Opposing Reinforcement Learning Agents in Checkers

Bryant Saunders

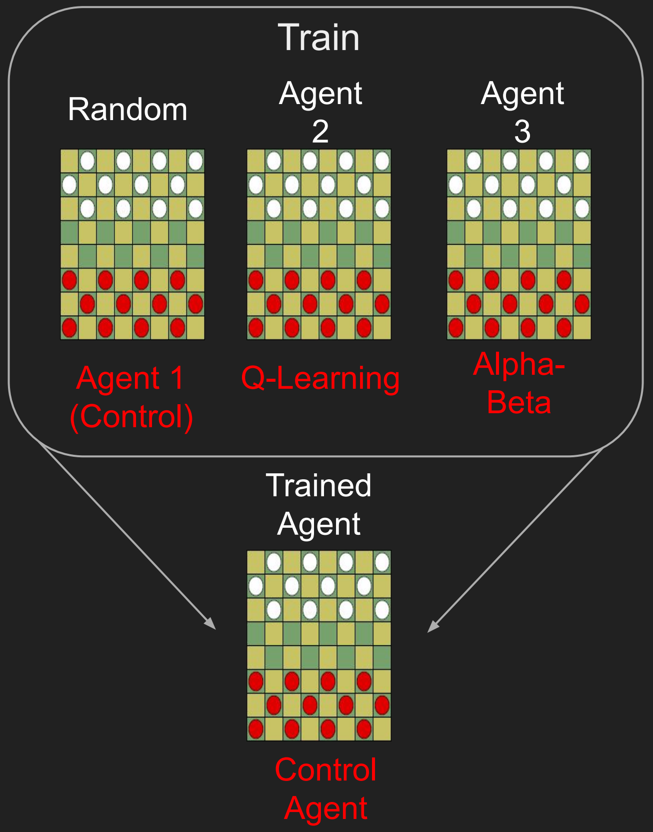
Overview

Project Goals:

- Compare different training methods in opponent processing

Motivations:

- Determine optimal training styles for a reinforcement learning
- Gain experience in multi-agent systems



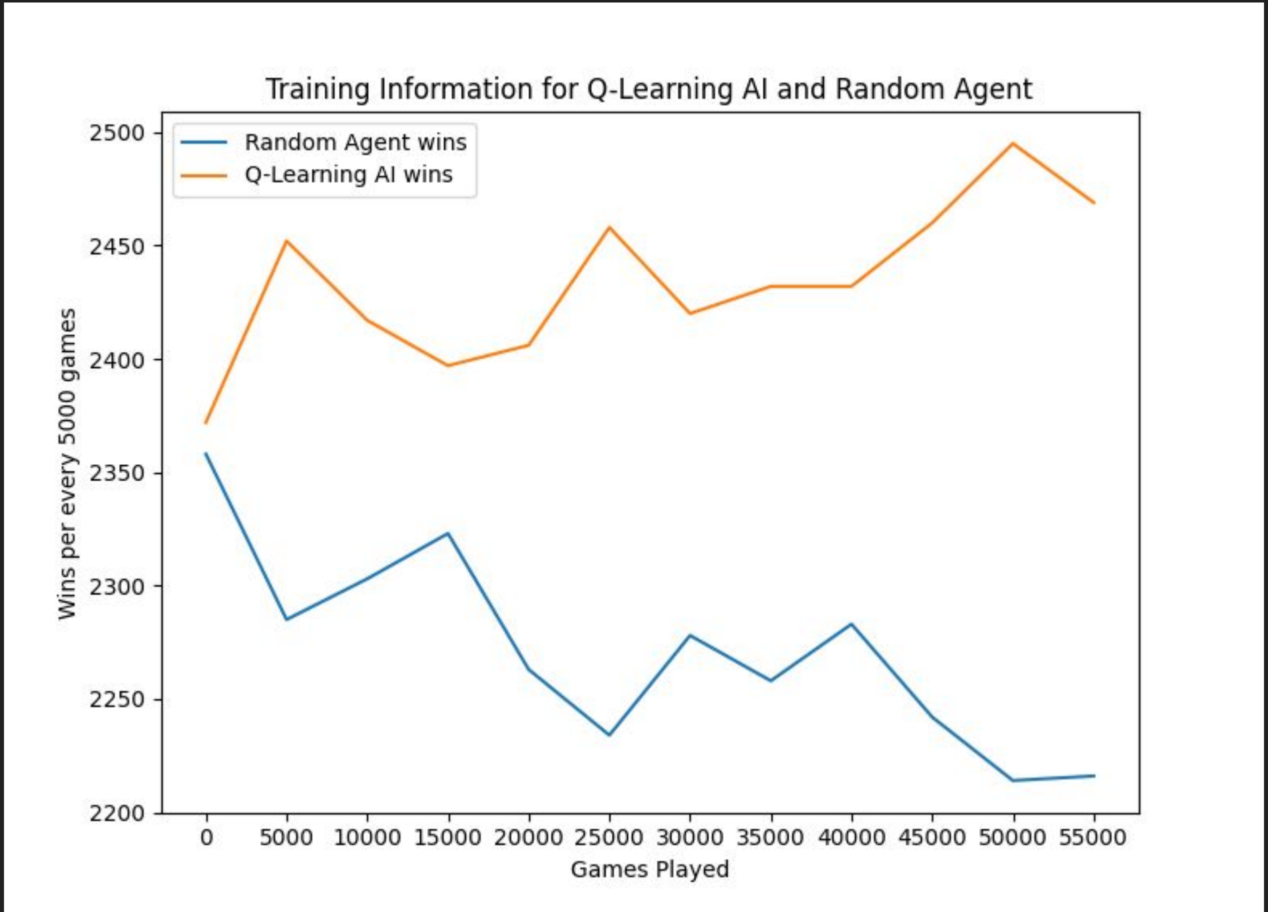
Methodology

- Train several Q-Learning agents independently
- Use a different training opponent for each agent
- After training, use a control agent as a benchmark to test training
- Compare win-rates between trained Q-Learning agents

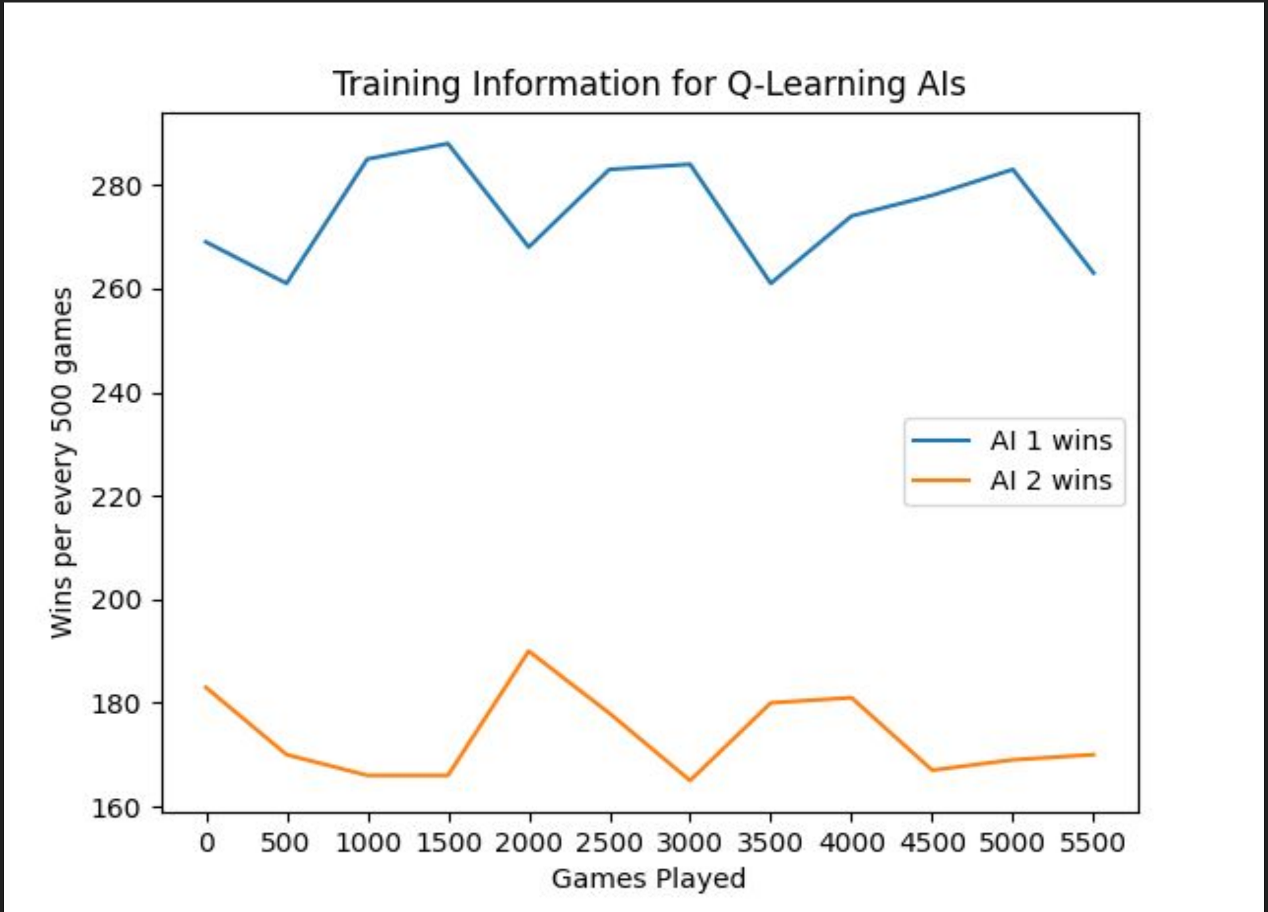
$$Q^{new}(s_t,a_t) \leftarrow \underbrace{Q(s_t,a_t)}_{\text{current value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Bigg(\underbrace{\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1},a)}_{\text{estimate of optimal future value}}}_{\text{new value (temporal difference target)}} - \underbrace{Q(s_t,a_t)}_{\text{current value}}\Bigg)}^{\text{temporal difference}}$$

All learners use the above value function

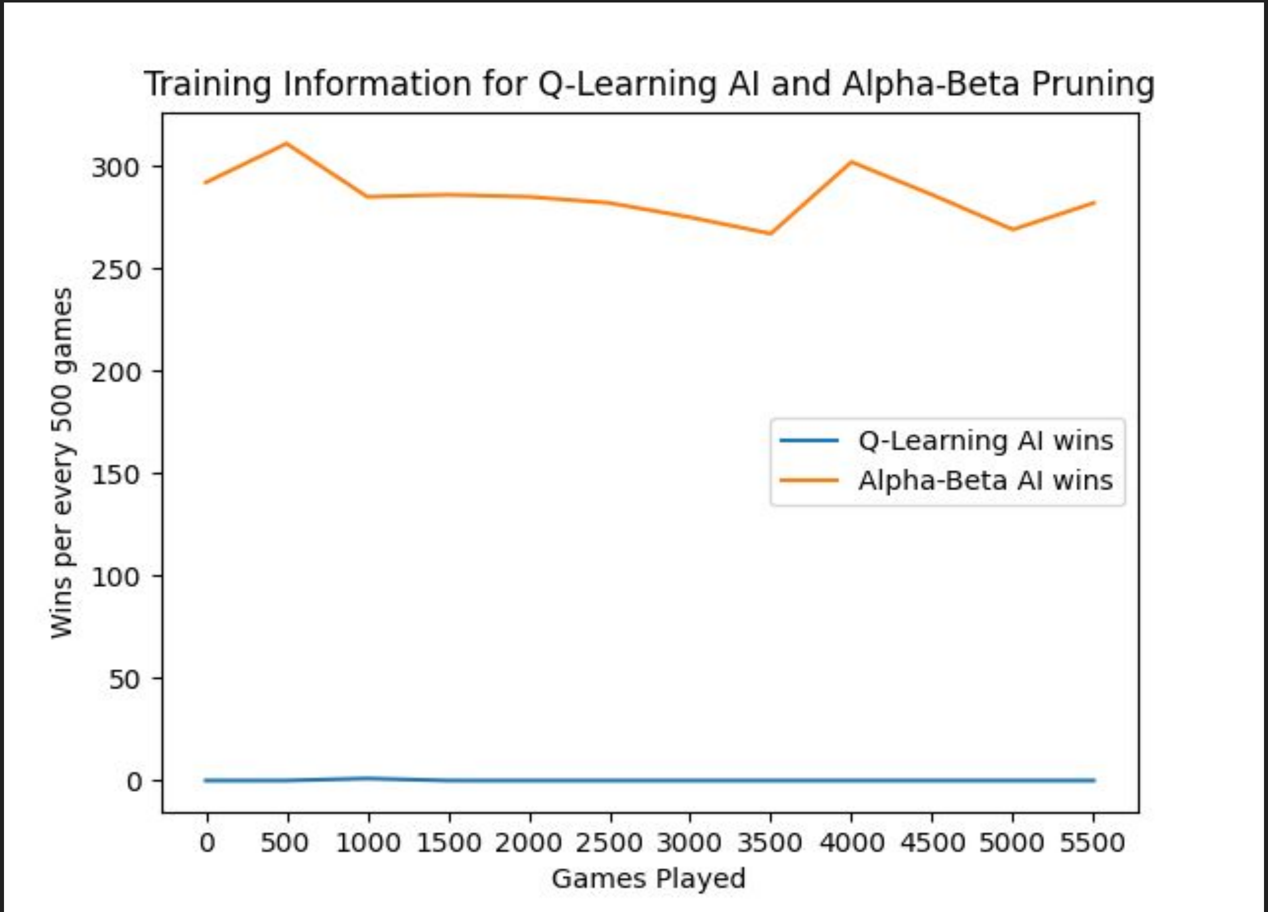
Control Agent



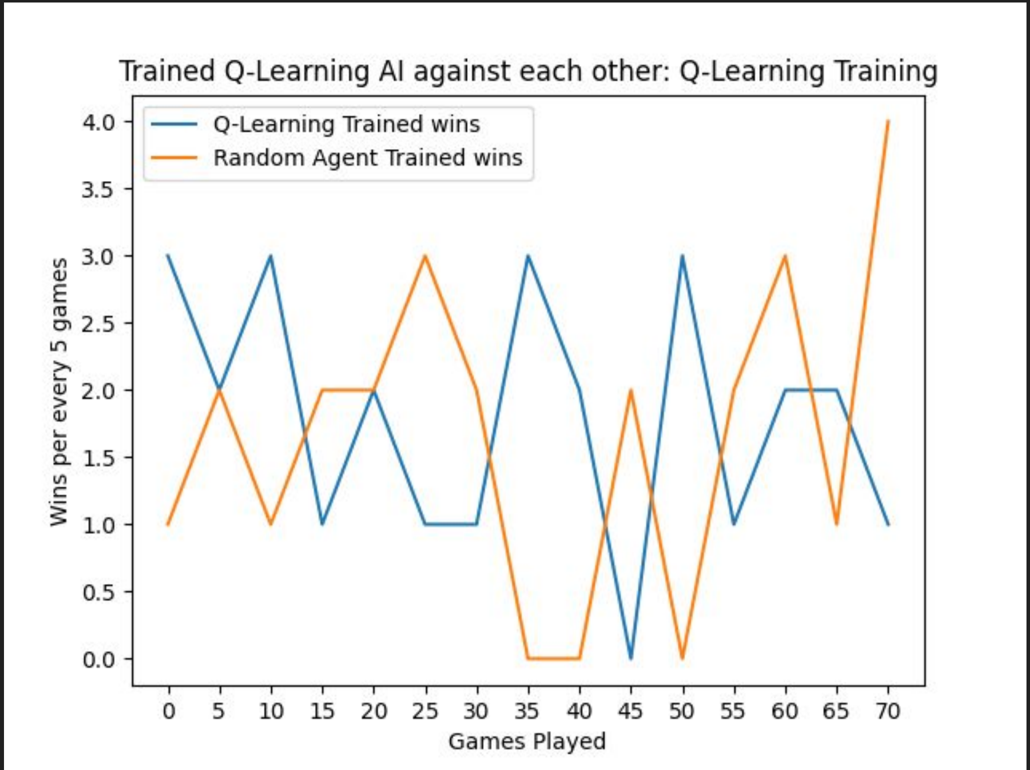
Q-Learning Trained Agent



Alpha-Beta Trained Agent

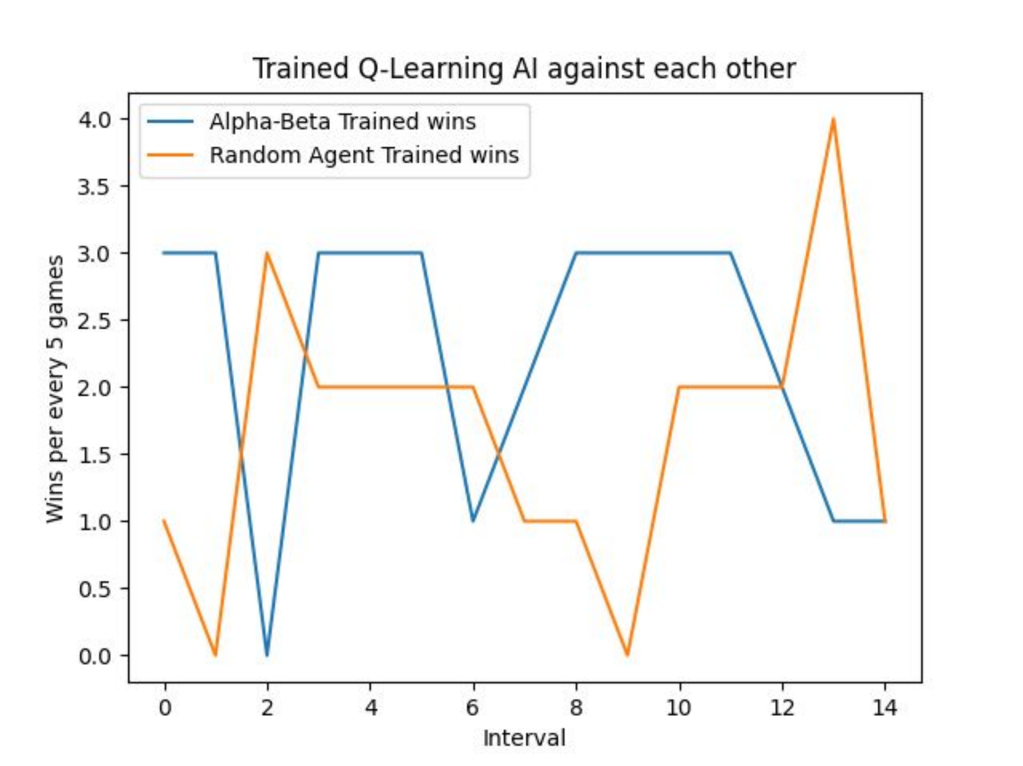


Q-Learning Results



Games Played:	75
Q-Learning Trained wins:	27
Random Agent Trained wins:	25
Games exceeded move limit:	17
Games tied:	6
Total moves made:	20471
Average moves made:	272.94666666666666
Max moves made:	1000
Min moves made:	32

Alpha-Beta Trained Results



Games Played:	75
Alpha-Beta Trained wins:	34
Random Agent Trained wins:	25
Games exceeded move limit:	12
Games tied:	4
Total moves made:	15930
Average moves made:	212.4
Max moves made:	1000
Min moves made:	35

Conclusion

- Agents who are trained against other learning agents tend to perform better than agents trained against random movesets

- Agent architecture is key determinant on the ability of an agent (i.e. trained agents do not hold up against Alpha-Beta Pruning)

Questions?