Evolutionary Optimization of Deep Learning Agents for Sparrow Mahjong

Jim O'Connor, Derin Gezgin, Gary B. Parker

Connecticut College

The 21st AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

Introduction

Evo-Sparrow is an AI agent for *Sparrow Mahjong*, a three-player variant of Mahjong that is much simpler yet preserves the game's core strategic complexity. We evolve a controller that utilizes a long short-term memory network using covariance-matrix evolution strategy.

To benchmark our evolved agent, we compare it against random, rule-based, and PPO-optimized agents. Our results show that, Evo-Sparrow is able to significantly outperform random and rule-based agents while having comparable performance to the PPO-optimized agent.

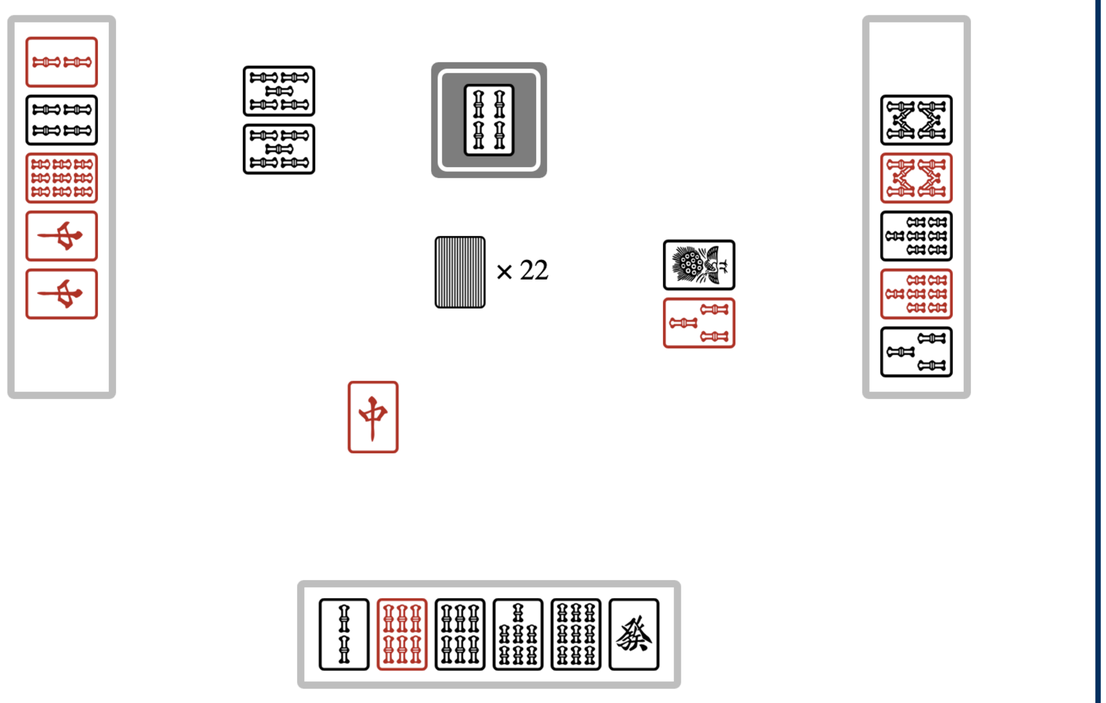

Figure 2: Example game state in Sparrow Mahjong.

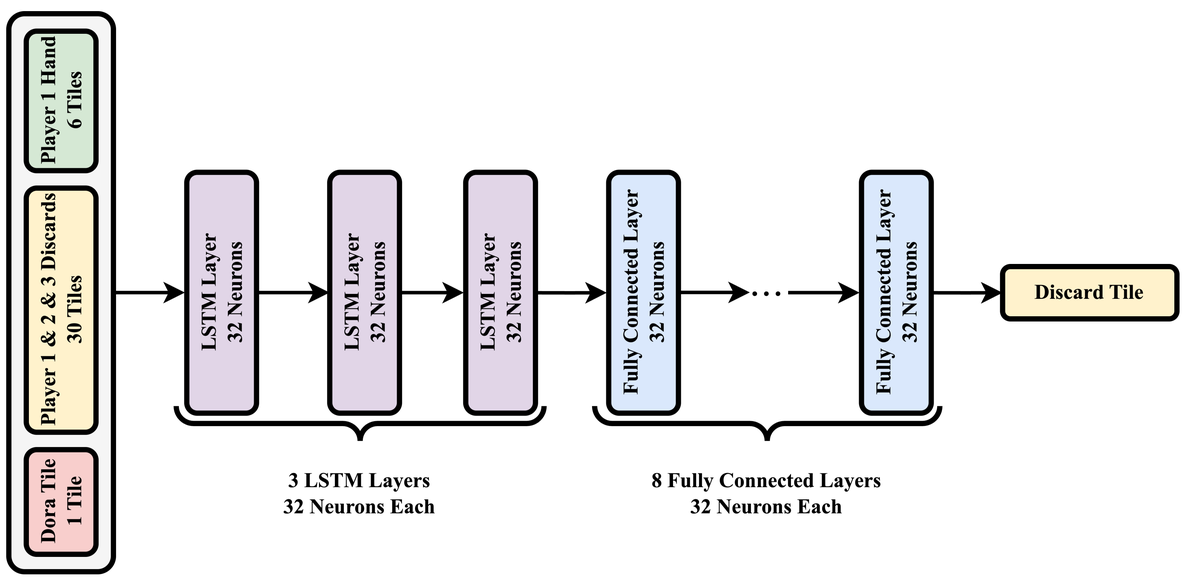

Figure 1: A high-level model of our LSTM network.

Methodology

Evo-Sparrow uses a long short-term memory (LSTM) network to process sequential game states in *Sparrow Mahjong*. Each 37-dimensional input vector includes the players hand, opponents discards, and the Dora indicator. This vector is passed through 3 LSTM layers followed by 8 fully connected layers, all with a size of 32. Finally the network outputs the tile to discard. Evo-Sparrow optimizes the network weights with the covariance matrix adaptation evolution strategy (CMA-ES). CMA-ES evaluates the candidate weights in parallel, keeps the best performers, and adjusts its search distribution toward them. Every generation, each candidate is tested over 200 simulated games, with the total score serving as its fitness. After 50 generations of training, best solution among the final set of candidate solutions is selected.

Results

Evo-Sparrow was evaluated against three baselines: a **random agent**, a **rule-based agent**, and a **PPO-optimized agent**. The rule-based agent applies simple heuristics to keep high-potential tiles and avoid helping opponents, while the PPO agent uses the same LSTM architecture and training setup for a fair comparison.

Across 1,000,000 games per match up, Evo-Sparrow achieved higher scores and win rates than both random and rule-based agents and maintained a lower deal-in rate, indicating balanced play. In self-play, identical Evo-Sparrow agents performed consistently, showing policy stability. Against the PPO agent, Evo-Sparrow reached similar performance while training about **2.6x faster**.

Table 1: Performance of Evo-Sparrow against random, rule-based, PPO-optimized agents and self-play.

	Avg. Score	Win %	Draw %	Loss %	Deal-in %
Evo-Sparrow	**0.8687**	**28.55**	44.17	**10.97**	**16.31**
Rule-Based	0.5051	20.91	44.17	12.65	22.28
Random	1.3738	6.64	44.17	18.66	30.54
Evo-Sparrow	**0.1934**	**22.80**	36.08	**17.60**	**23.52**
PPO-Sparrow	0.1868	22.62	36.08	17.74	23.57
Rule-Based	0.3802	19.07	36.08	9.98	34.87
Evo-Sparrow	0.0027	19.20	42.81	10.79	27.20
Evo-Sparrow	0.0062	19.25	42.81	10.77	27.17
Evo-Sparrow	0.0035	19.17	42.81	10.75	27.27