Team NanaBoshi

Katsuki Ohto (freelance) 2025.8.26

Overview

- NanaBoshi means “seven stars”
- Faster board implementation with C++
- Self-play reinforcement learning (policy gradient)
- Results vs baselines:
  - 100% against Random Trio (500 games)
  - 98% against Greedy Heuristic Agent (2000 games)

Methodology

- Policy & Value model
  - Model: Linear functions with ~100 parameters
  - Features: number of owned planets, amount of ships, simulated owner of each planet, simulated owner of each planet after each action, etc …

- Training
  - Off-policy policy gradient (V-trace based method)
  - Entropy regularization, KL regularization

Future Work

- Better policy & value model with more features or using neural networks
- Tree search
- Reinforcement learning with tree search (like AlphaZero does)