality of breadth-first search (BFS). The algorithm incrementally increases the search depth d, evaluating all possible moves up to d within the available time budget. Mathematically, the time complexity of IDS is expressed as:

$$T_{\text{IDS}} = \sum_{i=1}^{d} O(b^i) = O(b^d),$$

where b is the branching factor, and d is the maximum depth. This approach ensures that the agent always returns the best move discovered at the deepest fully evaluated level, even if time runs out during deeper iterations. In Reversi, the branching factor b depends on the number of valid moves, which varies dynamically based on the board state.

IDS is the backbone of our agent's decision-making process. It progressively deepens the search tree, evaluating board states at increasing depths until the **2-second per-move time limit** is reached. The primary advantage of IDS is that it ensures the agent always returns the **best move discovered** within the time constraint. If the agent cannot fully explore the search tree due to time limits, it will still provide a move based on the most recent completed depth. The IDS implementation operates as follows:

1. Start with a depth of 1 and incrementally increase the search depth.

2. Evaluate possible moves using **Minimax Search** and **Alpha-Beta Pruning**.

3. If time expires before completing the current depth, the agent selects the **best move identified in previous iterations**.

This method is particularly effective in **time-critical environments** like competitive Reversi matches, ensuring robust decision-making under constraints.

Minimax Algorithm with Alpha-Beta Pruning

The agent employs Minimax to evaluate game states by simulating all possible moves for both players under the assumption of perfect play. The utility function $U(s)$ for a board state s is defined as:

$$U(s) = \begin{cases} +1 & \text{if the agent wins,} \\ -1 & \text{if the opponent wins,} \\ 0 & \text{if the game is a draw.} \end{cases}$$

In practice, this binary evaluation is replaced with a heuristic function $H(s)$, which provides a more granular assessment of non-terminal states. The Minimax algorithm recursively computes the optimal value for each state:

$$V(s) = \begin{cases} \max_{a \in A(s)} V(s') & \text{if the agent is to move,} \\ \min_{a \in A(s)} V(s') & \text{if the opponent is to move,} \end{cases}$$

where $A(s)$ is the set of valid actions from state s, and s' is the resulting state after action a. **Alpha-Beta Pruning** optimizes the Minimax search by eliminating branches that do not influence the final decision. Specifically:

- α: Tracks the **best score** achievable by the maximizing player (agent).
- β: Tracks the **best score** achievable by the minimizing player (opponent).

Pruning occurs when $\beta \leq \alpha$, reducing computational overhead and enabling **deeper exploration** within the 2-second limit.

Heuristic Evaluation Function

The heuristic evaluation function estimates the quality of a board state based on three key factors:

1. **Coin Parity:** Measures the difference in discs controlled by the agent and opponent:
$$\text{Coin Parity} = \frac{\text{Agent Discs} - \text{Opponent Discs}}{\text{Total Discs}}.$$
2. **Mobility:** Evaluates the difference in valid moves available to each player:
$$\text{Mobility} = \frac{\text{Agent Moves} - \text{Opponent Moves}}{\text{Total Moves}}.$$
3. **Corner Control:** Rewards ownership of stable corner discs, which cannot be flipped for the remainder of the game.

The final heuristic value is computed as:

$$H(S) = 2.0 \times \text{Coin Parity} + 3.0 \times \text{Mobility} + 5.0 \times \text{Corner Control}.$$

Weights were determined through **grid search** and **random sampling** across various board sizes (6x6 to 12x12), optimizing for win rate, average discs captured, and mobility advantage.

Board Size and Performance Impact

The agent dynamically adjusts its depth and breadth of exploration based on board size:

- On **smaller boards** (e.g., 6x6), the agent reaches depths of 3-5.
- On **larger boards** (e.g., 12x12), the increased branching factor limits depth to 2-3.

Breadth varies with game state:

- Smaller boards: Evaluates 5-10 moves per state.
- Larger boards: Evaluates up to 20 moves per state.

Strategy and Dynamic Adaptation

Initially, our approach to developing a competitive Reversi agent involved segmenting the game into three distinct phases—**opening**, **midgame**, and **endgame**—based on the percentage of empty spaces on the board. Each phase was to be governed by a tailored strategy that reflected the unique priorities and dynamics of that stage of the game.

The **opening phase** prioritized mobility and strategic positioning, particularly emphasizing corners and avoiding risky moves near unstable edges. The **midgame** was designed to focus on balancing mobility and stability while limiting the opponent's options. Finally, the **endgame** sought to maximize disc captures and secure stable regions of the board, where moves are less likely to be flipped.

We implemented these strategies by defining threshold percentages of empty cells to determine the transition between phases. For each phase, we designed evaluation heuristics and decision-making logic tailored to its objectives. During testing, this phase-based design demonstrated reasonable success in addressing the specific challenges of each part of the game. For example, it effectively avoided early pitfalls like **corner-adjacent moves** in the opening phase and navigated the more tactical decisions of the midgame.

However, while this phase-based system provided a clear framework, it introduced significant complexity. Transitioning between phases required additional computational overhead, and in some cases, the boundaries between phases were too rigid. These transitions often overlooked opportunities or risks unique to specific game states, particularly when the board's configuration didn't neatly align with predefined thresholds. Furthermore, this approach lacked the flexibility to dynamically adapt across varying board sizes and opponent strategies, leading to inconsistent performance during testing.

Through **iterative design and testing**, we realized that a unified strategy, driven by a single evaluation function and adaptable heuristics, could achieve better results without the added complexity. By incorporating heuristics that naturally accounted for phase-specific priorities—such as **corner control**, **stability**, and **mobility**—we captured the essence of the phase-based approach while maintaining simplicity and computational efficiency.

For example, **dynamic weighting of heuristics** based on board size and game state allowed the agent to adapt organically, reducing the need for explicit phase transitions.

Quantitative Analysis

Depth of Search

On smaller boards (6x6), the agent typically reaches depths of 3 to 5. On larger boards (12x12), the depth is limited to 2 to 3 due to the increased branching factor. Alpha-beta pruning ensures depth variation across branches.

Search Breadth

The agent evaluates 5-10 moves per state on smaller boards and over 20 on larger ones. Move ordering and pruning reduce unnecessary computations, prioritizing high-value actions such as corner captures.

Impact of Board Size

Larger boards increase computational complexity, requiring customized heuristics emphasizing corner control. Smaller boards allow deeper searches due to reduced branching factors.

Advantages

The principal advantage of our agent is the integration of Iterative Deepening Search (IDS) with Minimax and Alpha-Beta Pruning. This hybrid approach allows for efficient exploration of the game tree within the strict computational constraints imposed by the single-threaded Mimi-Server paired with the tournament's two-second decision time limit. During midgame scenarios, when the branching factor is at its peak, this strategy allows the agent to focus computational resources on promising branches, effectively reducing the branching factor and maximizing search depth.

IDS ensures that the agent always has a feasible move prepared by incrementally deepening the search, even if the search does not reach a terminal state. This dynamic adjustment prevents suboptimal decisions, such as random moves, which would inevitably occur under a static Minimax approach constrained by Alpha-Beta pruning alone.

To further optimize the search, we incorporated a transposition table to store previously evaluated game states. This mechanism minimizes redundant computations by leveraging memoization principles. Caching is particularly effective in a game like Reversi, where many states recur due to overlapping gameplay scenarios. In the context of our Iterative Deepening Search (IDS) strategy, the transposition table is particularly valuable as it avoids recomputing state evaluations across successive iterations of increasing depth.

Disadvantages

Our agent's heuristic design is limited by the absence of advanced evaluation metrics, such as coin parity and stability heuristics. These metrics provide critical insights into midgame and endgame dynamics, potentially enhancing the agent's ability to make nuanced strategic decisions. [**SOURCE**].

The fixed size and hashing approach of the transposition table introduce significant inefficiencies. On smaller boards, the sparsity of the table results in underutilization. Conversely, on larger boards, hash collisions and limited capacity lead to frequent evictions of valuable state information, compromising the caching mechanism's effectiveness. These issues reduce the consistency of the agent's performance across varying board sizes, as the transposition table struggles to adapt to the dynamic demands of different game configurations.

Finally, the agent's performance is notably influenced by board dimensions. On larger boards, where complex evaluations and deeper searches are required, the agent performs well. However, on smaller boards with simpler states, the computational resources are not fully utilized, leading to diminished efficiency. This disparity underscores the need for a more adaptive approach to handling diverse board sizes effectively.

Evaluation and predicted win rates of our agent

To evaluate the quality of our agent's play, we conducted rigorous testing across multiple dimensions. First, we performed extensive head-to-head simulations against a variety of opponent agents that we constructed ourselves using different approaches (some with the three-phase approach, some with just a heuristic, some with Monte Carlo, some with Minimax but different parameters), as well as the provided agents. These simulations were conducted on even boards of varying sizes, ranging from 6x6 to 12x12, to ensure the agent's adaptability to different game scenarios. We also engaged in iterative design, tweaking weights in the heuristic evaluation function and observing the impact on performance through controlled experiments.
Beyond automated testing, we played directly against the agent as humans to assess its strategic depth and adaptability in real-time scenarios. This hands-on approach gave us unique insights into the agent's strengths and weaknesses, revealing its ability to effectively capitalize on human errors while maintaining a consistent and strategic playstyle. Additionally, we consulted existing

research on Reversi strategies and algorithms to refine our approach, incorporating insights from both academic sources and prior implementations of game-playing AI.
Based on its strategic design, the agent demonstrates strong quantitative performance:

- **Against the Random Agent:** We predict a win rate of approximately **98 percent**, as the agent's heuristic-guided decision-making vastly outperforms a purely random strategy. Autoplay tests corroborate this with nearly flawless win rates.
- **Against an Average Human Player:** Against an average human player, such as "Dave," the estimated win rate is between **80-90 percent**, as the agent's prioritization of stable positions like corners allows it to excel in structured play, and it has better lookahead depth than a normal human. However, an experienced Reversi player could occasionally exploit tactical weaknesses or force mistakes.
- **Against Classmates' Agents:** The predicted win rate is **70-80 percent**, depending on the sophistication of the competing strategies. While our agent's iterative deepening and heuristic evaluation provide a competitive edge, advanced methods such as Monte Carlo Tree Search or more complex heuristics could pose challenges.

Future Improvements

While our current agent demonstrates strong performance, several areas for further enhancement could significantly improve its strategic depth, adaptability, and computational efficiency. Below are detailed proposals for future improvements:

1. **Dynamic Heuristic Weights:** The current heuristic function relies on fixed weights for coin parity, mobility, and corner control, which remain constant throughout the game. However, the relative importance of these factors changes across the opening, midgame, and endgame. Future iterations could implement dynamic weights that adapt based on the current game phase. For instance:
   - **Opening Phase:** Prioritize *mobility* to maximize potential moves and control of the board.
   - **Midgame:** Focus on *corner control* and edge stability to secure strategic positions.
   - **Endgame:** Emphasize *coin parity* to ensure dominance in disc count.

   Dynamic weighting can be implemented through a linear or non-linear function, such as:

   $$w_i(t) = w_i^{\text{base}} + f(t) \cdot \Delta w_i,$$

   where w_i^{base} is the base weight, $f(t)$ is a phase-specific adjustment factor, and Δw_i is the change in weight across game phases. The function $f(t)$ could depend on metrics such as the percentage of empty spaces or the mobility difference between players. Techniques like reinforcement learning could also optimize these weights through self-play, allowing the agent to learn phase-specific strategies.
2. **Advanced Stability Metrics:** While the current heuristic indirectly accounts for stability through corner control, explicitly incorporating stability metrics could further refine decision-making. Stability measures could classify discs as:
   - **Stable:** Discs that cannot be flipped for the remainder of the game (e.g., corners and discs surrounded by stable pieces).
   - **Semi-Stable:** Discs that are stable under certain conditions, such as specific opponent moves.

- **Unstable:** Discs that are highly vulnerable to flipping.

A stability metric S could be integrated into the heuristic function:

$$S = \sum_i s_i \cdot v_i,$$

where s_i represents the stability classification score (e.g., 1 for stable, 0.5 for semi-stable, and 0 for unstable), and v_i is the positional value of the disc. Incorporating this metric would enable the agent to make more nuanced decisions, such as avoiding risky placements near unstable edges or prioritizing moves that cluster stable regions.

3. **Dynamic Transposition Tables:** The current transposition table design uses a fixed size and hashing approach, which introduces inefficiencies across varying board dimensions. A dynamic transposition table system could address these limitations by preconfiguring table sizes tailored to specific board dimensions:

   - **Smaller Boards (e.g., 6x6):** Use smaller table sizes to minimize overhead, as the game states are simpler and searches terminate faster.
   - **Larger Boards (e.g., 12x12):** Allocate larger tables to accommodate the increased number of states and deeper searches.

   The capacity C of the transposition table could be dynamically allocated as:

   $$C = k \cdot b^d,$$

   where b is the average branching factor, d is the expected depth, and k is a scaling factor based on board size. Additionally, techniques such as Least Recently Used (LRU) eviction policies could prevent valuable state information from being overwritten during gameplay.

4. **Move Ordering with Machine Learning (if we were allowed):** Enhancing move ordering could significantly improve the efficiency of alpha-beta pruning. Instead of relying on heuristic scores alone, a supervised machine learning model could predict the likelihood of a move being optimal based on features such as:

   - Board configuration.
   - Opponent's recent moves.
   - Stability and mobility metrics.

   These predictions could prioritize the most promising moves, maximizing pruning opportunities and allowing deeper searches within the time constraint. A gradient boosting classifier or neural network trained on self-play data could serve this purpose.

5. **Integration of Monte Carlo Tree Search (MCTS):** While Minimax with alpha-beta pruning is effective, Monte Carlo Tree Search (MCTS) could be integrated as a complementary or alternative approach. MCTS is particularly effective in midgame scenarios with high branching factors, as it uses random simulations to evaluate moves without requiring a complete search. An adaptive hybrid model could use:

   - Minimax for shallow depths or when reliable heuristics are available.
   - MCTS for deeper or highly complex states where heuristic evaluation is less effective.

6. **Parallelization and Hardware Optimization:** To maximize computational resources, future iterations could implement parallel search techniques:

- **Parallel Alpha-Beta Pruning:** Distribute branches of the search tree across multiple threads or processors.
- **GPU Acceleration:** Use GPUs for large-scale evaluations, particularly for heuristic calculations and simulations.

By implementing these improvements, the agent would achieve greater adaptability, efficiency, and competitiveness, particularly against advanced human players and sophisticated AI opponents.

Acknowledgements

This project was informed by class discussions, academic resources, and testing against various opponents. We consulted example code for alpha-beta pruning and utilized grid search for heuristic tuning.